

Structural studies of DNA G-quadruplexes and ligand complexes

A thesis submitted for the degree of
Doctor of Philosophy of the University of London

Pascale Hazel

Cancer Research UK Biomolecular Structure Group
The School of Pharmacy
University of London

November 2005



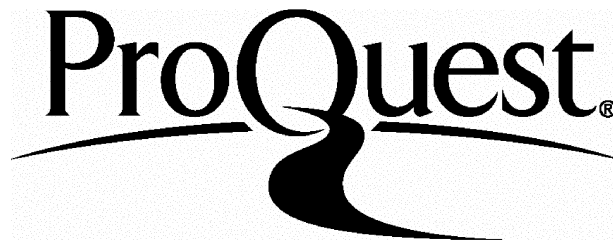
ProQuest Number: 10104789

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 10104789

Published by ProQuest LLC(2016). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code.
Microform Edition © ProQuest LLC.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

This thesis describes research conducted in the School of Pharmacy, University of London between October 2002 and November 2005 under the supervision of Prof. Stephen Neidle. I certify that the research described is original and that any parts of the work that have been conducted by collaboration are clearly indicated. I also certify that I have written all the text herein and have clearly indicated by suitable citation any part of this dissertation that has already appeared in publication.

A handwritten signature in black ink, appearing to be 'P. Neidle', written in a cursive style.

29/11/05

Acknowledgements

I would like to thank my supervisors, Prof. Stephen Neidle and Dr. Gary Parkinson, for guiding me through my research, and providing me with support. I would also like to thank the Association for International Cancer Research for providing me with the studentship that enabled me to carry out this work.

All members of the BMSG have contributed in some way to this work, and made my stay at the School of Pharmacy a productive, and fun three years. I would also like to thank Julian Huppert, who carried out the experimental work described in Chapter 4.

I would like to thank friends in London, and especially Jérôme, who, as a typical Frenchman, was the perfect person for exchanging the highs and lows of our respective research.

Finally, I would like to thank Andrew, for his invaluable help when dealing with computers, for teaching me Python, but mostly for always being there. This \LaTeX typeset thesis is for you !

Abstract

G-quadruplexes can form a wide variety of structures, depending, amongst others, on the number and length of the G repeats, the length and composition of the loops, and the cations binding within the central channel. Several X-ray and NMR structures of G-quadruplexes have been determined, however there are still no guidelines which enable the structure of a G-quadruplex to be determined from its sequence. A combination of X-ray crystallography and molecular modelling has been used to obtain more insight into this problem. Due to the large number of factors involved in quadruplex folding, this study was limited to the influence of loop length and composition.

The crystal structures of $(G_4T_3G_4)_2$ and $(G_4T_2AG_4)_2$ have been determined. Both sequences form antiparallel quadruplexes with lateral loops. Moreover, $(G_4^{Br}UT_2G_4)_2$ formed two different antiparallel structures, a head-to-tail and a head-to-head dimer within the same crystal, suggesting that these are very similar in energy. Prior to crystallisation, structure prediction was undertaken using molecular dynamics simulations. Although both lateral and diagonal T_3 and T_2A loops were equally possible, shorter T_2 loops restricted the quadruplex structures likely to form in solution.

Molecular dynamics simulations using $AG_3T_2AG_3T_2AG_3T_2AG_3$ X-ray and NMR structures as templates were able to show further effects of loop length on quadruplex folding. T and T_2 loops could form only parallel or lateral loops, however this was dependent on the length of all loops in the quadruplex. Longer loops could form equally stable parallel and antiparallel structures. Simulations of ligand-quadruplex complexes were dependent on the initial structure of the quadruplex, emphasising the importance of quadruplex structure knowledge for ligand design.

The combination of experimental and theoretical techniques has provided further insight into quadruplex structure, and has enabled the beginning of guidelines for quadruplex folding to be established.

Contents

1	Introduction	14
1.1	G-quadruplexes	14
1.2	G-quadruplexes within the genome	15
1.2.1	Telomeric G-quadruplexes	15
1.2.2	Non-telomeric G-quadruplexes	16
1.3	G-quadruplex formation in vivo	17
1.4	Structural diversity	19
1.4.1	X-ray and NMR structures	19
1.4.2	Biophysical techniques	23
1.5	Factors governing G-quadruplex folding	24
2	Molecular modelling	28
2.1	Molecular dynamics force fields	28
2.1.1	Amber force fields	29
2.1.2	Consistent and constant valence force fields (CFF and CVFF)	30
2.2	Conformational search in MD simulations	31
2.2.1	Simulated annealing	31
2.2.2	Locally enhanced sampling	32
2.3	Analysis of MD simulations	33
2.4	Applications of molecular dynamics	34
2.4.1	G-quadruplex structure	35
2.4.2	Ion binding	35
2.4.3	Loop conformation and folding	36
2.5	Limitations of molecular dynamics	37
2.6	Quantum mechanical calculations	39

3	(G₄T₂G₄)₂, (G₄T₂AG₄)₂ and (G₄T₃G₄)₂ structures	41
3.1	Predictive modelling	42
3.1.1	Computational methods	42
3.1.2	(G ₄ T ₃ G ₄) ₂	46
3.1.3	(G ₄ T ₂ AG ₄) ₂	58
3.1.4	(G ₄ T ₂ G ₄) ₂	61
3.1.5	Structure prediction	65
3.2	Crystal structures of G ₄ T ₃ G ₄ and G ₄ T ₂ AG ₄	65
3.2.1	Materials and methods	65
3.2.2	Crystal packing of the (G ₄ ^{Br} UT ₂ G ₄) ₂ P2 ₁ structure	68
3.2.3	Crystal packing in the C2 and C222 crystals	72
3.2.4	G-quartet structure	74
3.2.5	T ₃ and T ₂ A loop conformations	82
3.3	Comparison with other G-quadruplex structures	87
3.4	Comparison of X-ray structures and modelling	91
3.4.1	(G ₄ T ₃ G ₄) ₂ modelling	91
3.4.2	(G ₄ T ₂ AG ₄) ₂ modelling	97
3.4.3	(G ₄ T ₂ G ₄) ₂ modelling	101
3.4.4	Conclusion	103
4	Modelling of monomeric G-quadruplex folding	104
4.1	Computational methods	105
4.2	Simulation results	106
4.2.1	Single loop variants	106
4.2.2	Multiple loop variants	112
4.3	Experimental data	114
4.4	Comparison of simulations and experiment	119
4.4.1	CD spectroscopy results	119
4.4.2	MM-PBSA free energy calculations	119
4.4.3	Importance of interactions within the loops	124
4.4.4	G ₃ T structure	125
4.4.5	Limitations of the methods used	128
4.5	Loop-dependent folding	129
5	Modelling of G-quadruplex-ligand interactions	130
5.1	Ligand binding to G-quadruplexes	130
5.2	Computational methods	134

5.2.1	Docking	134
5.2.2	Molecular dynamics	135
5.3	Docking results	136
5.3.1	G ₃ T ₂ AG ₃ T ₂ AG ₃ T ₂ AG ₃ parallel receptor	136
5.3.2	(G ₄ T ₄ G ₄) ₂ antiparallel receptor	138
5.3.3	Energetic analysis	140
5.4	Molecular dynamics simulations	142
5.4.1	Parallel quadruplex complexes	142
5.4.2	Antiparallel quadruplex complexes	147
5.5	Free energy analyses	147
5.5.1	Single trajectory analysis	149
5.5.2	Separate trajectory analysis	151
5.5.3	Antiparallel quadruplex complexes	154
5.6	Conclusions	156
6	Ion binding within the G-quadruplex channel	158
6.1	Computational methods	159
6.1.1	Quantum mechanical calculations	159
6.1.2	Molecular dynamics simulations	160
6.2	Quantum mechanical calculations	160
6.3	Molecular dynamics simulations	162
6.4	Conclusions	163
7	Conclusions	166
	Bibliography	169
	Appendices	185
A	Clustering methodology	185
B	Ligand charges	187
B.1	JH ligand charges	187
B.1.1	JH001	188
B.1.2	JH002	189
B.1.3	JH003	190
B.1.4	JH004	191
B.2	B ligand charges	192

B.2.1	B124B	193
B.2.2	B128B	194
B.2.3	B120B	195
B.2.4	B123B	196

List of Figures

1.1	G-quartet and G-quadruplex	15
1.2	Telomeric single stranded overhang	16
3.1	Simulated annealing energy vs. cluster	49
3.2	SA lateral (narrow groove) T ₃ loops	50
3.3	SA diagonal T ₃ loops	50
3.4	LES lateral (wide groove) T ₃ loops	52
3.5	T ₃ loop rms deviations	53
3.6	T ₃ -L _n -2 MD LES simulation	55
3.7	T ₃ -D-1 MD simulation	55
3.8	T ₃ loop G _{total} variation over time	59
3.9	SA lateral and diagonal T ₂ A loops	60
3.10	SA T ₂ loops	62
3.11	T ₂ loop MD simulations	64
3.12	(G ₄ ^{Br} UT ₂ G ₄) ₂ P ₂ ₁ asymmetric unit	70
3.13	(G ₄ ^{Br} UT ₂ G ₄) ₂ numbering scheme	71
3.14	(G ₄ ^{Br} UT ₂ G ₄) ₂ P ₂ ₁ structure crystal packing	73
3.15	(G ₄ T ₃ G ₄) ₂ C222 structure crystal packing	74
3.16	(G ₄ ^{Br} UTAG ₄) ₂ C222 structure	75
3.17	(G ₄ T ₃ G ₄) ₂ G-quartet electron density	75
3.18	(G ₄ ^{Br} UT ₂ G ₄) ₂ P ₂ ₁ structure G-quartet stacking	76
3.19	(G ₄ ^{Br} UTAG ₄) ₂ Na ⁺ omit map	77
3.20	(G ₄ T ₃ G ₄) ₂ groove hydration	78
3.21	(G ₄ ^{Br} UT ₂ G ₄) ₂ narrow groove hydration	79
3.22	(G ₄ ^{Br} UT ₂ G ₄) ₂ wide groove hydration	80
3.23	(G ₄ ^{Br} UT ₂ G ₄) ₂ medium groove hydration	81
3.24	Type 1 loop conformation	84
3.25	P ₂ ₁ structure ^{Br} UT ₂ loop E	85

3.26	Type 2 loop conformation	86
3.27	(G ₄ ^{Br} UT ₂ G ₄) ₂ adjacent loop-loop interactions	88
3.28	(G ₄ T ₃ G ₄) ₂ and (GCGGT ₃ GCGG) ₂ T ₃ loops	91
3.29	T ₃ loop average MD structures	93
3.30	(G ₄ T ₃ G ₄) ₂ X-ray and average MD structure	94
3.31	Crystal structure simulations G _{total} variation over time	96
3.32	(G ₄ T ₂ AG ₄) ₂ average MD structure	99
3.33	(G ₄ T ₂ AG ₄) ₂ MD structure and Na ⁺ ion path	100
3.34	(G ₄ T ₂ G ₄) ₂ average MD structures	102
4.1	Parallel and antiparallel T ₄ loop clusters	107
4.2	Single parallel loop variant MD simulations	109
4.3	Experimental T ₂ A-G ₃ structures MD simulations	111
4.4	Single antiparallel loop variant MD simulations	113
4.5	Single mixed parallel/antiparallel loop variant MD simulations	114
4.6	Multiple loop variant MD simulations	115
4.7	CD spectra	117
4.8	Single loop variant G _{total} over time	122
4.9	G ₃ T antiparallel structure simulations	127
5.1	JH ligands	132
5.2	B ligands	133
5.3	Parallel quadruplex-B124B and JH001 docked structures	137
5.4	Parallel(4ns) quadruplex-JH001 and B124B docked structures	139
5.5	Lowest energy parallel quadruplex-JH ligand docked structures	139
5.6	Lowest energy antiparallel quadruplex-B124B docked structures	140
5.7	JH001 ligand MD simulations	144
5.8	JH004 ligand MD simulations	145
5.9	B ligand-parallel quadruplex MD simulations	146
5.10	B124B and B128B ligand antiparallel quadruplex simulations	148
5.11	B120B-3 antiparallel quadruplex MD simulation	149
5.12	Parallel quadruplex G _{receptor} variation over time	153
5.13	JH002-A complex free energy variation over time	155
6.1	(TG ₄ T) ₄ dimer X-ray structure	159
6.2	QM minimised ion-G-quartet structures	161
6.3	(TG ₄ T) ₄ dimer MD simulations	164

B.1	JH ligand numbering	187
B.2	B ligand numbering	192

List of Tables

1.1	Examples of genomic G-rich sequences	17
1.2	Tetrameric G-quadruplexes	19
1.3	Dimeric G-quadruplexes	21
1.4	Monomeric G-quadruplexes	22
3.1	T ₃ , T _{2A} and T ₂ loop types for structure prediction	43
3.2	SA lateral (wide groove) T ₃ loops	47
3.3	Predicted T ₃ loop free energies	57
3.4	Crystallisation conditions	67
3.5	Crystallographic data	69
3.6	P2 ₁ ^{Br} UT ₂ and C222 T ₃ loop B-factors	87
3.7	Crystal structure simulation G _{total} values	95
4.1	Monomeric quadruplex simulations	105
4.2	Single loop variant rms deviations	110
4.3	Multiple loop variant rms deviations	116
4.4	UV melting temperatures	118
4.5	Free energies of the monomeric quadruplexes	121
4.6	Free energies of monomeric quadruplexes using PARSE radii	124
5.1	ΔT_m values for the quindoline ligands	132
5.2	G-quadruplex-ligand MD simulations	141
5.3	Parallel G-quadruplex-ligand $\Delta G_{\text{binding}}$	150
5.4	$\Delta E_{\text{electrostatic}} + \Delta G_{\text{polar}}$ and $\Delta E_{\text{vdW}} + \Delta G_{\text{nonpolar}}$ for the parallel quadruplex complexes	152
5.5	Antiparallel quadruplex-B ligand $\Delta G_{\text{binding}}$	155
6.1	G-quartet-ion QM calculations	160
6.2	G-quartet-ion Mulliken and RESP charges	162

B.1	JH001 atomic charges	188
B.2	JH002 atomic charges	189
B.3	JH003 atomic charges	190
B.4	JH004 atomic charges	191
B.5	B124B atomic charges	193
B.6	B128B atomic charges	194
B.7	B120B atomic charges	195
B.8	B123B atomic charges	196

Chapter 1

Introduction

1.1 G-quadruplexes

The assembly of G-rich DNA sequences into multi-stranded quadruplex structures, formed from π -stacks of hydrogen bonded G-quartets, was first identified in 1962^{1,2}. G-quartets are planar arrangements of Hoogsteen hydrogen bonded G residues, as shown in Figure 1.1(a). G-quadruplex structures are formed by the successive stacking of G-quartets, with a 3.3 Å average separation, shown in Figure 1.1(b). The first G-quadruplex structures were solved in 1992, with the report of both X-ray³ and NMR structures⁴ of $(G_4T_4G_4)_2$. The former was later re-determined⁵. The $(TG_4T)_4$ structure shown in Figure 1.1(b) was also solved by NMR in 1992^{6,7}, and slightly later by X-ray crystallography⁸. Structures determined by NMR spectroscopy and X-ray crystallography are summarised in Section 1.4.1.

The formation of G-quadruplexes is cation-dependent, due to the strong negative electrostatic potential created by the G residue O atoms in the central channel of the G-quartet stack⁹. Quadruplexes require cations to bind within this channel for stability. The location of the cations between the quartets is ion dependent¹⁰, and also varies throughout the quadruplex. This is shown in Figure 1.1(b), with Na^+ ions within the channel. K^+ is another common G-quadruplex stabilising ion, however it occupies binding sites which are equidistant from the eight O atoms between each quartet plane.

G-quadruplexes can be formed by the assembly of four G-rich strands, each contributing one G residue to the G-quartets, or by the association and folding of two or a single strand. In the latter case, residues forming the G-quartets are separated by linker regions, which form the quadruplex loops. The combination of the number of G-stacks, the polarity of the strands, and the location and length of the loops leads to

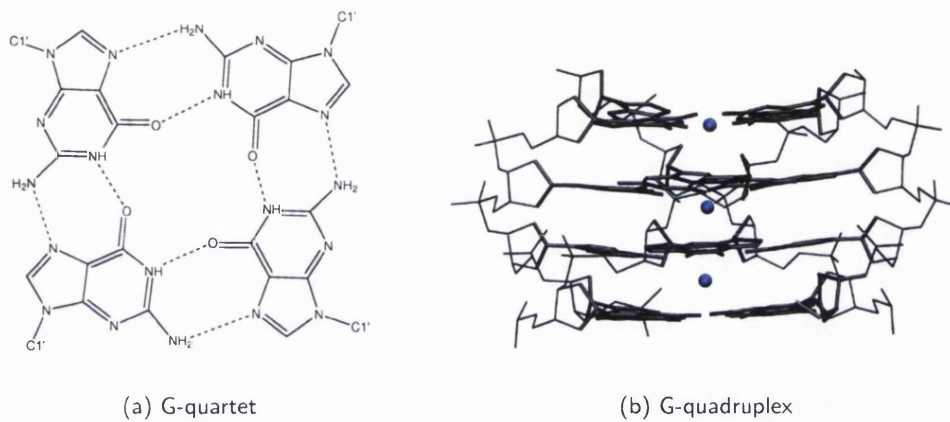


Figure 1.1: (a) G-quartet formed from Hoogsteen hydrogen bonded G residues and (b) $(TG_4T)_4$ structure formed from stacked G-quartets¹¹. The T residues have been omitted, and Na^+ ions within the central channel are shown as blue spheres.

the observed plurality of G-quadruplex structures. Potential monomeric G-quadruplex forming sequences can be described as follows:



where m is the number of G residues involved in G-quartet formation, and X_n , X_o and X_p can be any combination of residues, including G, forming the loops. It follows that if one of the runs of Gs is longer than the others, some of the G residues will be located in the loop regions.

1.2 G-quadruplexes within the genome

1.2.1 Telomeric G-quadruplexes

The most studied potential G-quadruplex forming sequences occur in the telomeric, or end region, of eukaryotic chromosomes. Telomeres are composed of non-coding tandem repeats of G-rich sequences. Most of telomeric DNA is double stranded, however the extreme 3' end consists of single-stranded G-rich DNA. The length of the single stranded overhang varies between species, but is around 100-200 bases in humans¹². Figure 1.2 shows a diagram of telomeric DNA, along with telomeric sequences found in various organisms. Telomeres protect the chromosome ends from events such as end-to-end fusions or formation of dicentric chromosomes through recombination. Telomeric DNA

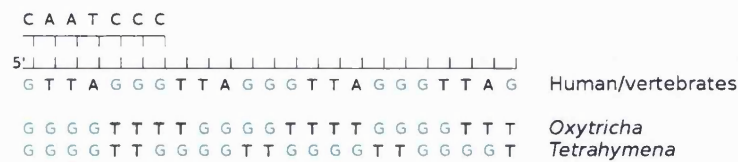


Figure 1.2: Schematic diagram of the telomeric single stranded overhang. Three examples of telomeric tandem repeats are listed, with G-runs in blue.

is extensively protein-bound, both on the double stranded portion (e.g. TRF1, TRF2)¹³ and along the single stranded overhang (e.g. POT1)¹⁴. These proteins prevent telomeric DNA from being recognised as damaged DNA, and eliciting an apoptotic response.

In normal somatic cells, telomeres are shortened at each cell division, due to the end-replication effect, whereby DNA polymerase is unable to fully replicate the lagging DNA strand. After 20 to 30 divisions, cells enter a growth arrest phase (replicative senescence), and apoptotic death can follow. In cancerous cells, there is no shortening of the telomeres at each cell cycle. Instead, the enzyme telomerase is expressed, which adds repeats onto the ends of the chromosomes, in order to maintain their length¹⁵. As telomerase is expressed in 80-85% of tumour cells, it is an attractive target for anti-cancer drugs¹⁶.

Telomerase inhibition can be achieved through targeting of the enzyme itself, or through targeting of its substrate, which is the single stranded G-rich overhang. The formation of a G-quadruplex type structure at the overhang could prevent telomerase from binding to its substrate, and inhibit telomere elongation¹⁷. Recently, Cech *et al* showed that G-quadruplexes are poor substrates of telomerase, and that the hPOT1 protein might be involved in disrupting them, enabling telomerase to bind and elongate its substrate¹⁸. Much effort is currently being put to developing ligands which promote the formation of G-quadruplexes, and these will be discussed further in Chapter 5.

1.2.2 Non-telomeric G-quadruplexes

G-rich sequences occur not only at the telomeres, but throughout the genome^{19,20}. There is growing interest in the potential of G-quadruplexes forming naturally elsewhere in the genome to perform functions such as gene transcription regulation. G-rich sequences have been found to occur in the promoter regions of some important genes, and within genes themselves, examples of which are shown in Table 1.1. Perhaps the most studied of these is the G-rich sequence found in the nuclease hypersensitivity element (NHE) III1 of the *c-myc* oncogene promoter. This is a 27 nucleotide long repeat with

Table 1.1: Examples of potential G-quadruplex forming sequences within the genome and their associated genes.

Sequence	Associated gene
TG ₄ AG ₃ TG ₄ AG ₃ TG ₄ A ₂ G ₂	<i>c-myc</i> oncogene ^{21,22}
G ₄ TGTG ₄ ACAG ₄ TGTG ₄	insulin-linked polymorphic region (ILPR) ²³
CG ₂ CG ₂ CG ₂ CG ₂	fragile X syndrome FMR1 gene ²⁴
TC ₂ G ₂ AG ₄ CAG ₂ CTGAG ₃ CG ₂ C	muscle creatine kinase (MCK) enhancer ²⁵
G ₁₆ CG ₃ TG ₂ TG ₂ T	chicken β -globin gene ²⁶

six G-runs, which has been shown to be able to form G-quadruplex structures²¹. It has been postulated that quadruplex formation is involved in *c-myc* transcription *in vivo*, although this has not yet been conclusively shown²².

The main difference between telomeric and non-telomeric G-rich sequences is that the telomere ends are single stranded. In contrast, genomic DNA is double stranded, and quadruplex formation is in competition with the duplex form. Whether quadruplex or duplex DNA is favoured has been shown to depend on sequence²⁷ and experimental conditions²⁸. A mixture of AG₃T₂AG₃T₂AG₃T₂AG₃ and C₃TA₂C₃TA₂C₃TA₂C₃T was shown to exist predominantly as a double helix in near-physiological conditions, however lowering the pH or temperature induced quadruplex formation²⁸. The C-rich strand can form an i-motif structure, which is stabilised at low pH values²⁹.

1.3 G-quadruplex formation in vivo

Possible biological functions of G-quadruplexes *in vivo* remain uncertain. However, the formation of G-quadruplexes has recently been shown to be controlled by telomere end-binding proteins *in vivo*, suggesting that these structures are involved in the capping of telomeres³⁰. Several quadruplex-binding proteins have also been reported. The yeast RAP1 protein^{31,32} and *Oxytricha* telomere binding protein^{33,34} bind telomeric G-quadruplexes, as well as facilitating their formation. Helicases in Bloom's³⁵ and Werner's³⁶ syndromes also have the ability to unwind G-quadruplex DNA. Bloom's syndrome helicase preferably unwinds quadruplex over duplex DNA, as shown by competition studies, suggesting that G-quadruplex removal is a function performed in the cell. The existence of several proteins that are capable of binding and unwinding G-quadruplexes suggests that these structures could form *in vivo*, and could have a role in telomere protection and/or gene regulation. The quadruplex-binding properties of certain proteins

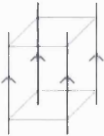
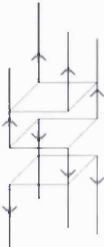
have been used in the development of G-rich sequences as therapeutic agents. Human α -thrombin is inhibited by the G-quadruplex forming $G_2T_2G_2TGTG_2T_2G_2$ aptamer³⁷, and the $G_4TG_3AG_2AG_3T$ aptamer has been shown to inhibit HIV-1 integrase³⁸.

Non-telomeric DNA exists as a nucleosome-packaged double strand during most of the cell cycle, however strand separation occurs during replication and transcription, at which point G-quadruplexes could form. Risitano and Fox have shown that some G-rich sequences such as the *Tetrahymena* and *c-myc* sequences preferentially form quadruplex over duplex DNA in K^+ containing solutions, even in the presence of excess complementary C-rich strand²⁷. Hence, separation of the two strands during transcription or replication could be sufficient to achieve quadruplex formation on the G-rich strand. On the other hand, the human telomeric sequence was found to adopt a quadruplex structure only at lower pH values, or higher temperatures, than the physiological conditions²⁸. Thermodynamic studies do show that the human telomeric sequence is in equilibrium between the duplex form and the quadruplex/i-motif form³⁹. Which structure dominates is dependent on pH and cationic species. The small difference in stability between duplex and quadruplex DNA however suggests that the equilibrium could be shifted towards quadruplex formation by ligand binding³⁹.

Whether G-quadruplexes occur naturally *in vivo* or not, their formation can potentially be induced by binding to quadruplex-stabilising ligands. Drugs are currently being developed to target both telomeric and non-telomeric G-quadruplex forming sequences. This allows the indirect targeting of proteins which would usually bind to the DNA single strand, and which become incapable of doing so. This approach has been encouraged by the results of cellular experiments. Many ligands for which G-quadruplex binding is the believed mode of action have been shown to inhibit telomerase. These include anthraquinones⁴⁰, porphyrins⁴¹, triazines⁴², acridines⁴³ and telomestatin⁴⁴. A G-quadruplex stabilising porphyrin ligand has also been shown to repress the *c-myc* gene transcription²². The characteristics of these G-quadruplex-binding ligands will be discussed in Chapter 5.

The design of selective G-quadruplex ligands requires a knowledge about the structure of their receptor. Selectivity is of paramount importance, due to the large number of G-rich sequences throughout the genome which could potentially form G-quadruplexes upon ligand binding. The G-rich sequences described above have varying G and loop regions, and can form a wide variety of quadruplex structures, which could be the key to their targeting.

Table 1.2: G-quadruplex structures formed from four or more strands. G-quartets at the extremities are shown with planes, any number can occur in between those.

Structural type	Examples
<p>Parallel</p> 	<p>$(TG_4T)_4$ X-ray^{8,11} and NMR^{6,7}</p>
<p>Interlocked</p> 	<p>possible structure for $(G_3T)_8$⁴⁵</p>

1.4 Structural diversity

1.4.1 X-ray and NMR structures

The simplest G-quadruplex structures are formed from the association of four single G-rich strands, as shown in Table 1.2. In $(TG_4T)_4$ X-ray^{8,11} and NMR structures^{6,7}, the strands are parallel to one another and the G residue glycosidic torsion angles are all *anti**. However, even tetrameric G-quadruplexes can form more complex structures, as shown by $(G_3T)_8$, in which eight strands form interlocked quadruplexes⁴⁵. An example of such a structure is shown in Table 1.2.

Dimeric quadruplexes are formed by two G-rich strands. Each strand can contribute two G residues to the G-quartets, however asymmetric dimerisation of three strands from one molecule and a single strand from a second has also been observed⁴⁶ (Table 1.3). Examples of dimeric structures are shown in Table 1.3. Structures are generally classified according to the chain polarities and the location of the loops which link the G residue strands. Four parallel strands require the loops to link the bottom G-quartet with the top G-quartet, leading to strand-reversal, or propeller type loops. This structural type

*In the quadruplex notation adopted here, sequences within brackets denote individual strands, and the composition of each strand is always written explicitly. The brackets and number correspond to the quadruplex stoichiometry, i.e. the number of strands which associate to form a whole quadruplex.

has been found both in crystal structures⁴⁷ and in solution⁴⁸.

Antiparallel quadruplexes are formed when at least one of the four strands is antiparallel to the others. This represents the majority of dimeric structures solved to date. Two types of loops exist in these structures. Lateral loops join adjacent G-strands, as in the $(TG_4T_2G_4T)_2$ structure⁵⁰. These can be located either on the same, or opposite faces of the G-quartets, corresponding to head-to-head or head-to-tail dimerisation, respectively. Strand polarities can vary, as in the case of the head-to-tail lateral loop dimers in which both all-antiparallel and a mixture of parallel/antiparallel strands has been found (see Table 1.3). The second type of loop joins diagonally opposed G-strands, as exemplified by the $(G_4T_4G_4)_2$ ^{5,51} structure. In this case the strand polarities alternate between parallel and antiparallel, as shown in Table 1.3.

The dimeric quadruplex structures have been described according to strand polarity and loop type, however nothing prevents a mixture of parallel and antiparallel loops in a single quadruplex. Mixed loop quadruplexes have been found, and $(G_3T_4G_4)_2$ is an example of a dimeric quadruplex in which both lateral and diagonal loops are formed⁵⁷.

Table 1.4 shows examples of monomeric quadruplex structures, which are formed from the folding of a single G-rich strand. The same three loop types (parallel, lateral and diagonal) found in dimeric quadruplexes occur in monomeric structures. An all-parallel stranded structure can form with three strand-reversal loops, as in the X-ray structure of $AG_3T_2AG_3T_2AG_3T_2AG_3$ ⁴⁷. Several other sequences, mostly derived from the *c-myc* promoter sequence, have been found to adopt a parallel structure in solution. Parallel strand-reversal loops are also found in conjunction with lateral or diagonal loops, as in the $T_2G_4T_2G_4T_2G_4T_2G_4$ ⁵⁸ and $G_2T_4G_2CAG_2GT_4G_2T$ ⁵⁹ NMR structures.


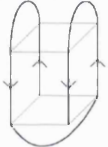
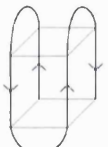
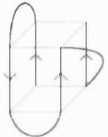
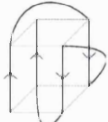
Tables 1.2-1.4 illustrate the structural diversity of G-quadruplexes. Only the structures solved with X-ray crystallography or NMR have been included, and it is clear that the possible variations of strand polarities and loop type combinations are not all represented. This by no means implies that these structures cannot be formed, as new structural types are continually being discovered.

The number of experimental structures determined is relatively small when compared to the conformational flexibility of G-quadruplexes. One of the main problems with determining G-quadruplex structures in solution is that in many cases there is no single structure formed. A G-rich sequence can fold in several different ways. This can be due to the formation of different loop conformations, such as lateral or diagonal loops. Many non-telomeric G-rich sequences contain unequal numbers of G residues in each G-run, meaning that the chain segment involved in G-quartet formation can vary. This is the case in the G-rich sequence found in the promoter region of the *c-myc* gene, for

Table 1.3: Dimeric G-quadruplex structures (X-ray or NMR) which have been solved to date. Strand polarity is indicated by arrows in the diagrams. Structures which exist in several forms are shown by $\downarrow\uparrow$.

Structural type	Examples	
Parallel loops	human telomeric sequence (TAG ₃ T ₂ AG ₃ T) ₂ x-ray ⁴⁷	(TAG ₃ UTAG ₃ T) ₂ NMR ⁴⁸
Antiparallel lateral loops	(G ₄ T ₃ G ₄) ₂ X-ray ⁴⁹	(UAG ₃ T ^{Br} UAG ₃ T) ₂ NMR ⁴⁸
	$\downarrow\uparrow$	
	(G ₄ ^{Br} UT ₂ G ₄) ₂ X-ray ⁴⁹	<i>Tetrahymena</i> telomeric sequence (TG ₄ T ₂ G ₄ T) ₂
	$\downarrow\uparrow$	
		(TG ₄ T ₂ G ₄ T) ₂ NMR ⁵⁰
	(G ₃ T ₂ AG ₃ T ₂ AG ₃) ₂ NMR ⁴⁶	
Antiparallel diagonal loops	<i>Oxytricha</i> telomeric sequence (G ₄ T ₄ G ₄) ₂ X-ray ^{5,51} and NMR ^{4,52,53}	(G ₄ T ₄ G ₃) ₂ NMR ⁵⁴ (G ₃ T ₄ G ₃) ₂ NMR ^{55,56}
Antiparallel mixed lateral/diagonal loops	(G ₃ T ₄ G ₄) ₂ NMR ⁵⁷	

Table 1.4: Monomeric G-quadruplex structures (X-ray or NMR) which have been solved to date. Strand polarity is indicated by arrows in the diagrams. Structures which exist in several forms are shown by $\downarrow\uparrow$.

Structural type	Examples	
Parallel loops		
	human telomeric sequence AG ₃ T ₂ AG ₃ T ₂ AG ₃ T ₂ AG ₃ X-ray ⁴⁷	<i>c-myc</i> structures TGAG ₃ TG ₃ TAG ₃ TG ₃ TA ₂ NMR ⁶⁰ TGAG ₃ TG ₄ AG ₃ TG ₄ A ₂ NMR ⁶¹ TG ₄ AG ₃ T ₅ AG ₃ TG ₄ A NMR ⁶¹
Antiparallel mixed lateral/diagonal loops	$\downarrow\uparrow$	
	AG ₃ T ₂ AG ₃ T ₂ AG ₃ T ₂ AG ₃ NMR ⁶²	<i>Oxytricha</i> telomeric sequence G ₄ T ₄ G ₄ T ₄ G ₄ T ₄ G ₄ NMR ⁶³
Antiparallel lateral loops		
	Thrombin binding aptamer G ₂ T ₂ G ₂ TG ₂ T ₂ G ₂ X-ray ³⁷ and NMR ⁶⁴	G ₃ TG ₃ TG ₃ TG ₃ NMR-based model ⁶⁵
Mixed parallel/antiparallel loops		
	<i>Tetrahymena</i> telomeric sequence T ₂ G ₃ GT ₂ GG ₃ T ₂ GG ₃ T ₂ G ₃ G NMR ⁵⁸	
	G ₂ T ₄ G ₂ CAG ₂ GT ₄ G ₂ T NMR ⁵⁹	

example⁶¹. The wild-type sequence, TG₄AG₃TG₄AG₃TG₄A₂G₂, contains six G-runs, one with two G residues, two with three Gs and three with four Gs. Many different quadruplex folds are possible, and indeed shown to be formed in solution by NMR spectra of the native sequence⁶¹. In order to obtain a single structure and interpretable NMR spectrum, the sequence generally has to be truncated and/or mutated in order to include only four runs of Gs. The modified structures which have been solved by NMR spectroscopy in this manner are shown in Table 1.4.

1.4.2 Biophysical techniques

In order to gain more information about quadruplex structures in solution, many studies have been carried out using biophysical techniques. These techniques can reveal structural information about quadruplexes, however they generally rely heavily on the availability of suitable structural models. The possibility of intramolecular quadruplexes with parallel loops was only discovered in 2002 with the publication of the X-ray structure of AG₃T₂AG₃T₂AG₃T₂AG₃⁴⁷. Until then, only lateral and diagonal loop structures were considered when envisaging possible models for quadruplex structures. This can affect the conclusions drawn from experiments reported before that date.

Combinations of techniques such as low-resolution NMR, circular dichroism (CD) spectroscopy, IV/UV melting or fluorescence resonance energy transfer (FRET) are frequently used to study G-quadruplexes. CD spectroscopy is commonly used to verify the presence of quadruplexes in solution. G-quadruplexes give characteristic peaks which can differentiate them from duplex and other DNA structures. Different quadruplex structures also give different CD profiles depending on the polarity of the G-strands, and thus parallel and antiparallel quadruplexes have different CD spectra^{32,66,67}. Parallel quadruplexes are associated with a maximum around 260 nm, and a minimum around 240 nm, and antiparallel quadruplexes with a maximum around 295 nm, and a minimum around 260 nm. The CD signal is correlated to G glycosidic torsion angles, with the all-*anti* G strands in parallel quadruplexes giving a different signal to the alternating *syn-anti* torsions in antiparallel quadruplexes. However, when there are multiple conformations present in solution, as is often the case, CD spectra can become difficult to interpret⁶⁸.

Thermodynamic data for the formation and dissociation of G-quadruplexes can be calculated using T_m measurements from temperature-dependent absorbance between 240 and 295 nm⁶⁹. T_m values are frequently used to determine the effect of sequence⁷⁰, cation and cation concentration^{66,69,71,72}, etc., on the stability of G-quadruplexes. These studies have shown that cation stabilisation generally follows the trend K⁺ > Rb⁺ > Na⁺ > Cs⁺ > Li⁺ for most quadruplexes. The ionic effects are,

however, not only limited to stabilisation as different cations can favour different structural types. Structural changes can be identified by CD spectroscopy, even if the structural characterisation is not straightforward. For example, both TG₃T and TG₃T₂G₃T show different CD profiles in K⁺ and Na⁺ containing solutions⁷³.

A combination of CD, chemical probing and gel electrophoresis was used to investigate structural features of human telomeric repeat sequences of various lengths⁷⁴. T₂AG₃T₂AG₃T₂AG₃T₂AG₃ and G₃T₂AG₃T₂AG₃T₂AG₃ were shown to form intramolecular antiparallel quadruplexes, while G₃T₂AG₃ dimerises to form an intermolecular antiparallel quadruplex. Chemical probing was further used to show that the loops of the dimer are adjacent on the same G-quadruplex face, and that two G-quadruplex dimers stack end to end, forming a dimer of dimers. The effect of loop length and sequence on the structure of the thrombin binding aptamer was also studied using UV melting, CD spectroscopy and differential scanning calorimetry (DSC)⁷⁰. Thermodynamic parameters showed that changing a single residue in the loops could greatly impact the quadruplex's stability. For example, changing a loop from TGT to TT caused a decrease in the T_m of around 25 °C.

The techniques described above have enabled several potential quadruplex structures to be added to those described in Section 1.4.1. These have allowed the impact of sequence on stability, such as the number of G-stacks or residues in the loops, to be studied, as well as the effect of changing the monovalent ions bound within the quadruplex channel. However, these methods are not always reliable in providing structural information. CD spectroscopy is widely used to determine whether quadruplex structures are parallel or antiparallel, however its interpretation is not always straightforward. For example, NMR and modelling data suggest that the HIV integrase aptamer G₃TG₃TG₃TG₃T adopts an antiparallel structure with three lateral loops, similar to the thrombin binding aptamer structure shown in Table 1.4⁶⁵. However, the CD spectrum of the same sequence is indicative of a parallel structure, with maxima around 210 and 265 nm, and a minimum around 240 nm⁷².

1.5 Factors governing G-quadruplex folding

X-ray, NMR and biophysical studies of G-quadruplexes have revealed the extent of G-quadruplex structural plurality. There are as yet no set of rules defining the folding of G-quadruplexes, however many influencing factors have been identified, and the most important are described below.

- The length of the loops has an effect on quadruplex structure. CD and gel elec-

trophoresis experiments have shown that $(G_4TG_4)_2$ folds into a parallel quadruplex, $(G_4T_3G_4)_2$ and $(G_4T_4G_4)_2$ fold into antiparallel quadruplexes, while $(G_4T_2G_4)_2$ forms a mixture of both conformations⁶⁷.

- The length of the loops affects quadruplex stability. Smirnov and Shafer systematically changed the loop lengths and residues in the thrombin binding aptamer, and obtained thermodynamic parameters for each modified sequence using UV melting experiments⁷⁰. All modifications, including shorter and longer loops, were found to destabilise the G-quadruplexes formed. On the other hand, Risitano and Fox found that $G_3TG_3TG_3TG_3$ was more stable than $G_3T_2G_3T_2G_3T_2G_3$ ²⁷.
- Changing the loop composition can change stability. Replacing the TTA loops with AAA loops in $TG_3T_2AG_3T_2AG_3T_2AG_3$ causes a decrease in stability, such that the T_m could no longer be measured²⁷.
- Interactions between the loop residues are important. $(G_3T_4G_3)_2$ folds into an antiparallel dimeric G-quadruplex in Na^+ solution, with hydrogen bonding interactions between certain T residues in the loops⁵⁶. Replacing each T in turn with a C residue causes structural changes, as identified by gel electrophoresis and CD spectroscopy⁷⁵.

Replacement of the T_4 loops in $(G_4T_4G_4)_2$ with non-nucleosidic linkers destabilises the structures, suggesting the importance of stacking and hydrogen-bonding interactions⁷⁶. Similarly, melting experiments on the HIV integrase aptamer $G_3TG_3TG_3TG_3$ in which the T loops were replaced by non-nucleosidic linkers of the same length shows that this sequence is less stable than the native⁷⁷. The T_m data suggests that both T and non-nucleosidic loop sequences fold in an analogous manner with fewer than 3 G-quartet stacks. This implies that the stacking interactions formed by the loop residues in $G_3TG_3TG_3TG_3$ contribute to the stability of the structure. On the other hand, the same study showed that when longer loops were replaced with non-nucleosidic linkers, the stability was increased.

- Increasing the length of the G-runs does not necessarily increase stability. For example, $G_4T_2G_4T_2G_4T_2G_4$ is less stable than $G_3T_2G_3T_2G_3T_2G_3$ ²⁷.
- The number of G-strand repeats can profoundly change the structure adopted. The *Tetrahymena* telomeric sequence $T_2G_4T_2G_4T_2G_4T_2G_4$ folds into the mixed parallel/antiparallel structure with three G-quartets shown in Table 1.4, whereas $(TG_4T_2G_4T)_2$ forms a dimeric structure with four G-quartets and lateral loops, shown in Table 1.3.

- Different cations can modulate quadruplex structure. Rujan *et al* have shown that quadruplexes formed from repeats of the human telomeric sequence T₂AG₃ form different structures in solutions containing high Na⁺ or high K⁺ concentrations⁶⁶. CD spectra in K⁺ suggest a parallel structure, compared to an antiparallel structure in Na⁺.

Cations can also induce changes in the conformation of the loops. (G₄T₄G₄)₂ forms the same antiparallel dimeric structure in Na⁺ and K⁺ solutions (shown in Table 1.3), however the loop conformations are different⁵².

- Temperature affects structure. Phan and Patel have shown that UAG₃T^{Br}UAG₃T preferentially forms an antiparallel structure at temperatures below 50 °C, and a parallel structure at higher temperatures⁴⁸. Both conformations have different rates of folding and unfolding, which are also temperature dependent.

Predicting the structure adopted by G-rich sequences is beyond the scope of present knowledge. The examples above show how many different factors can contribute to the folding of G-quadruplexes. Another difficulty when attempting to predict, and also determine, quadruplex structure is the possible coexistence of several structural forms in solution. It was mentioned in Section 1.4.1 that sequences with unequal numbers of G residues in each G-run can form several different structures, depending on which G residues form the G-tetrads. However, it has emerged that the formation of multiple structures is not limited to such cases, and that it is a relatively common occurrence.

The human telomeric repeat TAG₃T₂AG₃T forms several structures in solution, as shown by NMR spectroscopy⁴⁸. Substitution of selected T residues with U or BrU was used to shift the equilibrium from one form, a parallel stranded dimer with double-strand reversal loops, to another, a dimeric antiparallel quadruplex with lateral loops⁴⁸. Both structures are illustrated in Table 1.3. The longer sequence AG₃T₂AG₃T₂AG₃T₂AG₃ can also form parallel and antiparallel stranded structures, both of which are present in solution. This has been shown using cross-linking by mononuclear platinum complexes⁷⁸. (TG₄T₂G₄T)₂ forms two antiparallel structures in solution, with adjacent or opposite lateral loops (see Table 1.3). In this case, both structures are similar in energy, and were solved simultaneously by NMR⁵⁰. On the other hand, certain sequences form a single, well defined species in solution and in the crystal, such as (G₄T₄G₄)₂^{4,5,51-53}. The reason for these differences is not clear.

The aim of the work in this thesis is to derive guidelines for the folding of G-quadruplexes. As there are many factors which may influence structure, the study was restricted to loop length and composition effects. X-ray crystallography was used in

order to determine the structure of quadruplexes with varying loops, and experimental data was complemented with molecular modelling studies.

Chapter 2

Molecular modelling

Computational methods are increasingly being used to complement and interpret experimental data. Both molecular dynamics (MD) and quantum mechanical (QM) calculations have been used in this work. As discussed in Chapter 1, structural data for G-quadruplexes is not always readily available experimentally. Lower resolution techniques, such as chemical probing and CD, often require theoretical models for interpretation of the results. Classical force field methods such as minimisation and molecular dynamics are common techniques used to explore the conformation of molecules in solution.

Until approximately 1995, MD simulations of DNA suffered from instabilities, even though protein simulations were much more stable at that time. DNA structures distorted during the simulations, and base pairing was not maintained. This was due in part to the inappropriate handling of long-range electrostatic interactions. Simple truncation methods of dealing with these were unable to represent the electrostatics in highly charged systems such as DNA and RNA, but fared much better with proteins. In 1995, Kollman *et al* reported MD simulations of DNA double helices which were able to reproduce the experimentally observed B-DNA structure⁷⁹. This was possible due to the development of new methods to properly account for the long-range electrostatic interactions, such as the particle mesh Ewald (PME) method^{80,81}. Stable DNA and RNA simulations can now be routinely carried out on tens of nanosecond timescales⁸².

2.1 Molecular dynamics force fields

Several MD force fields have been developed to model protein and nucleic acid structure and function. Different force fields are parameterised to reproduce different molecular properties, and hence must be selected according to the system and properties to be

determined. Amber ff99⁸³ (for MD simulations), CVFF⁸⁴ (for docking) and CFF⁸⁵ (for simulated annealing) were used in the present study. The Amber force field, within the Amber 8 package⁸⁶, was used most extensively and is discussed in detail in Section 2.1.1. The Amber general atom force field (gaff)⁸⁷ was used for simulations of small ligand molecules, and will be discussed further in Chapter 5. The Amber force field has been shown to reliably simulate DNA systems⁸², and was chosen for consistency with previously reported simulations of G-quadruplexes^{88–94}.

2.1.1 Amber force fields

The general form of the pairwise potential energy function in the Amber force fields is shown in Equation 2.1. The bond and angle energies are approximated by quadratic potentials, where $K_x, x = (r, \theta)$ is the force constant, and r_{eq}, θ_{eq} are the equilibrium bond length and angle, respectively. The torsional energy is represented by a truncated Fourier series, where V_n, n, ϕ, γ are the torsional barrier, periodicity, dihedral angle and phase respectively. The final term includes all non-bonded interactions. A Lennard-Jones potential represents the van der Waals interactions, where R_{ij} is the distance between atoms i and j , and A_{ij}, B_{ij} are coefficients derived from the well depth and collision diameter of i and j . The electrostatic term is calculated from the Coulombic interactions of atom-centred point charges, with q_i, q_j the atomic point charges of atoms i and j . This is an effective pairwise potential, so that many-body effects are included in the model, by proper parameterisation, without requiring any further terms in Equation 2.1.

$$\begin{aligned}
 E_{\text{total}} = & \sum_{\text{bonds}} K_r (r - r_{\text{eq}})^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_{\text{eq}})^2 \\
 & + \sum_{\text{dihedrals}} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] \\
 & + \sum_{i < j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\epsilon R_{ij}} \right]
 \end{aligned} \tag{2.1}$$

The complete force field includes both the potential function, and the set of empirical parameters in Equation 2.1. The latter were determined by empirically fitting Equation 2.1 to reproduce certain properties, for example free energies of solvation or conformational energies⁹⁵. As ff99 was developed for the simulation of proteins and nucleic acids, it contains parameters for the common amino acids and nucleotides. The Amber force field was parameterised to reproduce the conformational energies and intermolecular interactions of proteins and nucleic acids, and is hence suited to the structural studies carried out in this work⁹⁵. The potential function contains fewer terms than

many other force fields, however the simulations carried out since its development show that this appears to be sufficient to model most biological systems. The Amber force field has been found to perform well for nucleic acid simulations compared to other force fields⁹⁶.

The set of parameters used in the force field are as important, if not more, as the form of the potential⁸³. Parameterisation of ff99 was carried out using a mixture of experimental data and simulations⁹⁵. For example, r_{eq} and θ_{eq} were obtained from X-ray structures, while the force constant K_{θ} was derived from vibrational analyses of simple systems, such as benzene. Dihedral parameters were kept simple, and in many cases only two atoms are used to define the torsion angles. This more general approach has the advantage of being less dependent on the set of molecules used for the initial parameter development. This should enable the force field to be applied to a greater range of biomolecular systems, without requiring any reparameterisation. Van der Waals parameters were derived from Monte Carlo simulations of small systems, and the values were adjusted in order to reproduce the density and enthalpy of vaporisation of liquids. Restrained electrostatic potential (RESP) fitting to quantum mechanically-calculated electrostatic potentials yielded the atomic charges used in the Coulombic term^{97,98}. The Hartree-Fock (HF) level of theory and 6-31G* basis set were used in the QM calculations⁸³.

As ff99 was mostly developed for the simulation of biomolecules, a separate general atom force field (gaff), was used for the ligand molecule simulations⁸⁷. Gaff has been parameterised specifically for small molecule ligands and is designed to be used in conjunction with ff99 in simulations of DNA- or protein-ligand complexes. The form of the potential is the same for both ligands and biomolecules (Equation 2.1), although the parameters are adapted to each.

The simplicity of the Amber force field means that it can be applied to a large range of biomolecules and small molecule ligands. Parameters for new molecule types can be derived and added to the existing model. Charges can be determined for ligand molecules using the RESP fitting approach, although this requires QM minimisation of each new molecule, before calculation of the electrostatic potential. RESP charges, however, can be conformation dependent, and the choice of which conformation to use is unclear⁸³.

2.1.2 Consistent and constant valence force fields (CFF and CVFF)

CFF is a class II force field, and its potential function includes many more terms than the ff99 potential in Equation 2.1. In CFF, the bond and angle terms are anharmonic and include cubic and quartic terms as well as the quadratic term in Equation 2.1. Moreover,

cross-terms between the bond lengths, angles and dihedrals are included. Class II force fields are complex, and hence require a large number of parameters. Parameterisation of CFF has been carried out using least-squares fitting of parameters to the potential. Most of the initial parameters were derived from *ab initio* calculations^{85,99}, giving the force field its consistency. This is in comparison to ff99, where the empirical derivation of parameters necessarily means that parameters come from many different sources.

CVFF was used when docking ligands onto quadruplexes, as CFF is not available in the Docking module of Insight II¹⁰⁰. This force field again uses a different potential function, and in this case the quadratic bond stretching term is replaced by a Morse potential. The potential also contains cross-terms, however these were not used as they can become unstable when structures are far from a minimum, which can be the case when docking molecules to a receptor.

2.2 Conformational search in MD simulations

MD simulations, using the force fields above, sample the dynamics of a system over time, yielding a trajectory from which time-averaged properties can be calculated. The conformational changes of a molecule through time can be monitored, and its energy can be calculated, for example. MD simulations are useful when simulating a system for which the coordinates are known, for example from X-ray or NMR data. However, when starting from a model-built structure which may be far away from the minimum energy conformation, methods for increasing the conformational sampling during simulations are often required. This can be achieved by giving more kinetic energy to the system, so it is able to overcome higher energy barriers, or by reducing the height of the barriers between conformations. Using these methods, a much greater range of conformations can be sampled during a simulation.

2.2.1 Simulated annealing

MD simulations can easily become trapped in a local potential energy well, which might be far from, and much less energetically favourable than, the global minimum. This is not a problem when simulating experimental structures which are already close to the minimum. However, in order to observe many conformations of a molecule when there is no experimental model, a much larger subset of conformational space needs to be sampled than can be achieved simply by room temperature simulations. In order to achieve this, the simulations can be carried out at higher temperatures, thereby

increasing the kinetic energy of the molecules. The latter are then able to overcome high potential energy barriers.

Generally, simulated annealing simulations involve a period of heating to a high temperature (for example 1000 K), followed by rapid cooling of the system. This enables the molecule to “fold” back into a close potential energy minimum. Simulations at lower temperatures such as 400 or 500 K can also be carried out for longer time periods, although there can be a danger of completely “unfolding” a molecule.

Simulated annealing allows the molecule to pass over high potential energy barriers, and hence sample configurational space more fully. However, the number of annealing runs necessary to find the global minimum is prohibitively high in practise when looking at biomolecules. Instead, simulated annealing runs are generally carried out for shorter time periods which are not guaranteed to find the global minimum, but rather yield a set of low energy conformations of the molecule. These conformations can then be analysed in terms of their energy, or their stability in further simulations.

2.2.2 Locally enhanced sampling

Enhanced conformational sampling can also be carried out with MD simulations without necessarily using the simulated annealing procedure described above. The Amber set of programs enables locally enhanced sampling (LES) to be carried out during simulations^{101,102}. The LES method is particularly useful when one part of the system (e.g. the quadruplex loops) is more flexible than the rest (e.g. the G-quartet stem). In LES, the flexible loop regions are replaced with multiple copies of themselves, which do not interact with one another. Each copy interacts in an average way with the rest of the system outside the LES region. In this manner, the non-LES part of the system interacts with the average energy of all the individual LES copies. As each LES copy can move independently from the others, they can sample different conformations within a single simulation, with only a limited increase in computational cost. Moreover, as the system interacts with an average of the LES copy energies, each copy has a reduced barrier, and can more easily move across the potential energy landscape. This enables each LES copy to sample more space than a single loop in a non-LES simulation. The LES method can also be combined with simulated annealing runs as described in Section 2.2.1

LES simulations can either be carried out in implicit or explicit solvent. The advantage of carrying out implicit solvent simulations in this case is that each LES copy can interact with its own continuum solvent approximation¹⁰³. In explicit solvent, the solvent molecules only see an average of the LES copies, and these cannot therefore drift too far apart or large gaps would be created in the solvent. This means that in

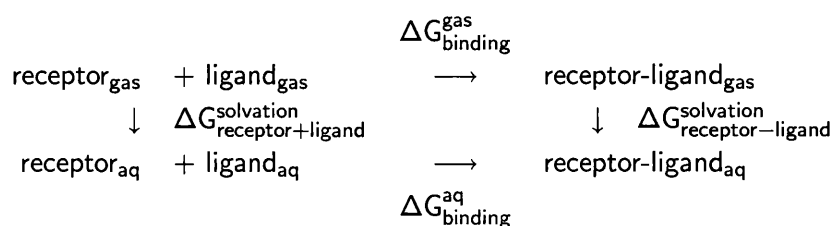
explicit solvent simulations the LES copies have less individual freedom. However, each copy still has lower energy barriers than in single-copy non-LES simulations.

2.3 Analysis of MD simulations

The stability of molecules during MD simulations can be assessed from simple visualisation of the trajectory over time, and by calculation of values such as root mean square (rms) deviations. This gives a first indication of whether a system is stable in solution. However, when comparing different possible conformations of a molecule, or ligand-receptor complexes, free energy information is the ultimate goal.

Binding free energy differences can be calculated accurately by several methods such as free energy perturbation (FEP) and thermodynamic integration (TI). These are numerical methods which can be used on biological molecules; however, as sufficient statistical sampling is necessary, they are very time consuming. Moreover, these methods only allow the calculation of binding energy differences and not absolute binding free energies. Recently, other methods have been developed which are less accurate, but enable absolute binding free energies to be estimated. Examples of these methods are the linear interaction energy method¹⁰⁴, and the molecular mechanics-Poisson Boltzmann/Surface area (MM-PBSA) method, which was used in this work^{105,106}. MM-PBSA is a post-processing method which allows the evaluation of absolute free energies and absolute binding free energies from an MD trajectory. The calculated energy is averaged over a subset of conformations sampled during the simulation.

Binding free energies are calculated according to the thermodynamic cycle below:



$\Delta G_{\text{binding}}$ in solvent is determined according to:

$$\Delta G_{\text{binding}} = G_{\text{aq}}(\text{complex}) - [G_{\text{aq}}(\text{receptor}) + G_{\text{aq}}(\text{ligand})]$$

Each of the above terms is calculated using the equations below. $G_{\text{aq}}(\text{receptor})$ can be calculated on its own in the same manner if absolute free energies are being estimated rather than binding free energies.

$$G_{\text{aq}} = E_{\text{MM}} + \Delta G_{\text{solvation}} - TS$$

$$\Delta G_{\text{solvation}} = \Delta G_{\text{polar}} + \Delta G_{\text{nonpolar}} = \Delta G_{\text{polar}} + (\gamma SA + b)$$

$$E_{\text{MM}} = E_{\text{internal}} + E_{\text{electrostatic}} + E_{\text{vdW}}$$

$$E_{\text{internal}} = E_{\text{bond}} + E_{\text{angle}} + E_{\text{torsion}}$$

In the above, E_{MM} is the absolute molecular mechanical energy, which can be further divided into E_{bond} , E_{angle} and E_{torsion} , the bond, angle and torsion energies, respectively, and the non-bonded energies $E_{\text{electrostatic}}$ and E_{vdW} . G_{polar} is the polar (electrostatic) contribution to the solvation free energy, and G_{nonpolar} the non-polar component. The latter can be approximated by a term dependent on the surface area SA , $\gamma SA + \beta$, where $\gamma = 0.00542 \text{ kcal.}\text{\AA}^{-2}$ is the microscopic surface tension coefficient and $\beta = 0.92 \text{ kcal.mol}^{-1}$ ¹⁰⁷. γ and β have been determined by fitting the molecular surface area to experimental $G_{\text{hydration}}$ values for small alkanes¹⁰⁸. G_{polar} can be calculated in several ways, but generally a continuum model is used, in which a high dielectric continuum solvent surrounds the solute with a low dielectric. The Poisson-Boltzmann finite-difference method enables the calculation of the reaction field energy of taking the solute from a vacuum to the solvent. Salt effects can also be included in the continuum approximation. T is the temperature and S is the entropy.

This method assumes the additivity of all energy components, an approximation which is alleviated by using solvation models which are compatible with the force field used. This includes using coherent atomic charge and radii values for the calculation of each term, for example. Although the MD simulations are generally carried out in explicit solvent, the ΔG_{polar} term is calculated using a continuum approximation. The large number of water molecules surrounding the systems during simulations would make the calculation of all pairwise electrostatic terms unfeasible.

2.4 Applications of molecular dynamics

MD simulations have been applied to many protein and DNA systems, as well as to G-quadruplex molecules. A variety of areas have been investigated, such as structural features of G-quadruplexes and stability in solution, ion binding and loop conformation.

2.4.1 G-quadruplex structure

The structure and dynamics of several quadruplexes in solution have been studied using MD simulations. Features of the $(G_3T_4G_3)_2$ NMR structure were studied in 1994 using MD¹⁰⁹. Simulations were limited to 150 ps, but were able to show energetic differences in stacking energies depending on the glycosidic angles of the G residues. *Syn-syn*, *anti-anti* and *syn-anti* G-quartet stacks were all equally stable, whereas the *anti-syn* stack was less stable. Early simulations were however limited to short timescales due to computational cost and the instability of DNA structures in longer simulations. Later simulations which used a more appropriate method of accounting for long range electrostatic interactions, such as PME, could provide much more information about the stability of quadruplexes.

MD simulations of both $(G_4)_4$ and $(G_4T_4G_4)_2$ showed that these structures were stable over the 2-3 ns length of the simulations⁸⁸. Some changes were however observed in the $(G_4)_4$ quadruplex, as bifurcated hydrogen bonds were formed in the two central G-quartets, replacing the Hoogsteen hydrogen bonding. This artifact was attributed to the use of a non-polarisable (additive) force field. $(G_4)_4$ was also simulated at 400 K, and remained in the same quadruplex conformation. The loops in the $(G_4T_4G_4)_2$ structure exhibited more flexibility during the simulations.

The stability of quadruplexes with mixed G-C-quartets has also been investigated using MD. Both $(GCGGT_4GCGG)_2$ and $(GGGCT_4GGGC)_2$ yielded stable trajectories, and the mixed quartets did not destabilise the structures⁹⁰. The simulations were able to reproduce two different GCGC quartet geometries depending on coordinating ions, as found in the NMR structures. The T_4 loops adopted different conformations, stable on nanosecond timescales. This suggests that the sampling of different quadruplex loop conformations requires simulations longer than 1-2 ns.

2.4.2 Ion binding

The binding of ions within the central channel of G-quadruplex molecules has elicited much attention due to their structural influence. Different ionic effects have been investigated using MD, however this has often revealed inadequacies in the force fields when dealing with ions. A comparison of quantum mechanical and MD calculations showed that only 60% of the binding energy of Na^+ to G residue O atoms is accounted for in the Amber potential function⁸⁸. The Amber potential is moreover inadequate in that the optimal Na^+ -G residue O6 distance is overestimated by between 0.25 and 0.3 Å, meaning that unlike in experimental structures, Na^+ ions never occupy in-G-quartet-plane positions during MD simulations. In the case of K^+ ions, only 70% of the binding

energy is represented in the MD potential, however in this case the optimal distance between K^+ and O6 corresponds to the QM results. These discrepancies are most likely due to the nature of the additive force field, which neglects any polarisation between atoms. The energetics of cation-G-quartet interactions were also studied using classical force fields¹¹⁰. Although these calculations could explain the binding of K^+ ions between quartet stacks, they could not explain why Na^+ stabilises G-quadruplexes whereas Li^+ does not. K^+ or Na^+ binding within the quadruplex channel was however found to affect the hydrogen bonding distances within G-tetrads, in order to accommodate the different ions⁹².

MD simulations do, however, show that ions are required for the stability of some G-quadruplex structures in solution. $(G_4)_4$ formed structures with slipped strands when no cations were placed within the central channel prior to simulation⁸⁸. On the other hand, $(G_7)_4$ did not undergo such distortions when simulated in solution with no coordinating ions⁹¹. However, Na^+ ions and water molecules did enter the channel to stabilise the structure over the 2 ns simulation. These simulations suggest that G-quadruplexes can be stable even when the central channel is not fully occupied with cations, and that Na^+ ions can diffuse from the surrounding solution into the channel.

Force field calculations were used to locate potential K^+ binding sites in the NMR structure of $G_2T_2G_2TGTG_2T_2G_2$ ^{111,112}. Minimisations of the NMR structures located two potential K^+ ion binding sites, however these have not been confirmed experimentally or with further MD simulations. In fact, a $G_2T_2G_2TGTG_2T_2G_2$ crystal structure in Na^+ containing conditions found no evidence of cation binding within the G-quadruplex³⁷. The conformation of the DNA in this structure is ambiguous, and this could be masking the weak Na^+ ion diffraction, especially if these are located within loop regions¹¹³.

2.4.3 Loop conformation and folding

MD simulations in combination with MM-PBSA free energy analyses can provide information about the preferred conformations of biomolecules. MM-PBSA was shown to be successful in comparing the relative energies of A- and B-form DNA in explicit solvent simulations in 1998¹⁰⁵. Since then, free energy calculations have been applied to the folding of small proteins and protein fragments, and the conformation of protein and RNA loops¹¹⁴. In several cases, correctly folded protein conformations could be identified compared to misfolded conformations^{115–117}. MM-PBSA calculations were used to show whether X-ray or NMR structures had identified the lowest energy conformation of a molecule¹¹⁸. The LES method has also been used when searching for possible loop

conformations in protein¹¹⁹ and RNA¹²⁰ molecules, in order to improve conformational sampling during simulations.

A few studies of G-quadruplex folding and loop conformation have been carried out using the methods described here. Stefl *et al* have investigated the folding pathway of a tetrameric G-quadruplex, $(G_4)_4$ ⁹³. MD simulations show that this sequence can adopt "slipped-strand" structures which are stable over nanosecond timescales. These are however less stable than the native quadruplex structure, as shown by MM-PBSA calculations, and could represent intermediates in the assembly process. The models also suggested that cations are involved in G-quadruplex formation, and contribute strongly to the stabilisation of the structure. Moreover, the free energy calculations provided meaningful results only when the cations within the quadruplex channel were included explicitly, rather than as an implicit salt contribution.

A combination of MD, LES and MM-PBSA was used in a detailed study of $(G_4T_4G_4)_2$ loops⁹⁴. The LES simulations allowed conformational transitions to occur in the loops. However, both LES and MM-PBSA calculations favoured a loop conformation different from the experimentally observed conformation. The mutual agreement of LES and MM-PBSA over loop conformational preference suggests that the inability to predict the experimental conformation is due to the force field. This is also suggested by the fact that the experimental $(G_4T_4G_4)_2$ diagonal loop is unstable in the MD simulations carried out. This study indicates that great care must be taken when trying to predict quadruplex loop conformations using MD.

2.5 Limitations of molecular dynamics

The above examples have already illustrated some of the shortcomings of MD simulations in structural studies. Some of the main limitations and possible ways of circumventing these are summarised here.

MD simulations have to be carried out with a time step of 1 or 2 fs, in order to sample the highest frequency motions, meaning that the length of simulations which are carried out is typically limited to the 10-100 ns range. Most biological processes occur over longer time periods, and hence cannot be simulated fully. Long timescale simulations have been carried out, such as a 1 μ s simulation of the folding of a small protein, the villin headpiece^{115,121}. However, these kind of timescales require large amounts of computational power. Most simulations are still restricted to sampling the folded and unfolded states in different simulations, without observing the transition pathway between the two.

The timescale of the simulations also restricts the extent of conformational sampling which can be carried out. This is a limitation when looking at possible conformations of loop regions, especially with many degrees of freedom. As already mentioned in Section 2.4.3, some methods are available which enable more of the conformational space to be sampled. However, most of the uncertainty in absolute free energies calculated comes from the fact that the whole conformational space of a molecule has not been sampled during the simulations. The energy of the substates which the molecule can occupy contribute to the value of the total free energy, however these substates are generally not all sampled during a simulation. This also affects the calculation of the entropic contribution to the free energy. One of two methods is commonly used to calculate the entropy of a biological system. Both methods compute the atomic fluctuation matrix, either from a normal mode calculation¹⁰⁵, or from snapshots collected over an MD simulation¹²². Each method has its limitations. The former uses a harmonic approximation to the normal modes, and hence the anharmonic contribution is neglected, which can mean that low frequency modes, leading to large movements, are not correctly accounted for. Moreover, snapshots from a simulation have to be minimised before normal mode analysis, which is time consuming. The latter method requires a proper sampling of configurational space, and often suffers from convergence problems, requiring longer simulations to be carried out.

One of the main drawbacks of current molecular dynamics simulations is the inability of the force fields to deal with ionic effects correctly. This is especially the case for simulations of G-quadruplexes, in which cations form an integral part of the system. The experimentally observed influence of cations on structure cannot be reproduced in MD simulations. For example, the $(G_4T_4G_4)_2$ X-ray structure shows K^+ ions binding within the loop regions, as well as in the central G-quadruplex cavity⁵. However, during MD simulations, the K^+ ions do not remain bound to the loops and diffuse into solution⁹⁴. Free energy perturbation calculations were also unable to reproduce the experimentally observed preference for K^+ over Na^+ in G-quadruplex structures¹²³. This is due to electronic effects of the ions which are not accounted for in the simulations. The development of suitable parameters for the treatment of ions has not followed the improvements in the force fields, which means that in many cases they are inadequate⁸². The parameterisation of ions in quadruplex channels was investigated in this work, and is detailed in Chapter 6.

Polarisable force fields are currently being developed which should enable ionic effects to be taken into account. Pairwise potentials represent the atoms as point charges only, which are fixed from the beginning of the simulation. However, if the environment

of an atom changes during the simulation, the atomic charges cannot respond to the change. In the Amber force field, 6-31G* RESP charges are used, which are known to overestimate the dipole moments of monomers. This is considered to account for some of the polarisation effects on the charge distribution.

Despite these limitations, molecular dynamics simulations have been shown in many cases to be useful in studying the structure and dynamics of biomolecules. The Amber force field has been extensively tested on DNA and G-quadruplex systems, as described above. From past performance, the instability of structures in simulations is unlikely to be due to artefacts in the force fields used, but rather represent real structural phenomena. It is however necessary to consider force field effects when looking at certain processes. This is certainly the case when looking at ion binding within G-quadruplex structures.

2.6 Quantum mechanical calculations

Classical MD simulations are unable to represent detailed features of cation interactions within the central channel of G-quadruplexes. Although MD simulations have been used to investigate these (described in Section 2.4.2), this has tended to reveal problems in the force field treatment of the interactions, rather than provide any real answers. On the other hand, quantum mechanical calculations have been able to explain the different quadruplex-stabilising abilities of cations. The drawback of these non-classical calculations is that the computational cost is much higher, and the system size which can be studied is accordingly limited. Currently, only one or two G-quartet stacks are considered in the calculations.

In *ab initio* calculations, the electronic structure of each atom is taken into account, and there are no empirical parameters which might bias the results. The choice of level of theory and basis set governs how the interactions will be calculated, which can influence the results obtained, as will be illustrated below.

The first QM study of G-quartets in 1999 identified a strongly electronegative region in the centre of the quartet¹²⁴. The G-quartet was minimised with no ions bound in the centre, and the structure stabilised with bifurcated hydrogen bonds, in which the amino group is closer to the O6 than N7 atom. However, the existence of bifurcated hydrogen bonds in the absence of a cation is still not clear as their formation depends on the basis set used for the calculation^{125,126}. Minimisations of G-quartets with ions in the centre do not exhibit any bifurcated hydrogen bonds, instead the normal Hoogsteen hydrogen bonds are formed¹²⁷. Several studies have compared the stabilisation properties of various cations binding within the central channel. Gas phase stability calculations

suggested that the stabilisation follows $\text{Li}^+ > \text{Na}^+ > \text{K}^+$. However, when the hydration energy was taken into account, the trend was $\text{K}^+ > \text{Na}^+ > \text{Li}^+$ ¹²⁷. This accounts for the necessary dehydration of the ions before entering the G-quadruplex channel. This trend was later confirmed experimentally using solid-state NMR¹²⁸. Smaller cations were shown to enable shorter hydrogen bonds between the G residues, however this was not a determinant of stability. Instead, the relative energies of hydration of K^+ and Na^+ determined the observed ion selectivity¹²⁹. Minimisations of a G-quartet with a Li^+ ion formed a non-planar complex, which would prevent further stacking of G-quartets, and could explain this ion's destabilising effect on G-quadruplexes¹³⁰.

QM calculations can also identify the most favourable binding sites of various ions within the G-quartets. K^+ ions within the plane of the G-quartets, rather than between successive quartets, lead to less stable structures due to lengthening of the hydrogen bonds^{126,130}. On the other hand, planar and non-planar structures with Na^+ had similar energies. This explains why, experimentally, K^+ ions are always observed between the G-quartet planes, whereas Na^+ ions adopt sequence-dependent positions (see Figure 1.1(b), page 15). Mulliken charges calculated from the minimised structures also show that there is a significant charge transfer from the O6 atoms to the cations in the channel¹³⁰.

QM calculations are useful in describing local phenomena, such as the binding of ions within the quadruplex channel. However, although individual and stacked G-quartets can be minimised, this is usually carried out with the bases only, and no phosphate backbone or sugar. The global effect of ions on quadruplex conformation, for example, cannot be studied using these methods, due to the computational cost. Currently, G-quadruplex structures therefore have to be studied using a combination of methods, depending on the properties of interest.

In this work, MD simulations were used to generate new loop structures, and compare the stability of various quadruplexes in solution (Chapters 3 and 4). Both simulated annealing and LES simulations were employed to search for loop conformations. Ligand binding simulations were carried out after docking of ligands onto the G-quadruplexes, and this will be detailed further in Chapter 5. In order to study the binding of Ca^{2+} ions within the quadruplex channel, a combination of MD and QM calculations was used (Chapter 6).

Chapter 3

$(G_4T_2G_4)_2$, $(G_4T_2AG_4)_2$ and $(G_4T_3G_4)_2$ structures

The length and composition of loops can have an important effect on the structure of G-quadruplexes. In order to gain more understanding of the loop-dependent folding of G-quadruplexes, a series of three G-rich sequences with different loops was investigated. Structural features of the $G_4T_2G_4$, $G_4T_2AG_4$ and $G_4T_3G_4$ sequences were studied using a combination of molecular dynamics simulations and X-ray crystallography. All three sequences are expected to form dimeric G-quadruplexes. The topology of the related $(TG_4T_2G_4T)_2$ quadruplex was solved recently by NMR⁵⁰, and was shown to form two different dimeric structures with lateral loops, each with four G-quartet stacks (see Table 1.3, page 21).

In conjunction with crystallographic trials, the folding of the three sequences into dimeric G-quadruplex structures was investigated using MD simulations. The purpose of the simulations was to determine whether theoretical methods could be used to gain information about G-quadruplex folding. The crystal structures of both $(G_4T_3G_4)_2$ and $(G_4T_2AG_4)_2$ were subsequently obtained, enabling a comparison with theoretical results.

The sequences were chosen for comparison purposes with the already well characterised NMR^{4,52,53} and X-ray^{5,51} structures of $(G_4T_4G_4)_2$. This dimeric G-quadruplex with T_4 loops forms an antiparallel structure with diagonal loops both in solution and in the crystal, and in K^+ and Na^+ containing conditions. No other structural form has been reported for this G-rich sequence. The aim of this study was to establish whether single nucleotide changes in the loop region could have an influence on the G-quadruplex structure, and whether this influence could be predicted.

3.1 Predictive modelling

As previously discussed in Chapter 1, a single G-rich sequence can fold into many different G-quadruplex structures. As current theoretical methods do not enable the modelling of the folding process from an unfolded sequence, simulations must be started from already folded quadruplexes. Possible models of the $(G_4T_3G_4)_2$, $(G_4T_2AG_4)_2$ and $(G_4T_2G_4)_2$ dimeric quadruplexes were constructed, however these were limited to structures containing four G-tetrad stacks, and only non-G residues in the loop regions. Structures with fewer than four G-tetrads and with G residues within the loops can also potentially form, however sampling of all possible structures would be too computationally expensive. Lateral, diagonal and parallel (strand-reversal) loop-type structures were considered.

3.1.1 Computational methods

MD simulations require a starting structure which is not too far removed from the minimum energy conformation. Starting conformations can be obtained either from model building, or derived from NMR or X-ray structures. A mixture of model and experimental templates was used here, in order to avoid bias due to different G-quadruplex stems on the loop conformations. When using experimental templates, only the G-quartets were retained, and loops of the desired length and composition were substituted.

The *Oxytricha* $(G_4T_4G_4)_2$ X-ray structure was used as a template for diagonal loop quadruplexes (PDB code 1JPQ)⁵. Lateral loop templates were generated from a previous (though incorrect) X-ray structure of $(G_4T_4G_4)_2$ (PDB code 1D59)³, and the GCGGT₃GCGG NMR structure (PDB code 1A6H)¹³¹. The latter has two mixed G-C quartets in the centre of the quadruplex. The C residues were replaced with G, and the quartet stem equilibrated for 2 ns to relieve any strain. Both lateral loop templates (1D59 and 1A6H) have alternating *syn-anti* glycosidic angles around the G-quartets, however the 1A6H template has *syn-syn-anti-anti* angles down each G-strand, compared to *syn-anti-syn-anti* angles for the 1D59 template. The 1D59 template was used both as-is, and after a 2 ns MD equilibration, as this incorrect X-ray structure is in a quite high energy conformation⁸⁸. A model-built antiparallel G-quadruplex stem was also included, containing two wide and two narrow grooves, and alternating *syn-anti* glycosidic angles down the G-strands. Parallel structures were generated from the $(TAG_3T_2AG_3T)_2$ X-ray structure (PDB code 1K8P)⁴⁷, in which a fourth G-quartet was added to the stem after removal of the loops and terminal T and A residues. The resulting structure was minimised to relieve any strain in the backbone. The various structural templates which

Table 3.1: T₃, T_{2A} and T₂ loop types which were included in the structure prediction simulations. The *syn* (*s*) and *anti* (*a*) glycosidic angles refer to G residues down the strands.

Loop sequence	Template	Loop type
T ₃	<i>s-s-a-a</i> 1A6H	lateral (wide groove)
	<i>s-a-s-a</i> 1D59	lateral (wide groove)
	<i>s-a-s-a</i> model	lateral (wide groove)
	1JPQ	lateral (narrow groove) diagonal
T _{2A}	<i>s-a-s-a</i> 1D59	lateral (wide groove)
	1JPQ	diagonal
	1K8P	parallel
T ₂	<i>s-a-s-a</i> model	lateral (narrow groove) lateral (wide groove)
	1JPQ	diagonal
	1K8P	parallel

were included in the structure prediction simulations are summarised in Table 3.1. The structures are referred to in the text using the PDB code from which the template was derived, or “model” to indicate the model-built template.

All the initial model building and structural modifications were carried out using the Insight II suite of programs¹⁰⁰. The T₄ loops in the 1JPQ and 1D59 templates were replaced with T₃, T_{2A} or T₂ loops, and the backbones were minimised to relieve strain. The 1A6H template already contains a T₃ loop, which was kept as a starting structure, or modified to T_{2A} or T₂. Loop conformational space was searched with simulated annealing (SA) procedures using the Discover module of Insight II. During all SA runs the G-quartets were kept fixed, and only the loop residues were allowed to move. The SA runs were carried out in implicit solvent, using a distance-dependent dielectric ($\epsilon = 4r$) to mimic the solvent. Two SA protocols were tested, in which the choice of initial structure at each new cycle differed. In the first protocol, the initial loop conformation was minimised, then during each cycle the loop was first heated to 1000 K over 2 ps, simulated at 1000 K for 2 ps, cooled to 300 K over 1 ps and finally minimised. The next loop conformation was generated from heating of the latest minimised conformation. In the second protocol, known as slowly-quenched SA, the starting loop conformations at each cycle were generated from previous structures at

300 K before minimisation. At the end of the run, all the 300 K structures were recalled and minimised. The second procedure should enable a faster search of conformational space, since the starting structure at each step has not been minimised, is in a higher energy state, and hence should be able to overcome higher energy barriers.

The structures obtained from the SA runs were clustered into conformational families, according to root mean square (rms) deviation calculations between all structure pairs. The clustering method is described in Appendix A. Briefly, pairwise rms deviations between all the structures were calculated, then clustered according to the method used by the NMRCLUST program¹³². NMRCLUST was designed to cluster NMR structures, and is therefore not able to handle the large number of structures generated here in each simulated annealing run. A Python script was written which uses the same clustering methodology, but with an unlimited number of input structures. Moreover, it was written to directly read in the output files from the Insight SA runs, and output Insight format archive files for each cluster.

Selected structures from the clusters were subjected to more lengthy MD simulations in explicit solvent using the Amber 7 program¹³³, or Amber 8 as it was released⁸⁶. The G-quadruplex equilibration procedure described here was used throughout this work, unless stated otherwise. 3 K⁺ ions were placed, one between each G-quartet stack, equidistant from the eight G O6 atoms, when these were not already present in the template. Further solution K⁺ ions were added using the Leap module of Amber to neutralise the system, which was then solvated in a pre-equilibrated box of TIP3P water. The box size depended on the system, but always extended at least 10 Å from the solute in every direction. The equilibration procedure consisted of 10 steps, beginning with 1000 steps of minimisation and 25 ps of dynamics of the solvent only. The whole system was then minimised for 1000 steps, followed by 3 ps of dynamics with a restraint of 25 kcal.mol⁻¹ on the DNA. The DNA restraint was lowered by 5 kcal.mol⁻¹ during each of the next 5 1000-step minimisations. Finally, the system was heated slowly to 300 K over 20 ps, with no further restraints. MD simulations were carried out at 300 K, using a 2 fs time step, with SHAKE applied to constrain the bonds containing hydrogen. The PME method was used to deal with long range electrostatic interactions, and Lennard-Jones interactions were cut off at 10 Å. Similar protocols were found to be reliable in previously reported simulations of G-quadruplexes^{93,94}.

LES simulations were carried out on a subset of loop conformations. After an equilibration period of between 500 ps and 1 ns of dynamics in explicit solvent, 5 copies of each loop were generated using the Addles module of Amber. Both the LES (loops) and non-LES regions (G-quartets) were maintained at 300 K, using separate water baths.

LES simulations were carried out in explicit solvent. After LES simulations, the final loop copies were averaged, and non-LES dynamics were carried out for at least 1 ns. The MM-PBSA energies calculated can then be compared to pre-LES energies. LES simulations were also attempted using implicit solvent methods (Generalised-Born(GB)-LES), however this resulted in unrealistic loop conformations, with all T residues completely exposed to the solvent. This is probably due to inadequacies in the solvent approximation by the GB method. Moreover, the GB method required the use of restraints on the G-quartet stem, which would hide any effect of the loops on the integrity of the G-quartets.

The Ptraj module of Amber was used to remove water molecules from the trajectories, overlap the snapshots for visualisation, and calculate average structures. The Carnal module was used to carry out rms deviation calculations. All loop rms deviations were obtained by first fitting the G-quartets over all the trajectory snapshots, and calculating the loop rms deviation with respect to that fit. This enables global loop movements with respect to the quartets to be accounted for, as well as variations within the loops. MM-PBSA was used to calculate approximate free energies, as described in Section 2.3. Snapshots were collected every 20 ps for energetic analysis. The electrostatic contribution to the solvation free energy was calculated using Delphi II¹³⁴. Dielectric constants of 1.0 and 80.0 were assigned to solute and solvent, respectively. A grid spacing of 0.5 Å was chosen, with the longest linear dimension of the molecule occupying 80% of this grid. The Amber parm99 charge set and BONDI radii were used¹³⁵. All MM-PBSA calculations included the three K⁺ ions within the quadruplex channel explicitly. The K⁺ radius was determined to be 2.025 Å, by adjusting it until ($\Delta G_{\text{polar}} + \Delta G_{\text{nonpolar}}$) was close to the experimental $\Delta G_{\text{solvation}}$ of $-80.6 \text{ kcal.mol}^{-1}$ ¹³⁶. The Na⁺ radius in the (G₄T₂AG₄)₂ crystal structure simulations (Section 3.4) was set to 1.656 Å, as determined previously⁹³. All other energy terms were calculated with the programs distributed with Amber. The solute entropic contribution was estimated with the NMODE program, using snapshots collected every 200 ps. Each snapshot was minimised in the gas phase, using a distance-dependent dielectric of $\epsilon = 4r$, before calculation of the vibrational, rotational and translational entropies. The minimisations caused some structural distortion, however this did not have a significant impact on the entropies calculated.

Figures in this thesis were rendered using either Coot¹³⁷ with Raster3D¹³⁸ or VMD¹³⁹ with POV-Ray¹⁴⁰. Diagrams were drawn in Dia¹⁴¹.

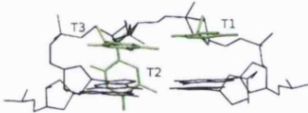
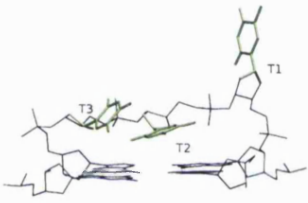
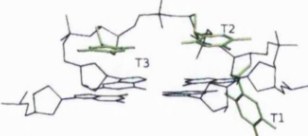
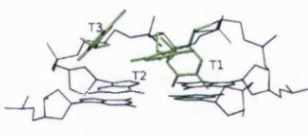
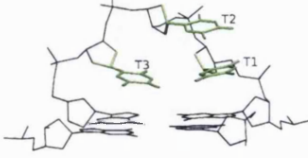
3.1.2 (G₄T₃G₄)₂

Simulated annealing

Lateral and diagonal T₃ loop conformations were generated from simulated annealing runs, using the templates listed in Table 3.1. Lateral T₃ loops bridging two different groove widths were considered. In the 1D59, 1A6H and model-built quadruplexes, the wide grooves measure on average 15.5 Å (C4' to C3' distance across the groove). Lateral loops over the narrow quadruplex groove were also considered using the model-built template, measuring 12.4 Å. The diagonal loops have to bridge an average C4' to C3' distance of 19.5 Å.

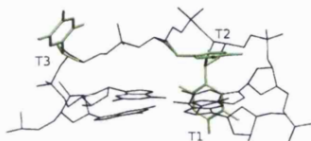
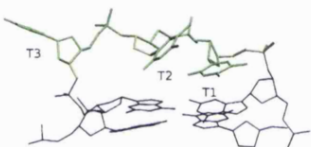
The simulated annealing runs generated large numbers of structures, and only the most frequently repeated conformations were considered. Results from the clustering of lateral T₃ loop conformations over the wide quadruplex groove are summarised in Table 3.2. A large number of loop structures generated were very solvent exposed, with no stabilising interactions between loop and G residues, probably due to the implicit solvent approximation used during the simulations. However, these loops were generally structurally very different from one another, and did not appear in the final clusters created. Clusters containing fewer than 5% of the total number of conformations were not considered. In this manner, the clustering method filters structures which appear very infrequently, and are therefore less likely to represent energy minima. The total number of structures in each simulated annealing run is indicated in Table 3.2, along with the occurrence frequency of each loop conformation. As mentioned in Section 2.2.1, the limited number of SA cycles cannot be guaranteed to find all possible loop conformations. Only a subset of the loop conformational space has been sampled, meaning that the loop occurrence frequencies in Table 3.2 cannot be directly related to the potential energy of the loops. In theory, with a sufficiently large number of SA runs, the number of structures in each conformational cluster should be directly related to the depth of the potential energy wells. This is unlikely to be the case here, however the SA runs did provide a set of possible loop structures. Loop conformations containing stabilising interactions, such as stacking or hydrogen bonding, were found to occur with greater frequencies than loop conformations with no such stabilising interactions.

Table 3.2: T₃ lateral loop structures over the wide groove of antiparallel quadruplexes obtained during simulated annealing runs. The number of times structures occurred during the SA runs is indicated. The first SA protocol was used in all cases, except for one slow-quenched SA protocol (column two of the "model" template). In the structure notation, L denotes a lateral loop, as opposed to D, diagonal or P, parallel. The w specifies that the loop bridges the wide, as opposed to n, narrow groove.

Structure	Model	1A6H	1D59	1D59 (MD)	
T ₃ -L _w -1 	29 (15%)	14 (11%)	30 (10%)	24 (22%)	
T ₃ -L _w -2 	23 (12%)	7 (6%)	21 (17%)	28 (26%)	
T ₃ -L _w -3 	13 (7%)	8 (6%)	7 (6%)	53 (17%)	8 (8%)
T ₃ -L _w -4 	11 (6%)	13 (10%)	19 (24%)		
T ₃ -L _w -5 	9 (5%)		27 (9%)		

Continued overleaf

Table 3.2: *Continued*

Structure	Model	1A6H	1D59	1D59 (MD)
T ₃ -L _w -6		6 (5%)	17 (5%)	
T ₃ -L _w -7			20 (6%)	
All structures	200	128	125	314
				108

Despite large structural variations between the lateral loops shown in Table 3.2, some structural features emerged at this early stage. In all but one conformation (T₃-L_w-1), the second T residue formed some kind of stacking interaction, either with the G-quartets (T₃-L_w-2, T₃-L_w-3, T₃-L_w-4 and T₃-L_w-6), or with another loop residue (T₃-L_w-5 and T₃-L_w-7). The first and third residues were less energetically confined by stacking interactions. Conformations with the second T stacking on the G-quartet were found with the first T either in the groove, at varying depths (T₃-L_w-3, T₃-L_w-4 and T₃-L_w-6), or extended in solution (T₃-L_w-2). The third T either stacked with the G-quartet (T₃-L_w-1, T₃-L_w-3 and T₃-L_w-5) or with the second T (T₃-L_w-2 and T₃-L_w-4), or was extended out into the solvent (T₃-L_w-6 and T₃-L_w-7). Conformation T₃-L_w-3 is similar to one of the 1A6H T₃ loops determined experimentally using NMR, in which the first T residue is in the quadruplex groove, and the other two T residues stack on the G-quartet surface. This provides evidence that such simulated annealing runs can generate experimentally relevant conformations.

The SA simulations generated several potential loop conformations. However it was not possible to classify them in terms of their relative stabilities. Figure 3.1 shows the potential energy for each conformation of the wide groove lateral T₃ loops generated using 1D59 as a template, grouped into clusters. The energies were calculated using the Insight II Discover module. There are large variations in energy within, as well as between, the clusters. Moreover, the most populated clusters do not always contain

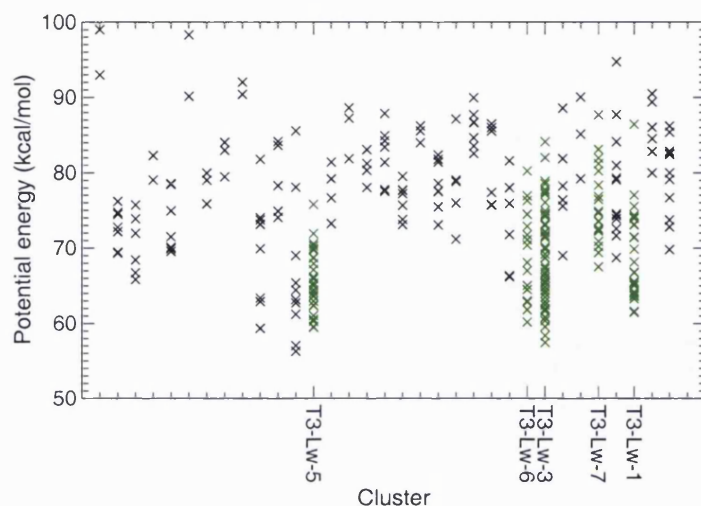


Figure 3.1: Potential energy of all T₃ lateral (wide groove) loops generated using the 1D59 template. For each cluster, the total potential energy of each loop conformation within that cluster is shown. The clusters (loop conformations) which are included in Table 3.2 are highlighted in green.

the lowest potential energy conformations found. This is primarily due to the incomplete sampling of all loop conformations, so that the greater number of structures in a particular cluster does not mean that this conformation is in a deeper potential energy well. The structures in each cluster are similar, as measured by pairwise rms deviations, however there are large energy fluctuations within each cluster, due to local differences in loop conformations. On average, however, the energy of the most populated clusters tends to be lower than the average energy of smaller clusters.

The same simulated annealing protocol was used to generate possible conformations of lateral T₃ loops over the narrow quadruplex groove (Figure 3.2). However, in this case, fewer recurrent conformations were found, and loop structures were less well conserved. The most populated clusters contained only 8 or 9 structures (out of 200 generated), with most conformations having two or more residues exposed to the solvent. This could indicate that the loop residues are able to form fewer stabilising interactions than over the wider groove. Structures similar to the T₃ wide groove loops were found, where the second and third T residues stacked on the G-quartets, and the first T residue was either located in the quadruplex groove or protruding into solution.

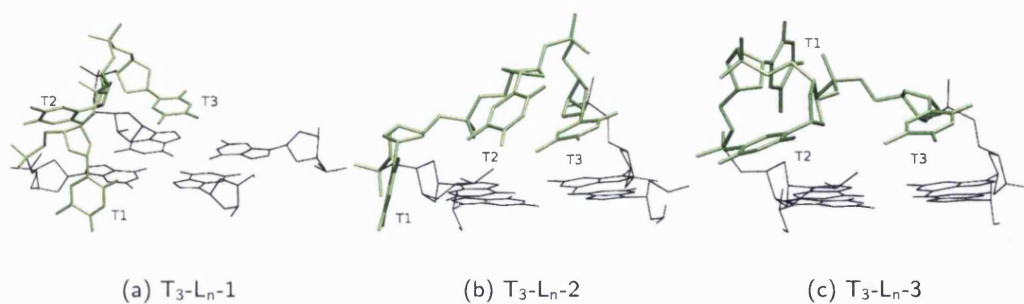


Figure 3.2: Most frequently occurring T_3 lateral loop conformations over the narrow quadruplex groove from simulated annealing runs.

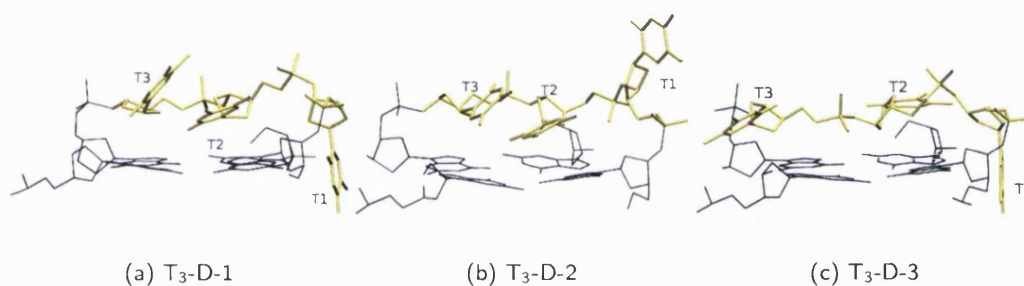


Figure 3.3: Most frequently occurring T_3 diagonal loop conformations from simulated annealing runs.

Diagonal T_3 loop conformations which occurred the most frequently during simulated annealing runs are shown in Figure 3.3. The diagonal loop conformations were similar to the lateral loops, as two residues generally stacked on the G-quartets. As with the lateral loops, the first T residue was most frequently in the quadruplex groove or in solution, but more rarely stacking on the G-quartet surface. The diagonal loop topology allows residues to stack on the G-quartets on either side of the loop backbone, such as in the T_3 -D-3 conformation (Figure 3.3(c)).

Although simulated annealing methods in implicit solvent were useful to generate many possible loop conformations, the stability of the resulting conformations could not be assessed. Even if the number of SA runs was large enough to be able to use the number of times a structure appeared as an accurate indicator of stability, no comparisons can be made between the lateral and diagonal loop types. Total potential energies of each system calculated using Discover are only approximate values, explaining the

spread in energies obtained within each cluster in Figure 3.1. The calculations included solvent effects using a distance-dependent dielectric, which is only a crude approximation. Moreover, only the loop residues were allowed to move during the conformational search, meaning that the G-quartets could not respond to any pressure caused by strained loop conformations. In order to further assess the loop stability, fully solvated MD simulations were carried out.

MD simulations

The T_3 lateral (over the wide and narrow grooves) and diagonal loops were further simulated using molecular dynamics in explicit solvent, accompanied with LES to enhance conformational sampling. With no restraints applied to the G-quadruplex stem, the effect of the loop conformations on the whole system can be studied. Due to the computational cost of fully solvated simulations, only a subset of structures were evaluated.

MD simulations of lateral (wide groove) loops were carried out on the three most frequently occurring loop conformations found for the alternating *syn-anti* model quadruplex, T_3-L_w-1 , T_3-L_w-2 and T_3-L_w-3 (Table 3.2). In the T_3-L_w-1 conformation, the middle T residue is in the quadruplex groove, while the first and third T residues stack on the G-quartet plane. A structure with two loops in the T_3-L_w-1 conformation showed no structural changes after 1 ns MD, apart from a slight movement of the middle T residues out of the quadruplex grooves (Figures 3.4(a) and (c)). A 4 ns LES simulation was carried out starting from the 1 ns equilibrated structure. The two loops behaved differently during the LES simulation. The first loop remained in the same conformation throughout the dynamics, with only some fluctuations in the rms deviation (Figures 3.4(b) and 3.5(a)). After about 2 ns, a structural rearrangement of the second loop occurred, so that the second T residue, previously in the quadruplex groove, moved to stack on top of the other two T residues (Figures 3.4(d) and 3.5(a)). This happened gradually over a few hundred picoseconds, and the second conformation was stable for the remainder of the LES run. The new loop conformation was similar to the T_3-L_w-5 structure in Table 3.2, sampled during simulated annealing. Lowering of the energy barriers enabled the third T residue to flip by 180° in both loops during LES simulations, allowing potential hydrogen bonds to form between the first and third T residues. However, the T residues were only within hydrogen-bonding distance in the second loop, and only after the conformational change to a T_3-L_w-5 type structure. During these simulations, each loop conformation was stable over nanosecond timescales, even when using the LES increased sampling method.

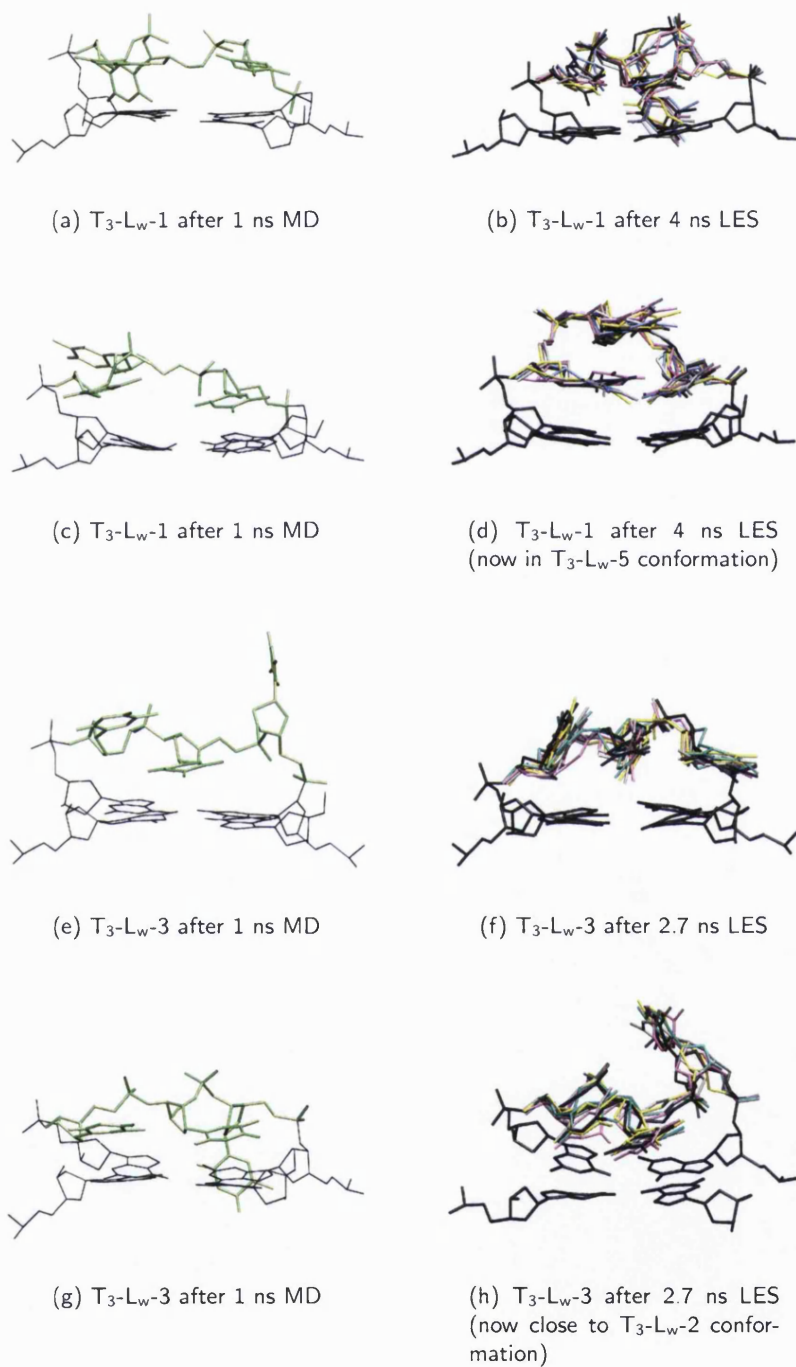


Figure 3.4: Structures of the T_3-L_w-1 (a)-(d) and T_3-L_w-3 (e)-(h) loop conformations after 1 ns MD and after LES simulations. The overlap of all 5 loop copies at the end of the LES simulations is shown.

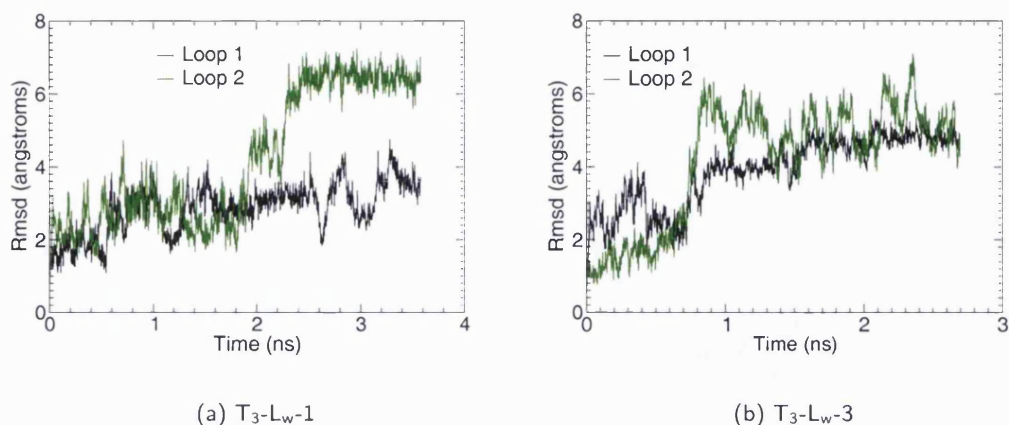


Figure 3.5: Loop rms deviations during the LES simulations of T_3-L_w-1 and T_3-L_w-3 conformations. Only the rms deviations of single LES copies are shown, as all LES copies showed the same behaviour.

The SA runs generated several structures in which the first T residue was positioned in the quadruplex groove (T_3-L_w-3 , T_3-L_w-4 and T_3-L_w-6). A 1 ns MD simulation of the T_3-L_w-3 loop types suggested that the latter were less stable than the previously sampled T_3-L_w-1 and T_3-L_w-5 conformations. The T_3-L_w-3 loop residues were very flexible, especially the first and third T residues in the first loop (Figure 3.4(e)). The third T residue spent most of the simulation protruding into the solvent, and only returned to stack with the middle T residue at the end of the dynamics. The first T residue moved gradually out of the quadruplex groove in the first hundreds of picoseconds, and then remained exposed to the solvent. However, the second loop in the same conformation was stable over the simulation, the only rearrangement being the stacking of the third T with the G-quartet, rather than on top of the middle T (Figure 3.4(g)). During the subsequent 2.7 ns LES simulation, both loops rearranged to the conformations shown in Figures 3.4(f) and (h). Stacking of the second and third T residues was conserved during the LES simulation, however the first T residue was much more flexible. The final structures obtained are similar to the T_3-L_w-2 conformation from the SA runs, in which the first T residue moves closer to stack with the other loop residues, thus minimising solvent exposure. The loop flexibility is apparent in Figure 3.5(b). The first loop rms deviation increased throughout the simulation, and never equilibrated into a specific conformation. This was also the case for the second loop, which showed

large fluctuations in rms deviation throughout the LES dynamics, after a conformational change around 800 ps. This differs from the second loop in Figure 3.5(a), where a major conformational change caused the rms deviation to increase sharply, but a stable conformation was reached and little variation in rms deviation was subsequently observed.

The native 1A6H T₃ loops were simulated for 4 ns, after mutation of the C residues to G in the quadruplex stem, in order to allow energetic comparisons with predicted loops. Multiple loop conformations are present in the NMR structure, and the first PDB entry was chosen for MD. In this conformation, the first T residue is in the quadruplex groove, the third T stacks on the G-quartets, and the second T is pointing into the solvent. Within the first 400 ps of dynamics, the second T residues of both loops formed stacking interactions with either the G-quartets, or the second T residue, and these were stable throughout the simulation. This is in accordance with the NMR structures, some entries having both second and third T residues stacking with the G-quartets. The first T residue remained within the quadruplex groove in only one of the loops. In the second loop, this T residue moved into the solution, corresponding to the second NMR structure in the 1A6H PDB entry. This simulation revealed that a loop with the first T residue in the quadruplex groove can be stable, although rearrangements to other conformations can occur (in the 1A6H simulation as well as the T₃-L_w-3 simulation described above). The T₃ loops were flexible in the NMR structure, and interchange between the different experimental conformations was observed during the simulations. The final loop conformations obtained were similar to the T₃-L_w-3 and T₃-L_w-4 structures in Table 3.2. The loop flexibility within the NMR structure suggests that the conformational changes which occur during the simulations are due to real structural flexibility of the loops, rather than force field effects.

Lateral loops over the narrow quadruplex groove in the T₃-L_n-2 conformation were simulated using MD. Over a 1 ns equilibration, the loop conformation changed little, with two residues forming a stack over the G-quartets, and the first T residue moving slightly out of the quadruplex groove (Figure 3.6(a)). Over a 4 ns LES simulation, the loops were more flexible, although no major rearrangements were observed (Figure 3.6(b)). The individual loop copy rms deviations were around 3.0 Å, after fitting of the G-quartets, however the final structure was very similar to the starting T₃-L_n-2 conformation. As with several other simulations, the first T residue, originally within the quadruplex groove, adopted a slightly higher position, in order to form interactions with the other two loop residues.

A quadruplex with two diagonal T₃ loops in the T₃-D-1 conformation was subjected to a 4 ns MD simulation. The two loops behaved differently during the dynamics. One

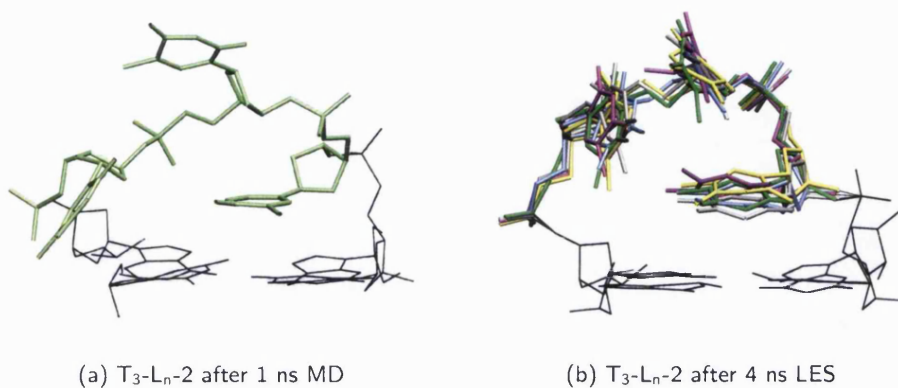


Figure 3.6: MD simulation of the T_3-L_n-2 loop conformation. (a) 1 ns equilibrated structure and (b) overlap of the 5 loop copies at the end of the 4 ns LES simulation.

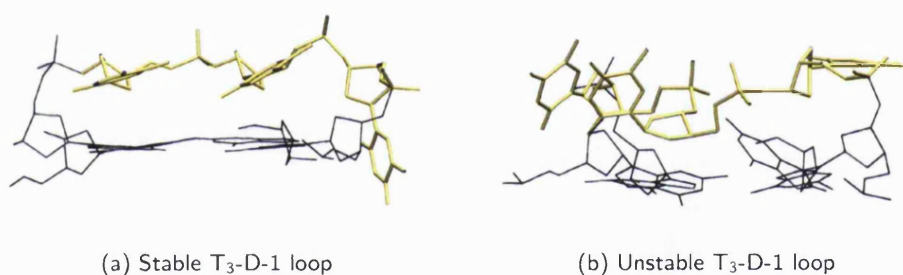


Figure 3.7: MD simulation of the T_3-D-1 loop structure. (a) Stable loop structure averaged over the final 2 to 4 ns MD and (b) final structure of the unstable loop occurring within the last 200 ps.

of the loops remained in the T_3-D-1 conformation, with the first T residue deep within the quadruplex groove. On the other hand, the first T residue of the second loop moved in and out of the quadruplex groove during the dynamics. The latter loop also caused a G-quartet to become distorted during the final few hundred picoseconds of simulation, when a G residue left the quartet plane, and formed a hydrogen bond with the loop backbone. Both loops are shown in Figure 3.7. This shows the importance of carrying out lengthy simulations, as the loop conformation was stable for almost 4 ns before causing G-quartet distortion. This also shows how simulations can be sensitive to small changes in loop conformation, as the first loop in essentially the same conformation was stable during the dynamics.

Many loop conformations were stable over nanosecond timescales, however interchange between different loop conformations was frequently observed. Similar loop conformations exhibited different behaviours within the same simulation, suggesting that several T_3 loop conformations are equally possible. Unstable loops did not generally affect the stability of the G-quadruplex stem, which always had a lower rms deviation. These results are encouraging in that there was generally a good agreement between the conformations found in the SA runs and during the dynamics. Conformations adopted during the MD simulations were structurally close to structures generated in the simulated annealing runs (for example in Figures 3.4(d) and (h)). The diagonal loop simulation outlined some of the main difficulties when using theoretical models as starting structures in MD simulations. Small differences in loop conformation can lead to major structural changes over long timescales, and it can be difficult to establish whether these differences are due to the general loop conformation or to the particular starting structure which was used. Both simulated lateral T_3 - L_w -3 loops suggested that conformations with the first T residue located within the quadruplex groove were not the most stable, as rearrangements occurred. However, this conformation was stable during a simulation of the 1A6H T_3 loops. The results from MD simulations are therefore highly dependent on the initial structure.

Comparison of diagonal and lateral loops

Using simulated annealing and molecular dynamics, stable conformations of both lateral and diagonal T_3 loops were found. Absolute free energy calculations using MM-PBSA were carried out in an attempt to energetically rank the different structures, and are shown in Table 3.3. The total free energy G_{total} was calculated for the structures simulated in explicit solvent. G_{total} was further subdivided into contributions from each of the loops, following the method of Fadrna *et al*⁹⁴. G_{stem} , $G_{stem+loop1}$ and $G_{stem+loop2}$ were each calculated, and G_{loop1} and G_{loop2} were obtained from $G_{stem+loop1} - G_{stem}$ and $G_{stem+loop2} - G_{stem}$, respectively. In this manner, the contribution of the loop-G-quartet interaction is included in the loop free energies.

Table 3.3 shows that all the loop conformations are close in energy. This was also observed during the dynamics as one loop type could rearrange to another, although both were stable for several nanoseconds. The LES simulations enabled more structural flexibility in the loops, and free energies after the LES simulations are generally lower than before the enhanced sampling. This suggests that during the LES simulations, more favourable loop conformations were formed. The T_3 - L_w -5 conformation was the most favourable T_3 loop obtained, with a free energy of $-461 \text{ kcal.mol}^{-1}$ (Table 3.3).

Table 3.3: Free energy values (kcal.mol⁻¹) for selected T₃ loop conformations, calculated over MD simulations before and after LES simulations (apart from 1A6H and T₃-D-1, for which LES simulations were not carried out). See text for a description of the simulations. Standard errors of the mean are in brackets.

Loop conformation	G _{total} [*]	G _{stem}	G _{loop1}	G _{loop2}	-TS _{total}
T ₃ -L _w -1 (before LES)	-4526 (4)	-3628 (4)	-449 (4)	-449 (4)	-585 (2)
T ₃ -L _w -1 (after LES)	-4545 (4)	-3632 (4)	-451 (4)		-581 (1)
T ₃ -L _w -5 (from T ₃ -L _w -1 LES)				-461 (4)	
T ₃ -L _w -3 (from 1A6H MD) [†]	-4522 (5)	-3620 (4)	-455 (4)		-582 (1)
T ₃ -L _w -4 (from 1A6H MD) [†]				-447 (4)	
T ₃ -L _w -3 (before LES)	-4517 (5)	-3626 (4)	-442 (4)	-449 (4)	-580 (2)
T ₃ -L _w -3 (after LES)	-4543 (5)	-3636 (4)	-456 (4)	-451 (4)	-586 (1)
T ₃ -L _n -2 (before LES)	-4525 (4)	-3633 (4)	-446 (4)	-446 (4)	-582 (1)
T ₃ -L _n -2 (after LES)	-4532 (4)	-3634 (4)	-451 (4)	-448 (4)	-582 (1)
T ₃ -D-1	-4525 (5)	-3623 (4)	-452 (4)	-451 (4)	-579 (1)

^{*}G values do not include the solute entropy

[†]Values calculated after a 2 ns equilibration

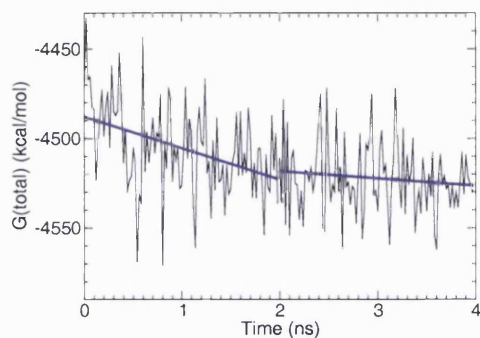
Unfortunately, the G_{loop} values are not significantly different enough to be used to compare the stabilities of various loop conformations. They are dependent on the particular loop conformation in each simulation, rather than on general structural features. Lateral T_3 loops over the narrow and wide groove of the quadruplex have similar free energies, suggesting that both topologies are equally possible. Moreover, the free energies of diagonal and lateral loops were also similar, which does not allow differentiation between the two. The free energy decomposition into loop and quadruplex stem contributions is an approximation, which adds further uncertainty to the energies calculated. However, as both loops often showed different behaviours during the simulations, the total free energy of the quadruplexes does not reflect the effect of each loop conformation.

The solute entropic contribution was not included in the G values in Table 3.3, although the solvent entropy is implicitly taken into account in the solvation energy term. The entropic contribution is the least reliably calculated component of the free energy, and was therefore included separately. The entropy was calculated for the whole quadruplexes, and was similar for the different quadruplexes considered. Some measure of the entropic component is, however, important, as its inclusion can alter the ranking of quadruplex energies.

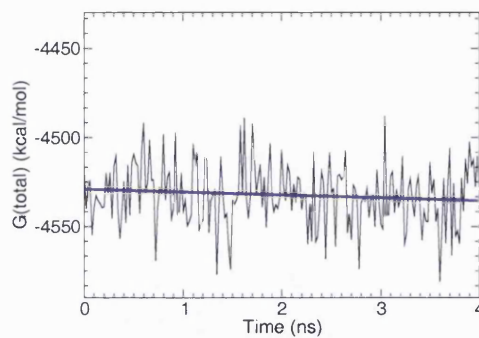
Free energy calculations can also provide information about the state of equilibration during trajectories. In most cases, plots of G_{total} over the length of the trajectories show that the structures did equilibrate over the timescale of the simulations. Figure 3.8 shows the G_{total} variation over the simulations of the 1A6H template with native T_3 loops (Figure 3.8(a)), and the model template with T_3-L_n-2 lateral loops after LES (Figure 3.8(b)). The latter structure is largely equilibrated, as indicated by the linear regression to the data. On the other hand, the first 2 ns of the 1A6H simulation show a strong decrease in the free energy as the structure equilibrated, followed by a stabilisation after 2 ns. This was mostly due to the backbone rearrangements necessary after modification of the G-C mixed quartets to G-quartets. Free energy plots are thus a good indicator of how long a simulation should be carried out for.

3.1.3 $(G_4T_2AG_4)_2$

The procedure followed to predict favourable conformations of T_2A loops was essentially the same as that described for the T_3 loops. Figure 3.9 shows the most frequently occurring T_2A loop conformations obtained during SA simulations. The T_2A loop structures were similar to the T_3 loop conformations. The replacement of T by an A residue does not appear to cause any changes in the possible loop conformations. A-T hydrogen bonds were observed in some of the loop conformations, however these were observed



(a) 1A6H template with native T_3 loops



(b) T_3 - L_{n-2} lateral loop (narrow groove), *sasa* model template after LES simulation

Figure 3.8: Variation of G_{total} over the trajectories of selected simulations. The purple line show the linear regression over (a) 0 to 2 ns and 2 to 4 ns and (b) 0 to 4 ns.

between co-planar T residues in the T_3 loops as well. MD simulations were carried out on a subset of these loop conformations, and confirmed that both lateral and diagonal loops were stable in solution during the dynamics.

Simulated annealing runs were also carried out using the 1K8P parallel quadruplex template, with strand-reversal loops. Results were less conclusive than for the antiparallel quadruplexes. The SA simulations were unable to identify any conformations which were particularly favourable compared to others. Instead, a wide range of different conformations were predicted, with average rms deviations of 5 Å between pairs of conformations. This is probably due to the nature of the interactions between loop residues and G-quartets. In the antiparallel quadruplexes, with either lateral or diagonal loops, the most favourable conformations generally include stacking interactions between loop residues and the G-quartets. In the parallel quadruplex structures, the loop residues can stack with each other, or interact with the quadruplex groove. However, these interactions appear to be much less specific, as particular loop conformations did not occur repeatedly. The results indicate that parallel strand-reversal loops might be more flexible in solution than antiparallel loops. This was illustrated during MD simulations, where some of the parallel T_2A loop residues had rms deviations of greater than 6 Å, after fitting of the G-quartets.

T_2A loops have been observed experimentally as lateral, diagonal⁶² and parallel

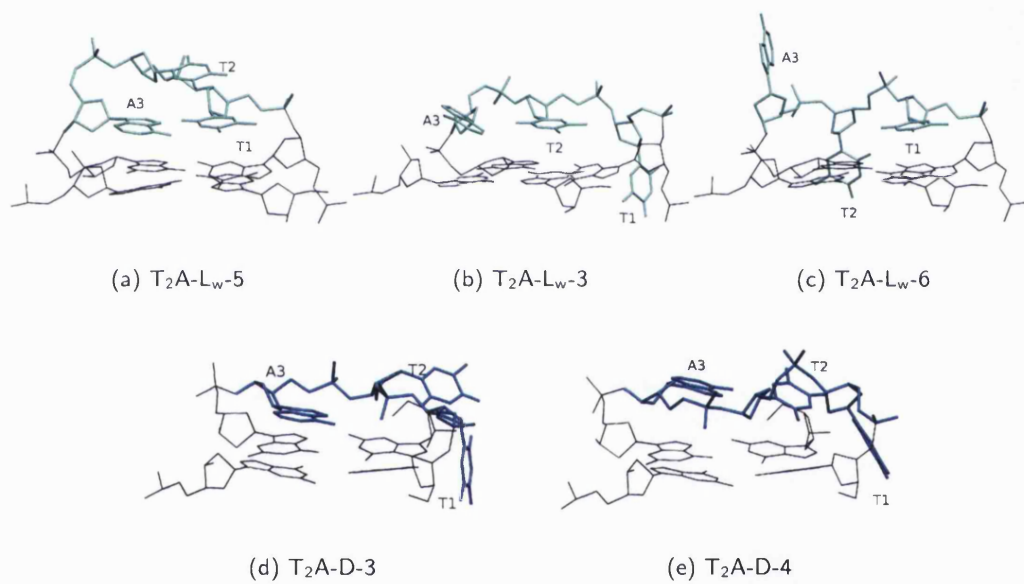


Figure 3.9: Most frequently occurring T_2A lateral and diagonal loop conformations from simulated annealing runs. The numbering of each conformational type is the same as in Table 3.2 and Figure 3.3 for the lateral and diagonal loops, respectively.

loops⁴⁷, in structures of the human telomeric sequence containing three, rather than four, G-stacks. The simulations show that this can probably also be the case for the G₄T₂AG₄ sequence. The occurrence of both lateral and diagonal loops within the quadruplex structure of AG₃T₂AG₃T₂AG₃T₂AG₃ indicates that one loop type is unlikely to be much more favourable than the other⁶².

3.1.4 (G₄T₂G₄)₂

Simulated annealing

Different results were obtained for the T₂ loop simulations compared to the three-nucleotide loops. Simulated annealing runs generated several possible lateral, diagonal and parallel loop conformations, which are shown in Figure 3.10. As only the loop residues were allowed to move during the conformational search, no effect of the short loop length could be observed on the G-quartets. However, the loops themselves were under some strain in the diagonal and lateral loops over the wider groove. This was manifested through increased bond lengths, for example P-O3' or P-O5' distances of 1.7 Å rather than around 1.6 Å. T₂ loops were much less conformationally flexible than T₃ and T₂A loops. Both diagonal and lateral T₂ loops over the wide quadruplex groove formed only four different conformations during 200-step SA runs. The same lateral loop conformations were found with both the 1A6H and 1D59 templates, and two examples are shown in Figures 3.10(a) and (b). Diagonal T₂ loops are shown in Figures 3.10(f) and (g). The limited number of possible conformations, and the reduced flexibility, suggests that the loop backbones are under strain. Lateral T₂ loops over the narrow groove were much more flexible, and many possible conformations were obtained, the most frequently repeated being shown in Figures 3.10(c)-(e). In the latter case, the flexibility of the backbone suggests that the structures were not under strain.

Parallel T₂ loops were quite flexible, although most structures obtained during the SA runs were of the type shown on the left in Figure 3.10(h), in which the two residues are located in solution, on either side of the backbone. Only very few structures obtained showed any interactions between the loops, such as the stacking shown in Figure 3.10(h), on the right.

The simulated annealing runs suggest that T₂ loops bridging distances of 15 Å and above are under strain. However, the SA protocol did not allow the G-quartets to respond to the strain in the loops. Longer MD simulations with no restraints on the G-quartets were carried out in order to assess whether local changes in the G-quartets could relieve some of the strain caused by the loops.

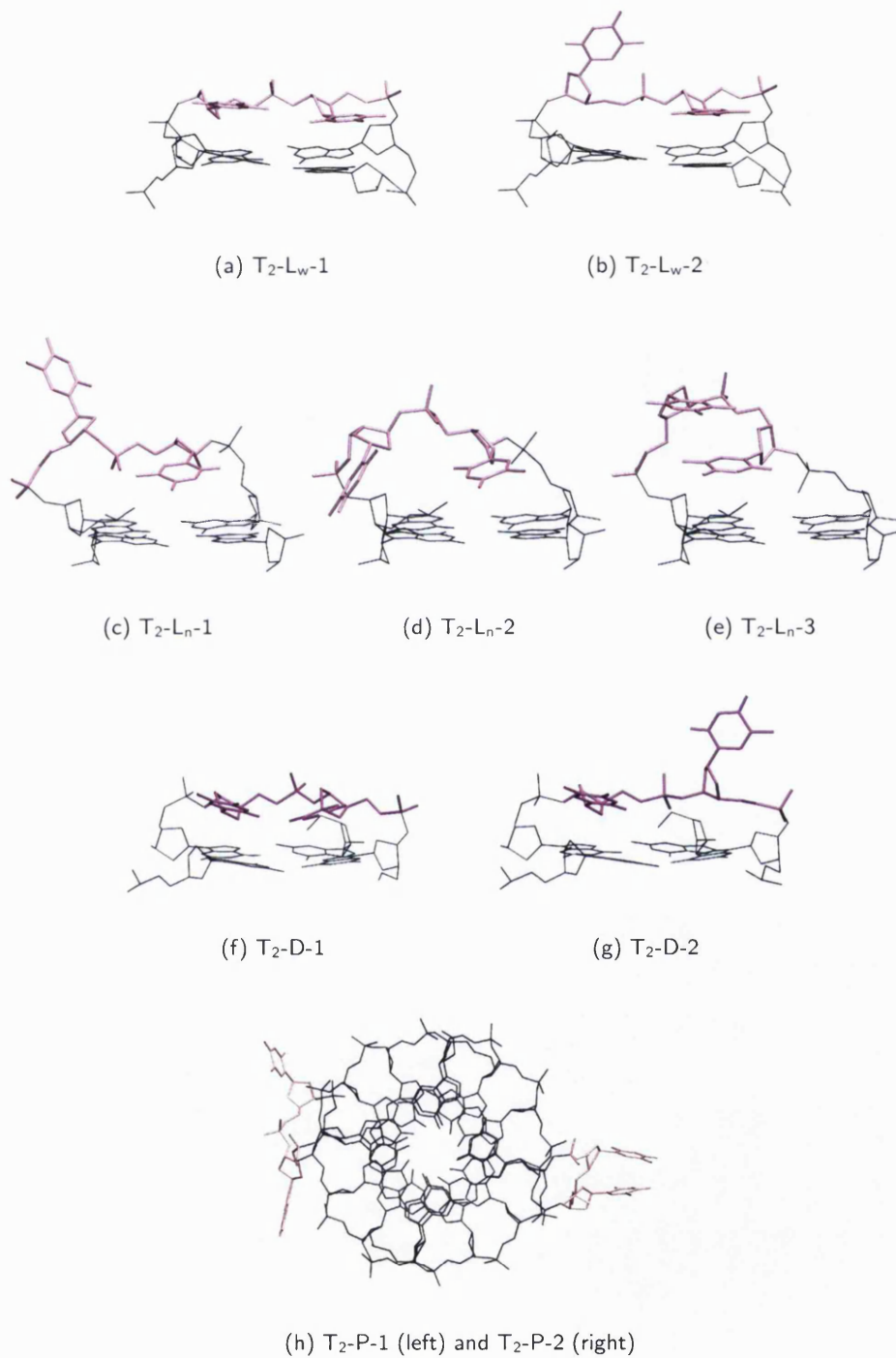


Figure 3.10: Possible T_2 loop conformations obtained during simulated annealing runs. The loop notation is the same as in Table 3.2. (a, b) lateral loops over the wide groove, (c, d, e) lateral loops over the narrow groove, (f, g) diagonal loops and (h) parallel loops (top view).

MD simulations

The diagonal-loop T₂-D-1 quadruplex was unstable during the simulation. After only 400 ps of dynamics, the G-quartets directly below the T₂ loops were severely distorted, as shown in Figure 3.11(b). The strain caused by the short loop was apparent from the very beginning of the simulation, as the upper quartet G residues were forced closer together, making them tilt inside the quadruplex, and causing a K⁺ ion to move out of the channel and into the loop region (Figure 3.11(b)). These simulations indicate that diagonal T₂ loops are unlikely to form in solution.

The same straining effect was observed for lateral T₂ loops over the wide quadruplex groove, although this was less pronounced than for the diagonal loops. Figures 3.11(c) and (d) show a selected lateral T₂ loop over the wider quadruplex groove at the beginning and after 4 ns of dynamics. Over the course of the simulation, the G-quartets became distorted, and Hoogsteen hydrogen bonds were not maintained between all the G residues. However, the three K⁺ ions remained within the central channel, and stacking between the G residues was mostly conserved. The lateral (wide groove) T₂ loop is less strained than the diagonal loop, however neither appears particularly favourable.

On the other hand, T₂ lateral loops bridging the narrow quadruplex groove caused no distortion of the G-quartets (Figure 3.11(f)). A parallel T₂ loop quadruplex was also stable over a 4 ns simulation (Figure 3.11(h)). A K⁺ ion moved out of the quadruplex channel, but remained at the channel exit for the duration of the dynamics. This did not cause any distortion to the G-quartets, or changes in the Hoogsteen hydrogen bonding.

MM-PBSA calculations were carried out to attempt to differentiate between the stable T₂ lateral and parallel loop quadruplexes which were obtained. Decomposition of the free energy into loop and stem components, as used above, is inappropriate in this case, due to the completely different manner in which lateral and parallel loops interact with the G-quartets. Antiparallel quadruplex loops (lateral or diagonal) interact primarily with the G-quartet surface. On the other hand, parallel quadruplex loops interact with the groove region of the G-quartets. Only overall free energies of parallel and antiparallel quadruplexes can therefore be compared. The lateral (narrow groove) quadruplex was more favourable than the parallel quadruplex, with $G_{\text{total}} = -4237$ and -4222 kcal.mol⁻¹, respectively. This was still the case when the solute entropy was taken into account, as $TS = 539$ and 543 kcal.mol⁻¹, respectively. For the particular loop conformations which were simulated, the antiparallel structure was more energetically favourable than the parallel structure. However, this does not imply that a more favourable parallel loop structure could not be found.

The T₂ loop results were in part confirmed by the publication of the topology of

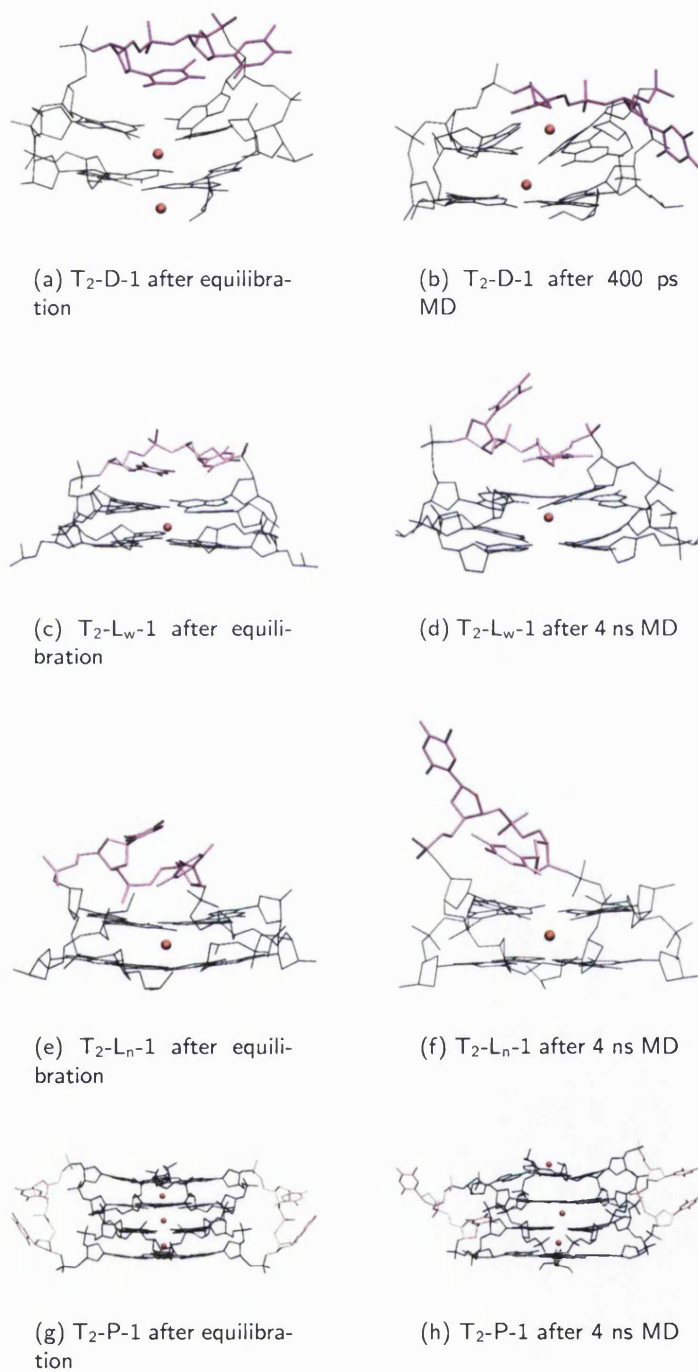


Figure 3.11: T₂ loop structures before and after MD simulations in explicit solvent. For the antiparallel quadruplexes, one loop is shown with two G-quartets and channel K⁺ ions in red. All residues and channel K⁺ ions are shown for the parallel quadruplexes.

the dimeric $(TG_4T_2G_4T)_2$ quadruplex⁵⁰. This sequence contains two extra terminal T residues which were not modelled here, but was shown to form two antiparallel dimers, both with lateral loops. The full structure and loop conformations were not, however, determined, and cannot be compared to the conformations suggested by the modelling studies.

3.1.5 Structure prediction

Initial simulated annealing runs enabled potential loop conformations to be identified. Loops in which hydrogen bonding or stacking interactions were formed were favoured over conformations which lacked such interactions. Simulated annealing generated several possible lateral, diagonal and parallel loops for each loop length. MD simulations of several T_3 and T_2A loops were, however, unable to distinguish between different conformations. Loops were very flexible and often changed conformations during the simulations. Free energy calculations carried out for the T_3 loop structures suggested that the most favourable loop conformation was T_3-L_w-5 (Table 3.3). In this conformation, all three residues stack on top of the G-quartets (Figure 3.4(d)), and this was more energetically favourable than any other lateral or diagonal T_3 loop. Free energy differences between loop types were however small, suggesting that even lateral and diagonal loop conformations might be similar in energy. Loop conformational preferences could be due to small differences in free energy which cannot be measured reliably using approximate methods such as MM-PBSA.

T_2 loop results were more conclusive, due to the strain caused by several of the loop types simulated. MD simulations showed that T_2 loops are unlikely to form diagonal or lateral loops over the wide quadruplex grooves. Both these loops caused distortion to the G-quartets, as shown in Figure 3.11. The instability of the G-quadruplexes was due to strain in the loop backbone, and is thus unlikely to be due only to force field artefacts. Free energy calculations suggested that lateral T_2 loops over the narrow quadruplex groove are energetically favoured over parallel loops.

3.2 Crystal structures of $G_4T_3G_4$ and $G_4T_2AG_4$

3.2.1 Materials and methods

Native and 5-^{Br}dU DNA oligonucleotides for crystallisation purposes were purchased from Eurogentec and used without further purification. $G_4T_3G_4$ and $G_4^{Br}UT_2G_4$ were annealed at 85°C for 5 minutes in solutions containing 2 mM DNA, 80 mM KCl and 50

mM K^+ cacodylate buffer, pH 7.0, then left to cool to room temperature. $G_4T_2AG_4$ and $G_4^{Br}UTAG_4$ were annealed in solutions containing 3 mM DNA and 40 mM Na^+ cacodylate buffer, pH 7.0, in the same manner. Initial crystallisation trials, both with Hampton Research and in-house developed screens, were set up as hanging-drops. After initial crystals were obtained, the conditions were refined using under oil experiments. For the latter, 96-well plates were used, and 4 μ L drops were covered with between 20 and 40 μ L of oil. Mixtures of paraffin and silicon oils were varied in order to control the evaporation rate of the drops. The crystallisation conditions are summarised in Table 3.4.

Crystals of $G_4T_3G_4$ from the original screen drop diffracted to 1.5 Å resolution, and diffraction data were collected on beamline ID14-4 at the ESRF, Grenoble. Despite high resolution data, molecular replacement with various models did not yield any structure solution. Crystals of $G_4^{Br}UT_2G_4$ were subsequently grown under oil. Crystals in space group C2 appeared within a few days, however they diffracted poorly, and single crystals were difficult to obtain. Data were collected on an in-house R-axis IV detector. After about 1 month, smaller crystals in space group P2₁ appeared in drops containing a higher ratio of paraffin to silicon oil (corresponding to slower evaporation). These crystals diffracted to 2.8 Å and data were collected on beamline ID14-4 at the ESRF, Grenoble. MAD data were collected at four wavelengths for Br heavy-atom phasing purposes. The peak (0.9199 Å), inflexion point (0.9201 Å) and remote wavelengths (high remote 0.9165 Å and low remote 0.9235 Å) were determined by carrying out a fluorescence scan.

Initially, the heavy atom search was carried out using all four wavelengths collected for the $G_4^{Br}UT_2G_4$ P2₁ crystal. However, due to poor merging statistics between the four wavelengths, an interpretable heavy-atom phased map was obtained with SHELXD/E¹⁴² only when using the peak data set on its own. This electron-density map was of poor quality, however two rows of three and seven discrete spherical density peaks were seen and interpreted as K^+ ions within the quadruplex central channel. The three quadruplexes in the asymmetric unit were positioned around the K^+ ions, and $^{Br}UT_2$ loops were built into the remaining density. The poor quality of the initial map, despite the strong anomalous signal, was explained by subsequent calculation of the heavy-atom Patterson map from the refined Br atom positions in the structure. Several Br peaks in the Patterson map overlap onto a single peak in the Harker Section, creating a very large peak near the origin. The various phasing programs used could not separate out these peaks, and the initial heavy atom phased map was calculated from other heavy atom positions. Model building and density visualisation were carried out using Coot¹³⁷,

Table 3.4: Crystallisation conditions.

Sequence	Drop solution	Well solution	Method	Space group
$G_4T_3G_4$	2 μ L of 1 mM DNA 2 μ L of well solution	0.3 M NaI 0.05 M Na ⁺ cacodylate pH 6.5 15% PEG 400)	hanging drop	C222
$G_4^{Br}UT_2G_4$	2 μ L of 1 mM DNA 2 μ L of (0.3 M NaI 0.05 M Na ⁺ cacodylate, pH 6.5 15% PEG 400)	NA*	under oil 20 μ L of 1:2 paraffin: silicon	C2
$G_4^{Br}UT_2G_4$	2 μ L of 1 mM DNA 2 μ L of (0.3 M NaI 0.05 M Na ⁺ cacodylate, pH 6.5 15% PEG 400)	NA	under oil 20 μ L of 2:1 paraffin: silicon	P2 ₁
$G_4^{Br}UTAG_4$	2 μ L of 1 mM DNA 2 μ L of (0.05 M NaCl 0.05 M CaCl ₂ 0.1 M Na ⁺ Hepes, pH 7.0 30% PEG 400)	NA	under oil 30 μ L of 1:1 paraffin: silicon	C222

*Not applicable

and the refinement was carried out using Refmac5¹⁴³ within the CCP4 package. The final R factor and R_{free} were 16.8 and 24.2% respectively, and the structure coordinates have been deposited in the Protein Data Bank with the code 2AVJ.

The $G_4^{\text{Br}}\text{UT}_2\text{G}_4$ $P2_1$ solution was used as a model for molecular replacement of the C2 $G_4^{\text{Br}}\text{UT}_2\text{G}_4$, $G_4\text{T}_3\text{G}_4$ and $G_4^{\text{Br}}\text{UTAG}_4$ structures. The C2 $G_4^{\text{Br}}\text{UT}_2\text{G}_4$ structure contains a single quadruplex molecule in the asymmetric unit, and was solved using two strands of a $P2_1$ quadruplex. The EP MR¹⁴⁴ program found the correctly packed solution, despite this having a low correlation coefficient of 20% and a high R factor of 84%. The solution was only partially refined using Refmac5 to an R factor of 25%, and an R_{free} of 36%, due to the low quality of the data collected.

The $G_4\text{T}_3\text{G}_4$ structure contains a single strand (i.e. half a quadruplex molecule) in the asymmetric unit, and was solved by molecular replacement using a single strand from the $P2_1$ solution structure using EP MR. A correctly packed solution was found with a correlation coefficient of 28% and an R factor of 58%. This solution was refined using Refmac5 to a final R factor and R_{free} of 19.8 and 20.8% respectively. The structure factor data and structure coordinates have been deposited in the Protein Data Bank, code 2AVH.

Diffraction data from $G_4^{\text{Br}}\text{UTAG}_4$ C222 crystals were collected on beamline 10.1, at SRS, UK. The $G_4^{\text{Br}}\text{UTAG}_4$ structure contains a single chain in the asymmetric unit and was solved by molecular replacement using a single strand from the $P2_1$ solution. The first EP MR solution had a correlation coefficient of 79%, and an R factor of 41%. Structure refinement was carried out with CNS¹⁴⁵, after mutation of a T to an A residue. Due to the small number of reflections which were collected, the final R factor and R_{free} were only refined to 25.0 and 28.6%, respectively.

Crystallographic data for all the structures is shown in Table 3.5.

3.2.2 Crystal packing of the $(G_4^{\text{Br}}\text{UT}_2\text{G}_4)_2$ $P2_1$ structure

The $(G_4^{\text{Br}}\text{UT}_2\text{G}_4)_2$ structure was solved in space group $P2_1$. The $P2_1$ asymmetric unit contains three crystallographically and structurally distinct quadruplex molecules. Figure 3.12 shows the contents of the asymmetric unit. Two strands form a head-to-tail dimer, with lateral loops at either end of the quadruplex (Figure 3.12, on the left). This dimer is very similar to the asymmetric unit in the $(G_4^{\text{Br}}\text{UT}_2\text{G}_4)_2$ C2 structure. The $P2_1$ asymmetric unit contains two other G-quadruplex molecules, both head-to-head dimers, with lateral loops on adjacent sides of the G-quartet stem (Figure 3.12, on the right). The two dimers stack end-to-end via their free G-quartet face, forming an eight G-quartet stack. Seven K^+ ions are equally spaced between the G-quartets, including

Table 3.5: Crystallographic data.

Sequence	G ₄ ^{Br} UT ₂ G ₄	G ₄ ^{Br} UT ₂ G ₄	G ₄ T ₃ G ₄	G ₄ ^{Br} UTAG ₄
Data collection				
Space group	P2 ₁	C2	C222	C222
Cell (a, b, c (Å), α, β, γ (°))	31.7, 33.3, 79.2, 90, 91.5, 90	30.1, 37.2, 45.0, 90, 86.8, 90	27.9, 37.1, 43.7, 90, 90, 90	30.1, 36.4, 40.2, 90, 90, 90
Wavelength (Å)	0.9199 (peak)	1.5418	0.9792	0.9200 (peak)
Resolution range (Å)	79.1 - 2.4	44.9 - 2.8	43.6 - 1.5	23.0 - 2.6
Completeness (%)	94.6	93.2	98.9	77.1
R _{merge} on I	0.04	0.06	0.05	0.07
I/σ(I)	12.1	27.8	23.2	27.2
Number unique reflections	6534	1102	4066	576
Refinement statistics				
R (%)	16.8	25.0	19.8	25.0
R _{free} (%)	24.2	28.6	20.8	28.6
Rmsd bond distances (Å)	0.008	0.005	0.004	0.005
Rmsd angles (Å)	1.73	0.79	1.15	0.79
DNA strands / asymmetric unit	6	2	1	1
Water molecules	155	8	54	8
K ⁺ or Na ⁺ ions	14	4	4	4
PDB code	2AVJ	—	2AVH	—

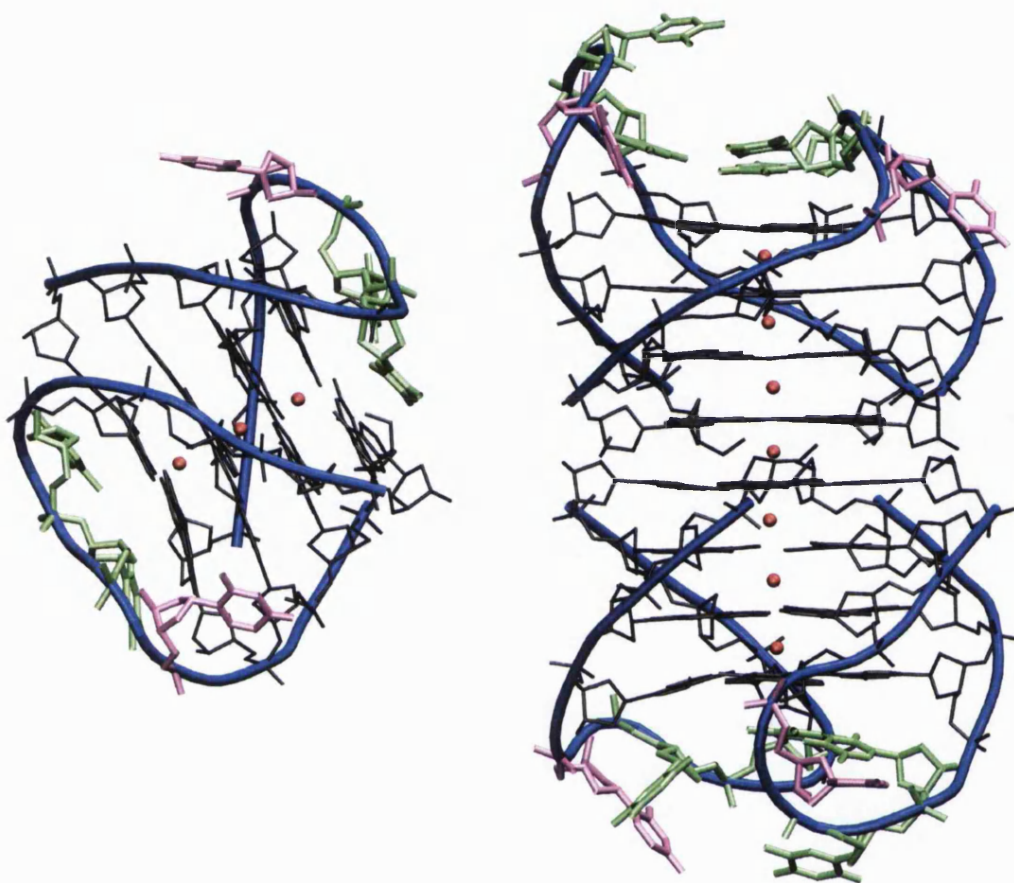


Figure 3.12: Contents of the $(G_4^{Br}UT_2G_4)_2$ $P2_1$ asymmetric unit. The backbone is traced in blue, G residues are black, T residues are coloured green, ^{Br}U residues purple and K^+ ions red.

at the junction of the two dimers. The dimers stack 5' to 3', and are aligned so that the strands from one quadruplex to the next are nearly continuous. Linking phosphate groups could be added to the strands without requiring any relative rotation of the dimers.

Apart from the dimer stacking, crystal packing interactions are formed by the loop residues. The residue numbering scheme for all the structures is detailed in Figure 3.13, and dimers are referred to in the text according to the strand names given in this Figure. The loop-loop interactions contributing to crystal packing in the $(G_4^{Br}UT_2G_4)_2$ $P2_1$ structure are shown in Figure 3.14. Loop stacking interactions are very similar between the three quadruplexes. Figure 3.14(a) shows how loop B interacts with two symme-

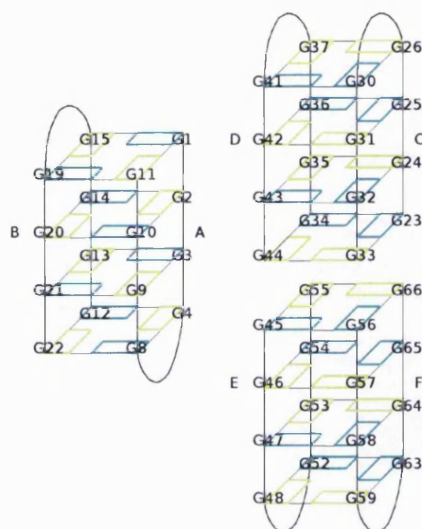


Figure 3.13: Numbering scheme used for the $(G_4^{BrUT_2G_4})_2 P2_1$ structure. The numbering scheme for the $(G_4T_3G_4)_2 C222$ crystal is the same as for strand A of the head-to-tail $P2_1$ quadruplex, on the left. The loop residues have been omitted for clarity. G residues with *anti* glycosidic angles are shown in yellow, and *syn* in blue.

try related head-to-tail dimers. The $^{Br}U16B$ residue is folded back into the quadruplex groove, and stacks on top of G1A, from a symmetry related molecule. This stacking is shown twice between three molecules in Figure 3.14(a). No hydrogen bonding interactions are formed to the T17B residue, as the Br atom of $^{Br}U16B$ points towards the former. The T17B and T18B residues are not involved in packing interactions. Chain A of the head-to-tail quadruplex interacts in a completely different manner with symmetry related molecules. This is shown in Figure 3.14(c), in which the $^{Br}U5A$ residue is inserted into a stack formed by the loop F residues T61F and T62F. The same stacking interaction occurs between loops C and D and their symmetry-related selves (Figure 3.14(b)). Residues T28C and T29C stack with $^{Br}U38D$ from a symmetry-related molecule. This extends the G-quartet stem by forming a three layer stack of loop residues, at both ends of the dimer of head-to-head dimers. The head-to-head dimers are related by a pseudo two-fold axis through the centre of the stack. The eight G-quartet stack has almost perfect two-fold symmetry, as have the last two residues of loops C and F. Loops D and E break the symmetry, and form different packing interactions. Loop D is shown in Figure 3.14(b), and loop E in Figure 3.14(c). In the latter, one residue stacks with the G-quartets (T51E), but the other two residues are flexible and not involved in any crystal packing interactions.

The six loops in the $(G_4^{Br}UT_2G_4)_2$ $P2_1$ structure only form two distinct types of packing interactions, as described above. Packing interactions between each pair of dimer loops are through formation of a three-layer stack of loop residues, extending the G-quartet core. Only strand B interacts with symmetry-related B strands in a different manner, by forming a stack of only one residue on top of the G-quartets (Figure 3.14(a)). The repeated packing interactions are further reflected in the loop conformational similarities between the various quadruplexes. The loop structures will be discussed in Section 3.2.5.

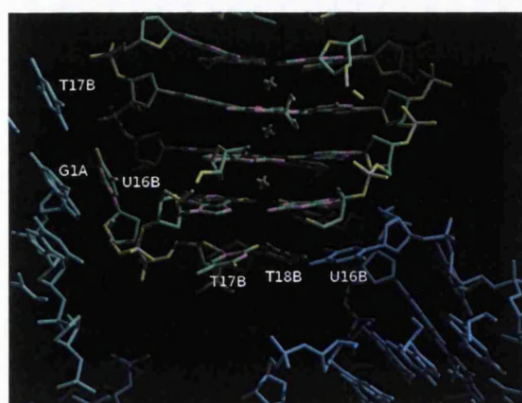
3.2.3 Crystal packing in the C2 and C222 crystals

The $(G_4^{Br}UT_2G_4)_2$ structure also crystallised in space group C2. The C2 structure is very similar to the $(G_4^{Br}UT_2G_4)_2$ head-to-tail dimer in the $P2_1$ structure. $(G_4T_3G_4)_2$ was solved in space group C222, using chain A from the $P2_1$ $(G_4^{Br}UT_2G_4)_2$ structure as a molecular replacement model. Only very small changes in the loop region were necessary in order to refine the structure. Both C2 and C222 crystals pack in essentially the same manner. The C2 crystal data were of much poorer quality than the $P2_1$ and C222 data, and will not be discussed in detail, other than to note the structural similarities.

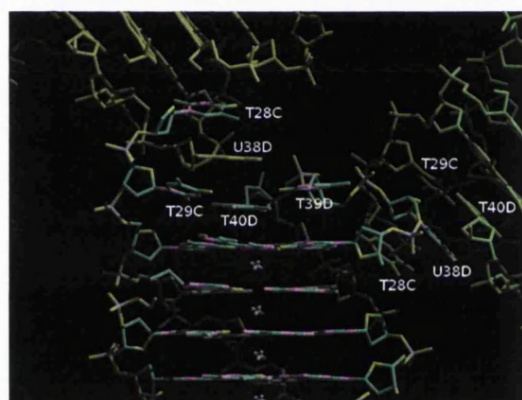
Crystal packing interactions are shown for the C222 $(G_4T_3G_4)_2$ structure in Figure 3.15. The $(G_4T_3G_4)_2$ C222 asymmetric unit contains a single strand, and the two strands of each quadruplex molecule are related by a two-fold crystallographic axis. Further packing interactions come from the loop residues. Figure 3.15 shows that a loop T residue (T1) is flipped out into the quadruplex groove, and stacks with a G residue (G10) from a symmetry related molecule. Unlike residue $^{Br}U16$ of loop B in the $(G_4^{Br}UT_2G_4)_2$ $P2_1$ structure, which stacks with a terminal G residue, in the C222 crystal, T1 stacks with a G residue within the G-quartets. This stacking interaction causes the G residue to bend slightly out-of-plane with the other G residues, causing all four G-quartets to deviate slightly from planarity. This, however, does not have any effect on the Hoogsteen hydrogen-bonding integrity within the quadruplex. Figure 3.15 also shows the three K^+ ion binding sites between each of the G-quartets.

The C2 $(G_4^{Br}UT_2G_4)_2$ and C222 $(G_4T_3G_4)_2$ structures are very similar, and form identical crystal packing interactions. The presence of the brominated residues therefore does not appear to have had any effect on the crystallisation or folding of the sequences considered.

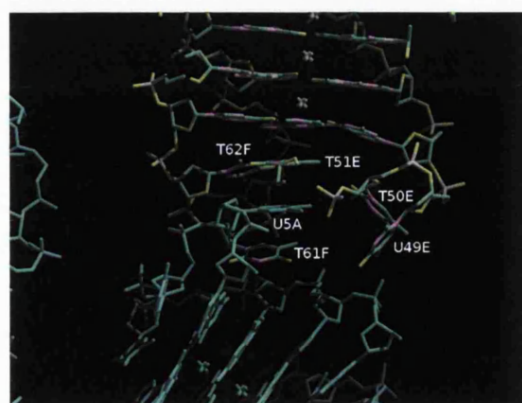
Crystal packing in the C222 $(G_4^{Br}UTAG_4)_2$ structure is identical to the $(G_4T_3G_4)_2$ C222 structure. The ^{Br}U residue is folded back into the quadruplex groove and stacks



(a) loop B



(b) Loops C and D



(c) Loops A, E and F

Figure 3.14: Crystal packing interactions for (a) the single G-quadruplex, (b) and (c) the stacked G-quadruplexes in the $(G_4^{Br}UT_2G_4)_2 P2_1$ structure. The residue numbering (atom name, number and chain) corresponds to Figure 3.13.

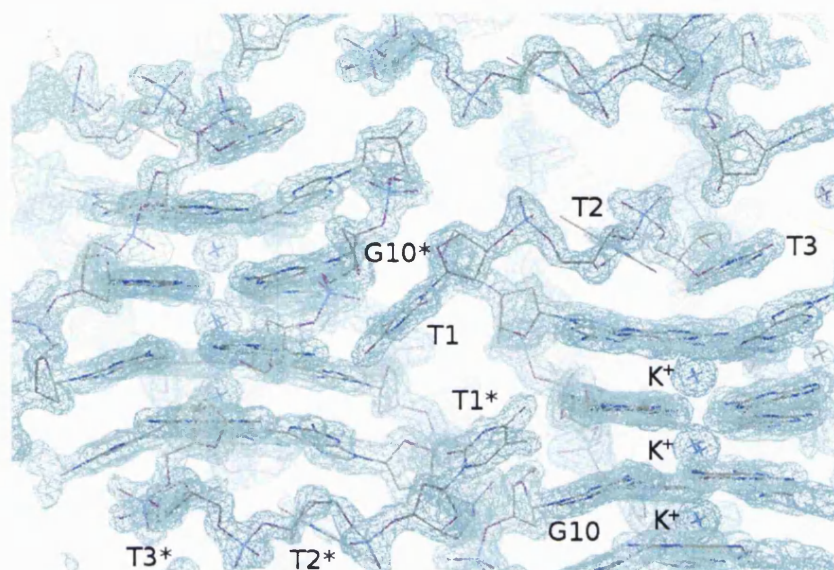


Figure 3.15: Crystal packing interactions in the C222 $(G_4T_3G_4)_2$ structure. The $2F_o - F_c$ electron density has been contoured at the 1.8σ level. K^+ ions are labelled in one quadruplex molecule, loop residues are labelled in two symmetry-related strands.

with a neighbouring G residue. The quartets are again slightly unplanar, due to this stacking interaction, as shown in Figure 3.16.

3.2.4 G-quartet structure

All the G residues were clearly defined, and formed Hoogsteen hydrogen bonded G-quartets. Figure 3.17 shows an example of G-quartet electron density in the C222 $(G_4T_3G_4)_2$ structure. The K^+ ion density below the plane of the G-quartet is also visible.

Two different quadruplex topologies were formed in the structures which were crystallised, head-to-tail dimers in the $P2_1$, C2 and C222 crystals, and head-to-head dimers in the $P2_1$ crystal. The head-to-tail dimers have a pattern of alternating *syn-anti* G glycosidic angles, both down the strands and around each of the G-quartets. This is illustrated in Figure 3.13. Two alternating groove widths are thus created, a narrow and a wide groove. All G-strands are antiparallel with respect to each other. The T_3 (and $^{Br}UT_2$) loops bridge the wide quadruplex grooves. The second topology is formed by the head-to-head dimerisation of strands in the $(G_4^{Br}UT_2G_4)_2$ $P2_1$ structure. Each G-strand in this dimer is neighbored by one parallel and one antiparallel strand. G

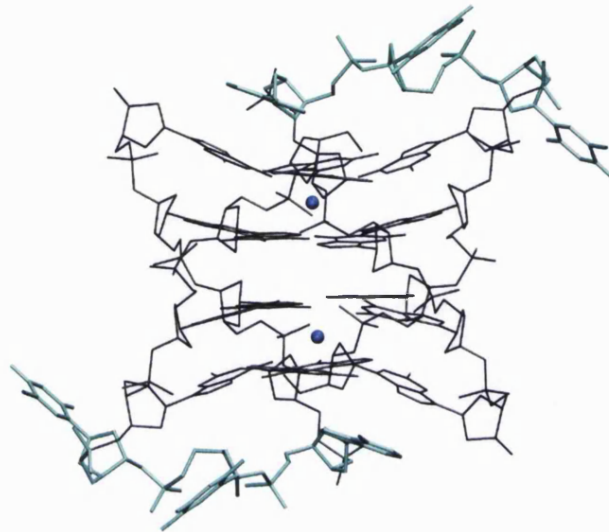


Figure 3.16: C222 ($G_4^{\text{BrUTAG}_4}$)₂ crystal structure. The T₂A loops are shown in cyan, and Na⁺ ions as blue spheres. Two asymmetric units, forming a whole quadruplex are shown.

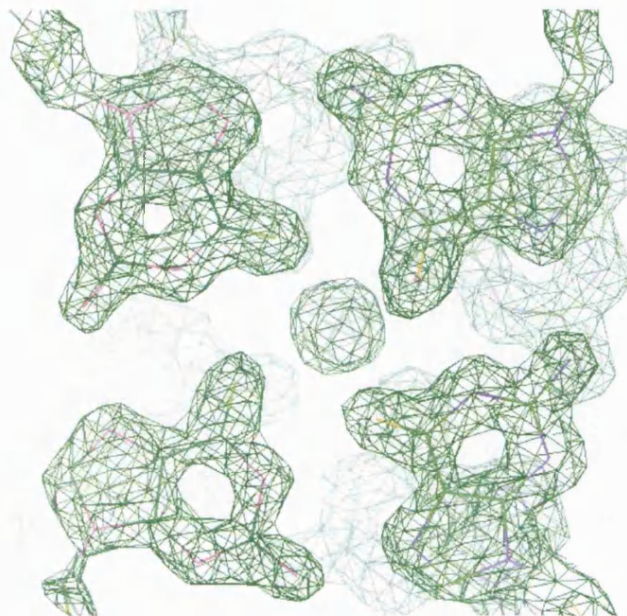


Figure 3.17: C222 ($G_4T_3G_4$)₂ G-quartet electron density, contoured at 1.5σ .

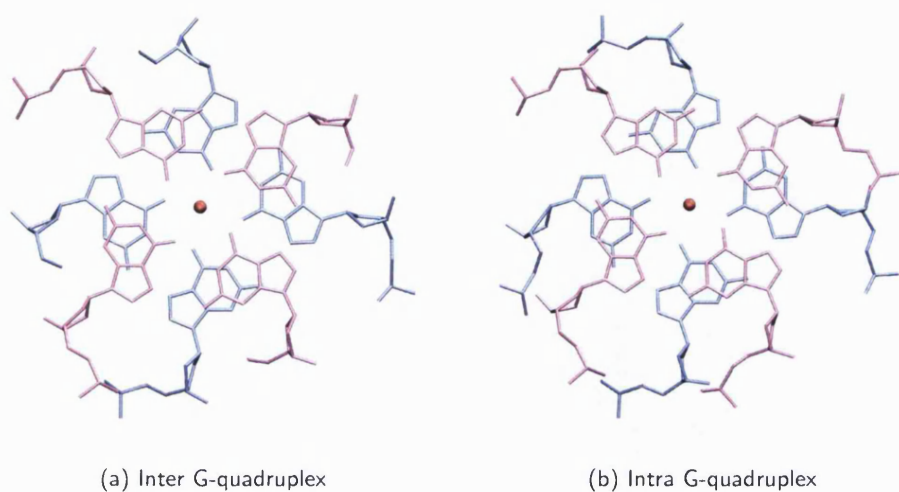


Figure 3.18: Inter- and intra-quadruplex G-quartet stacking in the $P2_1$ dimer of head-to-head dimers. K^+ ions between the two G-quartets are shown as red spheres.

residue glycosidic angles still alternate *syn-anti* down the strands, however around each G-quartet, the pattern is now *syn-syn-anti-anti*. Different groove widths are thus created, a narrow, two medium and one wide. The loops bridge each of the narrow and wide grooves on each of the two stacked dimers.

The stacking interaction formed by the two G-quartets at the interface of the two stacked dimers in the $(G_4^{Br}UT_2G_4)_2$ $P2_1$ structure is identical to that between G-quartets within a quadruplex. This is illustrated in Figure 3.18. This kind of quadruplex stacking interaction has not been seen in crystal structures before, although it has been suggested in solution, for example in the related sequence $G_3T_2AG_3$ ⁷⁴. G-quadruplex forming sequences used in X-ray crystallography (and often NMR) generally have terminal T or A residues attached to the G-rich ends, in order to promote packing interactions between molecules. These structures then generally form 5' to 5' quadruplex stacks, rather than 5' to 3', as observed here¹¹.

The $(G_4^{Br}UT_2G_4)_2$ and $(G_4T_3G_4)_2$ structures were annealed and crystallised in K^+ containing solutions, and K^+ ions were found between each of the G-quartet stacks in the $P2_1$, C2 and C222 crystals. The ions are equidistant from the G O6 atoms in all cases. Several K^+ ions were also identified in the quadruplex grooves, replacing water molecules. $(G_4^{Br}UTAG_4)_2$ was annealed in a Na^+ containing solution. Only two Na^+ ion binding sites were identified in the G-quadruplexes, one between each of the two terminal G-quartets. No extra density was found between the two central G-quartets.

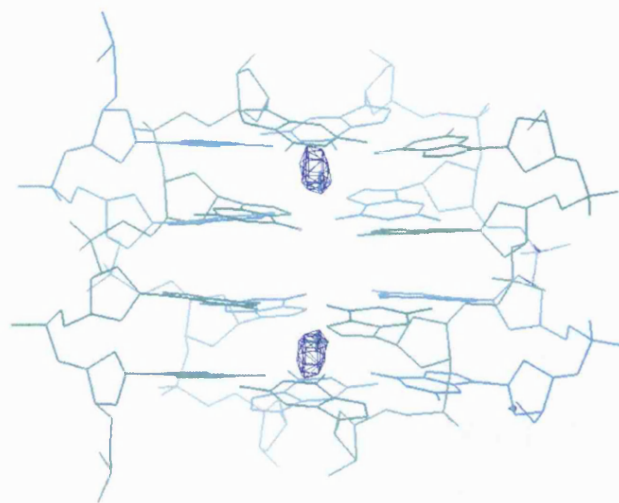


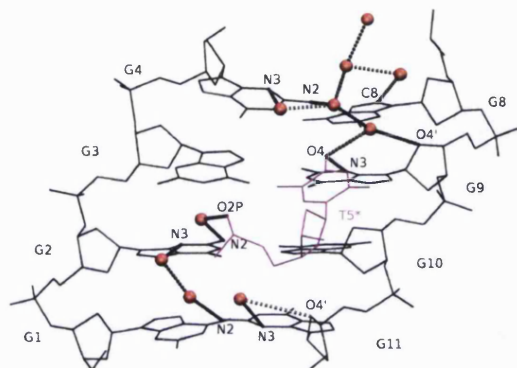
Figure 3.19: Na^+ ion omit map in the $(\text{G}_4^{\text{Br}}\text{UTAG}_4)_2$ structure. The density was contoured at 3.7σ .

Figure 3.19 shows the Na^+ ion omit map for the $(\text{G}_4^{\text{Br}}\text{UTAG}_4)_2$ structure. The two density peaks are located close to the outer G-quartet planes, confirming that the ions are Na^+ , rather than K^+ . The latter are too large to occupy in-plane positions. Due to the rather poor $(\text{G}_4^{\text{Br}}\text{UTAG}_4)_2$ data quality, it is difficult to determine whether the missing density in the centre of the G-quadruplex is due to the absence of any ions, or whether a very mobile Na^+ ion is present. Na^+ ions can occupy various positions within the quadruplex channel¹¹, and diffract only weakly.

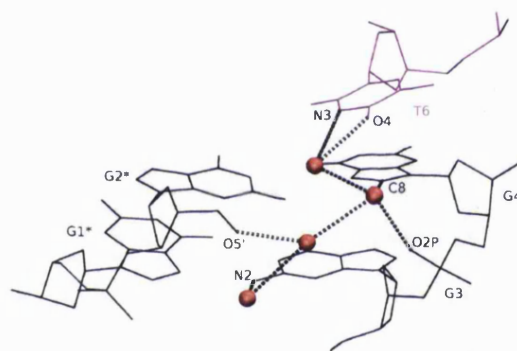
Clearly defined water molecules were located in most of the quadruplex grooves (Figures 3.20-3.23). The general hydration patterns are similar to those previously described for the $(\text{G}_4\text{T}_4\text{G}_4)_2$ structures^{5,51}. Even though the latter structure has diagonal, rather than lateral loops, the groove widths are similar to those of the head-to-head dimers in the $(\text{G}_4^{\text{Br}}\text{UT}_2\text{G}_4)_2$ $\text{P}2_1$ structure (a wide, narrow and two medium grooves).

In the narrow groove, the phosphate backbone faces towards the inside of the groove, and hence few water molecules can approach the G residue atoms. Hydration is generally limited to one water molecule interacting with the G N2 or C8 atoms at each G-quartet step. Spines of hydration are formed from the G residues to the phosphate backbone, and extend to the loop residues located above the narrow grooves. Figure 3.20 shows the water molecules present in the $(\text{G}_4\text{T}_3\text{G}_4)_2$ grooves, and Figure 3.21 shows the water molecules within the four narrow grooves formed in the $(\text{G}_4^{\text{Br}}\text{UT}_2\text{G}_4)_2$ $\text{P}2_1$ quadruplexes.

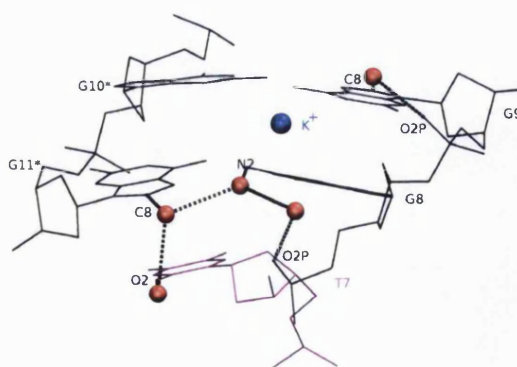
In the wide quadruplex grooves, the phosphate backbone faces outwards, which pro-



(a) Wide groove



(b) Narrow groove 1



(c) Narrow groove 2

Figure 3.20: Hydration in the $(G_4T_3G_4)_2$ grooves. Loop residues are coloured purple, water molecules red and K^+ ions blue. Hydrogen bonds are represented with dotted lines. Symmetry related residues are indicated with *.

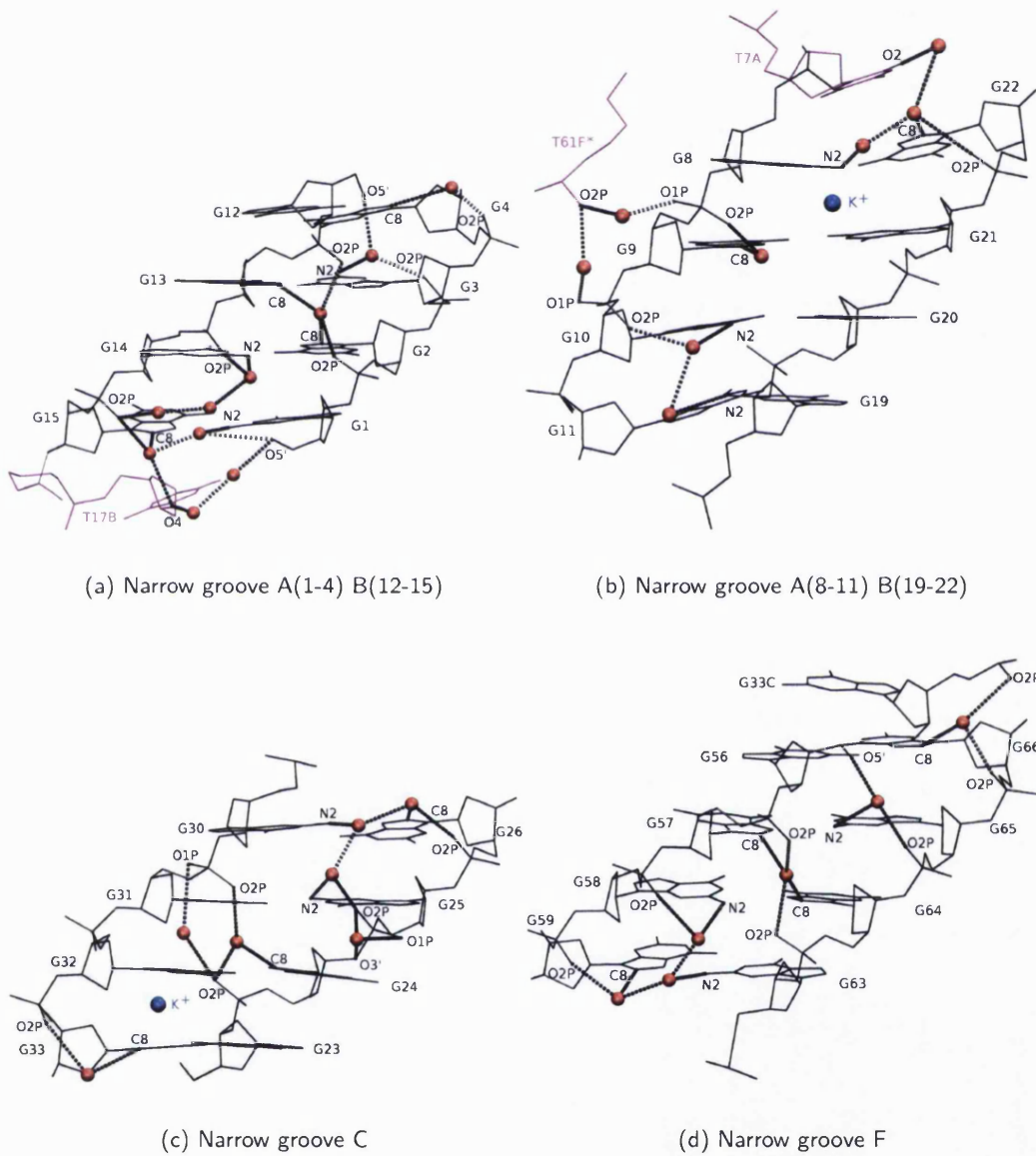
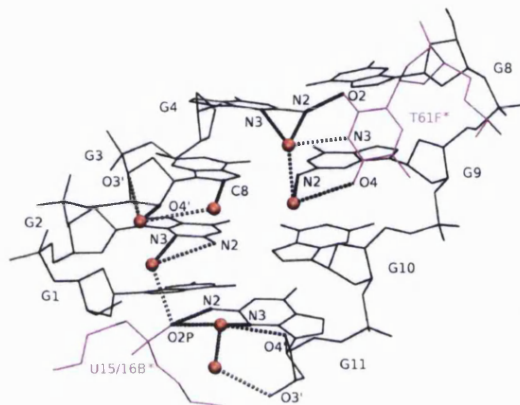
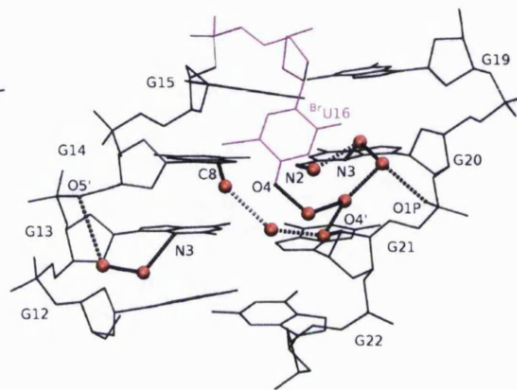


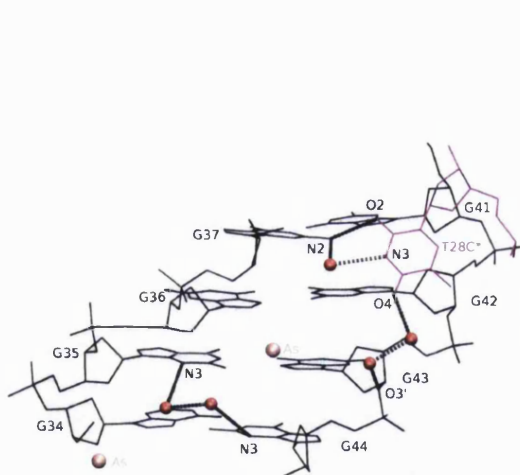
Figure 3.21: Hydration in the narrow grooves of the $(G_4^{Br}UT_2G_4)_2 P_21$ structure. Loop residues are coloured purple, water molecules red and K^+ ions blue. Hydrogen bonds are represented with dotted lines. Symmetry related residues are indicated with *.



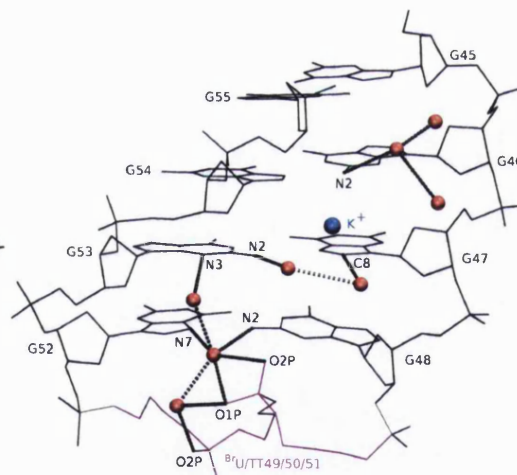
(a) Wide groove A



(b) Wide groove B



(c) Wide groove D



(d) Wide groove E

Figure 3.22: Hydration in the wide grooves of the $(G_4^{Br}UT_2G_4)_2 P2_1$ structure. Loop residues are coloured purple, water molecules red, K^+ ions blue and As pink. Hydrogen bonds are represented with dotted lines. Symmetry related residues are indicated with *.

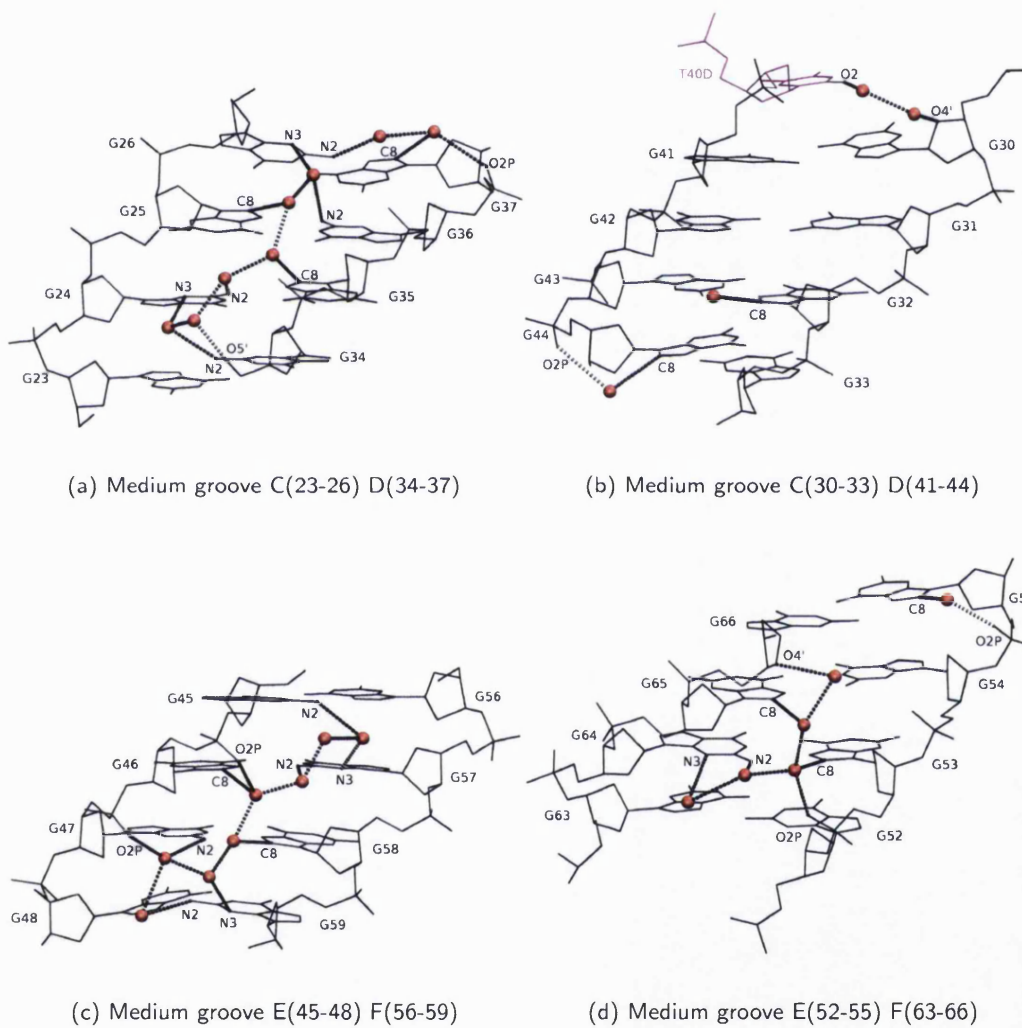


Figure 3.23: Hydration in the medium grooves of the $(G_4^{Br}UT_2G_4)_2$ $P2_1$ structure. Loop residues are coloured purple and water molecules red. Hydrogen bonds are represented with dotted lines.

vides greater access to the floor of the groove for water molecules to bind. Hydration of the wide grooves is shown in Figures 3.20 and 3.22. Out of the five wide grooves considered here (one in the $(G_4T_3G_4)_2$ C222 structure, and four in the $(G_4^{Br}UT_2G_4)_2$ $P2_1$ structure), four have a symmetry related residue located within the groove. This residue either hydrogen bonds directly to a G residue (such as T28C to G37D in the $(G_4^{Br}UT_2G_4)_2$ $P2_1$ structure, Figure 3.22(c)) or more frequently forms water-mediated interactions with the G residues in that groove. Only the wide groove formed by chain E in the $(G_4^{Br}UT_2G_4)_2$ $P2_1$ structure has no interactions with a loop residue (Figure 3.22(d)). The presence of loop residues within the grooves prevents water molecules from binding to certain inaccessible residues. Water molecules form hydrogen bonds to N2, N3 and C8 G atoms, which is in contrast to the $(G_4T_4G_4)_2$ structure, where only contacts to N2 and C8 were found^{5,51}.

The medium grooves in the head-to-head $(G_4^{Br}UT_2G_4)_2$ dimers have similar hydration patterns to those described above (Figure 3.23). Continuous spines of hydration were found in three of the medium grooves, however only very few water molecules were detected in the groove formed by G-strands C30-C33 and D41-D44 (Figure 3.23(b)). The groove is probably hydrated in the crystal, but the water molecules might be less ordered, explaining the lack of density.

Figures 3.20-3.22 also show the location of several K^+ ions bound within the wide and narrow grooves. Two arsenic ions (from the K^+ cacodylate buffer) were also located within the wide quadruplex groove in the $(G_4^{Br}UT_2G_4)_2$ $P2_1$ structure. These were unequivocally identified due to the strong anomalous peak. However, the density was spherical and did not allow unambiguous positioning of the O and C atoms in the cacodylate ion.

3.2.5 T_3 and T_2A loop conformations

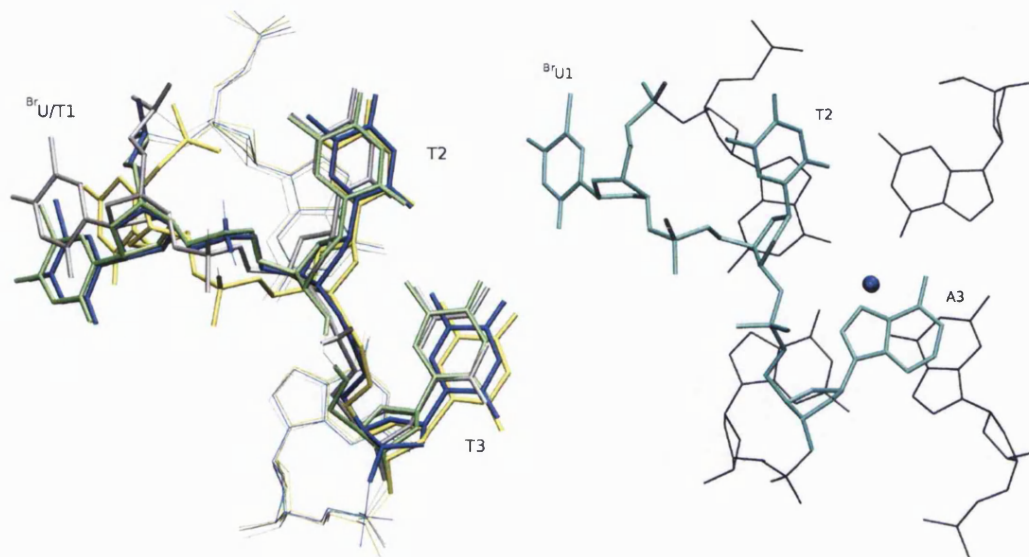
The $(G_4^{Br}UT_2G_4)_2$ $P2_1$ and $(G_4T_3G_4)_2$ C222 structures provided seven independent T_3 (or $^{Br}UT_2$) loop conformations. Five of these (the single T_3 C222 loop, two $^{Br}UT_2$ loops from the head-to-tail dimer and one $^{Br}UT_2$ loop from each of the head-to-head dimers in $P2_1$), bridge wide quadruplex grooves. A further two $^{Br}UT_2$ loops (from the head-to-head $P2_1$ dimers) bridge narrow grooves. The loops are all in crystallographically different environments, however there are considerable similarities between them. All but one of the loops bridging a wide groove are in the same conformation (type 1), where the first T (or ^{Br}U) residue is folded back in the quadruplex groove, and the other two T residues stack on the G-quartet plane. Figure 3.24(a) shows the overlap of $^{Br}UT_2$ loops A, B and D in the $P2_1$ crystal, and the T_3 loop in the C222 crystal. The groove T residue

position is variable, due to different crystal packing interactions between the different loop environments. This residue either stacks between two other residues (${}^{\text{Br}}\text{UT}_2$ $\text{P}2_1$ loops A and D), in which case it points out of the groove (green and blue loops in Figure 3.24(a)). On the other hand, in the ${}^{\text{Br}}\text{UT}_2$ $\text{P}2_1$ loop B, it stacks with a single G residue, and is deeper into the quadruplex groove (yellow loop in Figure 3.24(a)). The C222 T_3 loop T1 residue stacks with a G residue from a neighbouring quartet, and is therefore also quite far out of the groove (grey loop in Figure 3.24(a) and side view in Figure 3.15).

The type 1 loop is found in both the head-to-tail dimers in the C222 and $\text{P}2_1$ crystals (loops A and B), and the head-to-head dimer in $\text{P}2_1$ (loop D). The A, B and D G-strands are identical hairpins, with the same G residues and loop conformations. The second and third T residues in the type 1 loop are in very similar positions, despite the lack of any hydrogen-bonding interactions between them. This suggests that only the stabilisation gained by stacking with the G-quartets governs their position. These two T residues are not involved in crystal packing interactions in the $(\text{G}_4{}^{\text{Br}}\text{UT}_2\text{G}_4)_2$ $\text{P}2_1$ crystal, however in the $(\text{G}_4\text{T}_3\text{G}_4)_2$ C222 structure there is a water-mediated hydrogen bond between the third T residue and its symmetry-related self. The type 1 loop conformation is therefore independent of the Br atom (it occurs for both ${}^{\text{Br}}\text{UT}_2$ and T_3 loops), of the G-quartet topology (head-to-head or head-to-tail) and crystal packing interactions. This suggests that this loop conformation is due to real structural preferences, and not only to crystal packing effects.

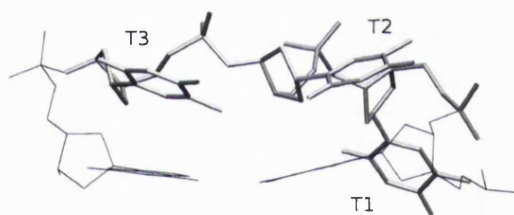
The fourth loop over a wide quadruplex groove (${}^{\text{Br}}\text{UT}_2$ $\text{P}2_1$ loop E) is less well defined (Figure 3.25). However, the third T residue also stacks on the G-quartet, in a similar position to that shown in Figure 3.24(a). The first and second ${}^{\text{Br}}\text{U}$ and T residues in loop E are much more flexible (${}^{\text{Br}}\text{U}49\text{E}$ and $\text{T}50\text{E}$ are shown in Figure 3.14(c)). Their B-factors are amongst the highest obtained for the loop residues in the $\text{P}2_1$ structure (Table 3.6).

A second conserved loop conformation is formed over the narrow quadruplex grooves in the $(\text{G}_4{}^{\text{Br}}\text{UT}_2\text{G}_4)_2$ $\text{P}2_1$ structure (type 2 loop), and is shown in Figure 3.26(a). In this case, the loop conformation is governed by crystal packing interactions. The second and third T residues form a stack on the G-quartet, with space in between them for a symmetry related residue (from loops A and D). This loop conformation is of course unlikely to be stable in solution in the absence of stacking of another residue. In the type 2 loop, only one T residue stacks on top of the G-quartets, which may be due to the narrower groove width prohibiting two residues from lying side-by-side. The first residue is displaced into the solution, and is much more flexible than the other two, as

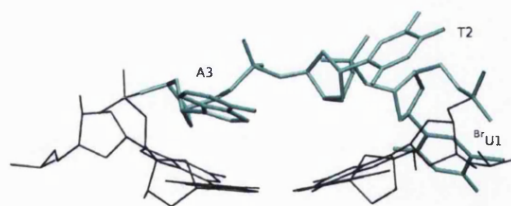


(a) P21 BrUT₂ loops A, B, D and C222 T₃ loop

(b) BrUTA C222 loop



(c) T₃ C222 loop



(d) BrUTA C222 loop

Figure 3.24: Type 1 loop conformation in the crystal structures. (a) Top view of the overlap of P21 BrUT₂ loops A (blue), B (yellow) and D (green) and the C222 T₃ loop (grey). The structures were overlaid using the two upper G residues only (shown in thin lines). (b) Top view of the BrUTA loop. (c) and (d) side view of the T₃ and BrUTA loops.

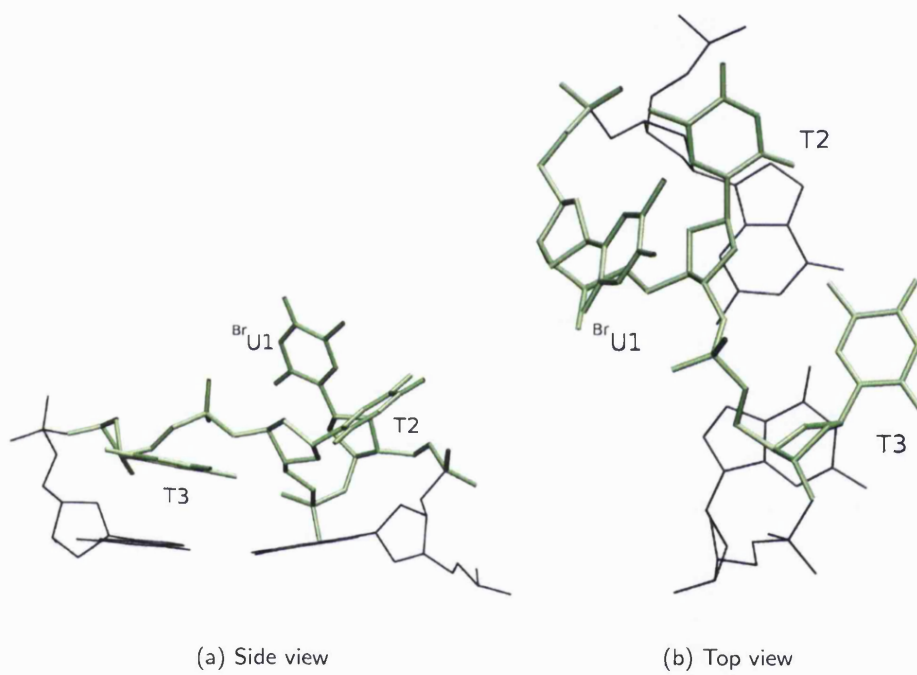


Figure 3.25: P21 BrUT₂ loop E. (a) Side view and (b) top view.

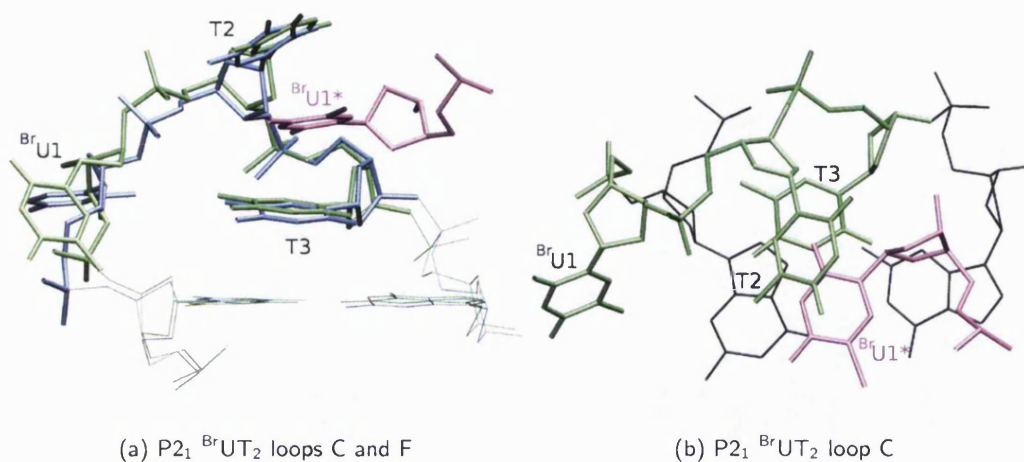


Figure 3.26: Type 2 loop conformation in the crystal structures. (a) Side view of the overlap of the $P2_1$ $^{Br}UT_2$ C (green) and F (blue) loops. The structures were overlaid using the two upper G residues only. (b) Top view of $^{Br}UT_2$ loop C. The symmetry related $^{Br}U38D$ residue is also shown in pink.

shown by the B-factors in Table 3.6.

Crystallographic B-factors for all loop residues in both the $(G_4^{Br}UT_2G_4)_2$ $P2_1$ and $(G_4T_3G_4)_2$ C222 structures are shown in Table 3.6. Loops in the head-to-tail dimers (C222 T_3 loop and $P2_1$ $^{Br}UT_2$ A and B loops) have the lowest B-factors. The loops in the $P2_1$ head-to-head dimers have higher B-factors, matching the higher average B-factors of the G-strands in these dimers. Loop F residues T61F and T62F form stacking interactions with loop A residue $^{Br}U5A$, and have lower B-factors than the other head-to-head dimer loop residues.

The $(G_4^{Br}UT_2G_4)_2$ $P2_1$ head-to-head dimers have adjacent loops on the same G-quadruplex face, enabling stabilising interactions to form between the loop residues. Similar hydrogen-bonding interactions are formed at both ends of the eight G-quartet stack (Figure 3.27). Three T residues from loops C and D form a planar hydrogen bonded triplet on the G-quartet surface (Figure 3.27(a)). Residues T29C and T40C interact via two hydrogen bonds, and form a further water-mediated hydrogen bond with residue T39D. In loops E and F, two residues hydrogen bond and stack on the G-quartet plane (T51E and T62F), however the third T residue (T50E) is further removed, and does not stack with the G-quartet. Instead, the latter forms a water-mediated hydrogen bond with a symmetry-related $^{Br}U5A$ residue, forming a second stacked layer on top of the first (Figure 3.27(b)). These interactions contribute to stabilising the two adjacent

Table 3.6: B-factors for the $P2_1^{\text{BrUT}_2}$ and C222 T_3 loops.

Loop	First T/ ^{Br} U	Second T	Third T
C222 T_3	18	25	17
$P2_1^{\text{BrUT}_2}$ A	22	27	28
$P2_1^{\text{BrUT}_2}$ B	19	18	18
$P2_1^{\text{BrUT}_2}$ C	57	44	41
$P2_1^{\text{BrUT}_2}$ D	35	47	35
$P2_1^{\text{BrUT}_2}$ E	56	52	29
$P2_1^{\text{BrUT}_2}$ F	30	24	22

lateral loops in the head-to-head dimers. In the $P2_1^{\text{BrUT}_2}$ loop D, the hydrogen bonding interactions with loop C on the G-quartet surface do not affect the type 1 loop conformation shown in Figure 3.24(a). Residues T39D and T40D are in the same type 1 conformation as residues which are not involved in hydrogen bonding interactions (such as loops A and B in the $(G_4^{\text{BrUT}_2}G_4)_2$ $P2_1$ head-to-tail dimer).

A single independent structure of the ^{Br}UTA loop was observed in the $(G_4^{\text{BrUTAG}_4})_2$ C222 crystal. This adopts a conformation nearly identical to the type 1 T_3 loops, with the ^{Br}U residue located within the quadruplex groove, and the T and A residues stacking on top of the G-quartets (Figures 3.24(b) and (d)). The groove ^{Br}U residue adopts the same position as the groove T residue in the $(G_4T_3G_4)_2$ C222 structure, and these form identical crystal packing interactions. In common with the ^{Br}UT₂ and T_3 loops, the second residue is the least well defined of the loop residues in the $(G_4^{\text{BrUTAG}_4})_2$ structure. The ^{Br}UTA loop conformation is governed only by the formation of stacking interactions, as there are no hydrogen bonds between the second T and A residues. T_2A and T_3 are therefore able to adopt similar conformations, independently of the nature of the third loop residue.

3.3 Comparison with other G-quadruplex structures

$(G_4^{\text{BrUT}_2}G_4)_2$ formed two distinct quadruplexes, a head-to-tail and a head-to-head dimer, within the same crystal. This suggests that both conformations are energetically similar. It remains unclear, however, whether the extra stabilisation gained by forming a stack of two G-quadruplexes is necessary for the formation of a dimer with adjacent lateral loops. The adjacent loops are able to form hydrogen-bonding interactions, which are not possible with the opposite-sided lateral loops (Figure 3.27). However,

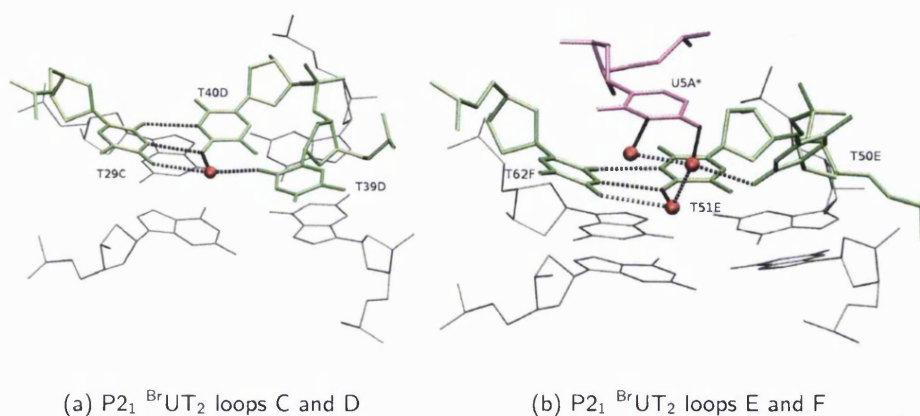


Figure 3.27: Hydrogen bonding and water-mediated hydrogen bonding interactions formed between residues on adjacent G-quadruplex faces in the head-to-head $P2_1$ dimers. A * indicates a symmetry related residue. Water molecules are drawn as red spheres and hydrogen bonds are indicated with dotted lines.

crystals containing a single quadruplex structure, such as the C2 and C222 forms of $(G_4^{BrUT_2}G_4)_2$ and $(G_4T_3G_4)_2$, formed only head-to-tail dimers. It is difficult to assess whether the formation of two stacked head-to-head dimers in the $P2_1$ crystal is due to crystal packing forces, or whether these structures exist, stacked or not, in solution. Evidence in favour of the potential coexistence of both dimeric forms in solution comes from the $(TG_4T_2G_4T)_2$ NMR structure, and chemical probing of the $G_3T_2AG_3$ sequence. The former was shown to form both head-to-head and head-to-tail dimers in solution, both having very similar energies⁵⁰. A head-to-head dimer of the $G_3T_2AG_3$ sequence was also found to form in solution, by detection of interactions between adjacent loop residues⁷⁴. It was further suggested that these dimers aggregated, presumably by stacking of the G-quartet planes. In order to gain a further understanding of the energetics of the head-to-tail and head-to-head $(G_4T_3G_4)_2$ dimers, MD simulations and free energy calculations were carried out to compare the relative stabilities of the two structural forms. The simulation results are described in Section 3.4.

The two $(G_4^{BrUT_2}G_4)_2$ structures obtained here are reminiscent of the $(TG_4T_2G_4T)_2$ head-to-head and head-to-tail dimeric solution structures⁵⁰. Both $(G_4^{BrUT_2}G_4)_2$ and $(TG_4T_2G_4T)_2$ form head-to-head dimers of an identical topology. Each strand is neighbored by a parallel and an antiparallel strand, and glycosidic angles are *syn-syn-anti-anti* around the G-quartets (Figure 3.13, page 71). On the other hand, the head-to-tail dimer formed by $TG_4T_2G_4T$ differs from that of the $G_4^{BrUT_2}G_4$ sequence illustrated in Fig-

ure 3.13. In $(TG_4T_2G_4T)_2$, the strand orientation is still alternating parallel/antiparallel, whereas the $(G_4^{Br}UT_2G_4)_2$ head-to-tail structure has all adjacent strands antiparallel. In the $(TG_4T_2G_4T)_2$ structures, both head-to-tail and head-to-head dimers have identical G-quartet stems, which is not the case for the $(G_4^{Br}UT_2G_4)_2$ structures. Both $(G_4^{Br}UT_2G_4)_2$ and $(TG_4T_2G_4T)_2$ structures are assembled from the dimerisation of two different G-strand topologies only. Two different hairpins of the $G_4^{Br}UT_2G_4$ sequence are formed, either creating a wide groove ($P2_1$ strands A, B, D and E) or a narrow groove ($P2_1$ strands C and F). The $G_4^{Br}UT_2G_4$ strands dimerise by forming either homodimers of the wide groove strands (head-to-tail dimer), or heterodimers of one wide groove and one narrow groove strand (head-to-head dimers). On the other hand, the $TG_4T_2G_4T$ sequence dimerises only by forming heterodimers of one wide groove and one narrow groove strand, either forming head-to-tail or head-to-head dimers.

Due to the different experimental conditions between the crystal and NMR structures, it is difficult to ascribe structural differences to the effects of loop length only, however there is some evidence that this might be having an effect. In the $(G_4^{Br}UT_2G_4)_2$ head-to-tail dimer, the loops bridge the wide grooves of the quadruplex. However, quadruplexes with lateral T_2 loops over the wide grooves were shown to be unstable during molecular dynamics simulations (Section 3.1.4). In the head-to-head $(G_4^{Br}UT_2G_4)_2$ dimer, the loops bridge a wide and a narrow groove, rather than two wide grooves. In this structure, the strain caused by one T_2 loop bridging a wide groove could be compensated for by the second loop bridging a narrow groove. The coexistence of two different G-tetrad cores within the same $(G_4^{Br}UT_2G_4)_2$ $P2_1$ crystal structure implies that both have similar stabilities. This is in contrast to the NMR structure of $(TG_4T_2G_4T)_2$ in which the same tetrad core was formed for both types of quadruplex coexisting in solution⁵⁰. The authors suggested that this reflected the greater stability of the *syn-syn-anti-anti* arrangement of G residues around the G-quartets, compared to *syn-anti-syn-anti*. This is not the case for the three-nucleotide loop structures in the present study, as both types of G-quartet core are present. It is possible that in the NMR study, the determining factor for the type of structure formed was not G-quartet core stability, but rather the effect of the shorter loops. As mentioned in Chapter 1, loops have been shown to have a profound effect on the G-quadruplex structure formed.

The structural plurality of G-quadruplexes is such that different structures are formed even within the same crystal. This raises the question of how feasible the targeting of specific G-quadruplex structures in solution is. The structures presented here do offer some scope for specifically targeting G-rich structures with T_3 loops over structures with T_4 loops. The latter form antiparallel quadruplexes with diagonal loops in all the

structures which have been solved to date. On the other hand, all the T_3 loops observed here were of the lateral type. Drugs could be designed to differentiate between lateral and diagonal loops. However, differentiating between T_3 and T_2 loops could be more difficult as both form lateral loop structures.

Both T_3 and ^{Br}UTA were able to form lateral loops of the same conformation (type 1), irrespective of the ions within the quadruplex channel. The $(G_4^{Br}UTAG_4)_2$ structure was solved in Na^+ ions, which did not alter the loop conformation. This is in contrast to the T_4 diagonal loop structure, which was shown to be dependent upon the ion in solution⁵². Of course, both the T_3 and ^{Br}UTA loops could adopt different conformations in different ionic conditions. However, the T_4 loop was also shown to be stabilised by the binding of a K^+ ion within the loop region⁵, whereas such ion-specific interactions were not observed in any of the structures solved here.

There are similarities between the type 1 T_3 loop structure obtained here, and that formed in the $(GCGGT_3GCGG)_2$ NMR structure¹³¹, shown in Figure 3.28. Both T_3 loops adopt the same general conformation, in which the first T residue is flipped out into the quadruplex groove, while the other two residues are positioned on top of the G-quartet plane. In the NMR structure, the third T residue is the most well defined, and stacks with the G-quartet plane. The second residue is located further away from the G-quartets, and is less well defined. This is similar to the T_3 loop conformation in the $(G_4T_3G_4)_2$ C222 crystal structure, where the second T residue has a higher B-factor than the other two T residues (Table 3.6). The first T residue, in the quadruplex groove, is however well defined in the crystal structure, because it is involved in crystal packing interactions, unlike in the NMR structure where it is more mobile. Lateral T_2A loops have also been observed previously, in the NMR structure of $AG_3T_2AG_3T_2AG_3T_2AG_3$, which contains a diagonal and two lateral loops⁶². However, in the latter, the T residues were ill-defined, and the structures cannot be compared.

G-quadruplex folding is not only determined by loop length. $(TAG_3T_2AG_3T)_2$ forms a dimeric parallel quadruplex with strand-reversal loops in the crystal in the presence of K^+ ions⁴⁷. However, T_2A loops are not sufficient to drive the formation of a parallel quadruplex, as the crystal structure of $(G_4^{Br}UTAG_4)_2$ shows (although this was obtained in the presence of Na^+ ions). Crystal structures of G-quadruplexes can hide the conformational flexibility of these molecules. NMR studies of G-quadruplexes have often revealed the presence of several structures in solution^{48,50}, which were not identified in the crystal.

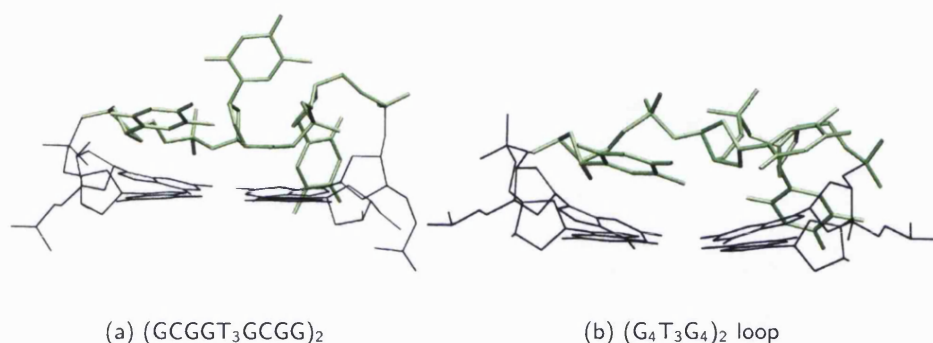


Figure 3.28: $(GCGGT_3GCGG)_2$ (1A6H PDB entry 1) and $(G_4T_3G_4)_2$ T₃ loops. A single G-quartet is also shown.

3.4 Comparison of X-ray structures and modelling

Theoretical methods were used in an attempt to predict the structure formed by the different sequences in this study (Section 3.1). No crystals of the $G_4T_2G_4$ sequence were obtained, however the simulation results were compared to the NMR-determined topology of the closely-related $TG_4T_2G_4T$ sequence (Section 3.1.4). Both the $G_4T_3G_4$ and $G_4T_2AG_4$ sequences formed dimeric quadruplexes with lateral loops in all the structures obtained. The most favourable T₃ loop predicted was a lateral, rather than diagonal loop, however diagonal loops had similar free energies (Table 3.3, page 57). Lateral loops bridging different groove widths also had similar free energies. The crystal structure of $(G_4^{Br}UT_2G_4)_2$ in space group $P2_1$ contains $^{Br}UT_2$ loops bridging both wide and narrow quadruplex grooves, which suggests that these are indeed energetically similar.

3.4.1 $(G_4T_3G_4)_2$ modelling

A range of possible lateral T₃ loop conformations were obtained using simulated annealing and MD simulations. Conformations similar to the T₃ type 1 loop (Figure 3.24(a)) were obtained using simulated annealing for both the T₃ (conformation T₃-L_w-3 in Table 3.2) and the T₂A loops (conformation T₂A-L_w-3 in Figure 3.9). However, these loop conformations were not identified as being more favourable than other conformations. In fact, during LES simulations of T₃-L_w-3, the loops adopted different conformations, with all three residues above the G-quartets, and no residue within the quadruplex groove. During the simulation of the 1A6H template with native T₃ loops, the type 1 conformation was maintained, although the first T residue within the quadruplex groove was

very flexible.

In order to further compare theoretical and experimental results, simulations of the X-ray structures were carried out. Head-to-tail and head-to-head dimers were simulated after replacing the ^{Br}U with T residues. Five separate, fully solvated, 4 ns simulations were carried out. Each of the three dimers in the $(G_4^{Br}UT_2G_4)_2$ $P2_1$ asymmetric unit was simulated separately, and the two head-to-head dimers were also simulated stacked as in the crystal structure. The $(G_4T_3G_4)_2$ $C222$ crystal structure was also simulated. Figure 3.29 shows selected average loop conformations over the final 2 to 4 ns of simulation. The $(G_4^{Br}UT_2G_4)_2$ head-to-tail dimer, formed by strands A and B, was stable during the simulations, and the average structures of the loops, shown in Figures 3.29(a) and (b) are very close to the X-ray conformation. This suggests that the type 1 loop, in which two residues stack with the G-quartets, and the first residue is located in the groove, is stable in solution. The head-to-tail dimers, in both $(G_4^{Br}UT_2G_4)_2$ and $(G_4T_3G_4)_2$ crystal structures, had slightly buckled G-quartets (Figure 3.30(a)). During the simulations, in absence of crystal packing interactions, the G-quartets became planar, as shown by the $(G_4T_3G_4)_2$ structure averaged over the last 2 to 4 ns of simulation in Figure 3.30(b). These simulations confirm that the slight distortion of the G-quartet planes in the crystal is caused by crystal packing interactions, rather than any loop effects. Moreover, in simulated solution, the first T residue of each loop moved closer to the floor of the wide quadruplex groove, again due to the absence of crystal packing interactions.

Not all the type 1 loops simulated remained in the starting conformation, as shown in Figures 3.29(c) and (d). Loop D of a head-to-head dimer was more flexible during the simulations. The general loop conformation remained similar, although the third T residue stacked with a loop residue rather than with the G-quartets. This second conformation was stable over the course of the dynamics. Loops C and F formed type 2 loops, shown in Figure 3.26(a) (page 86), which were involved in crystal packing interactions. During the simulations, the second and third T residues formed a stable stack over the G-quartets (Figures 3.29(e)-(h)). The first T residue was more flexible and generally more exposed to the solvent. Despite the fact that loops over the narrow groove in the head-to-head dimers are involved in crystal packing interactions, stable conformations were adopted in the simulations, showing that such loops can be stable even outside the context of the crystal. The small changes in loop conformations during the dynamics affected the loop-loop interactions between adjacent loops of the head-to-head dimers. Hydrogen bonding between loops E and F (Figure 3.27(b)) was not maintained during either of the strand E and F head-to-head dimer simulations. This

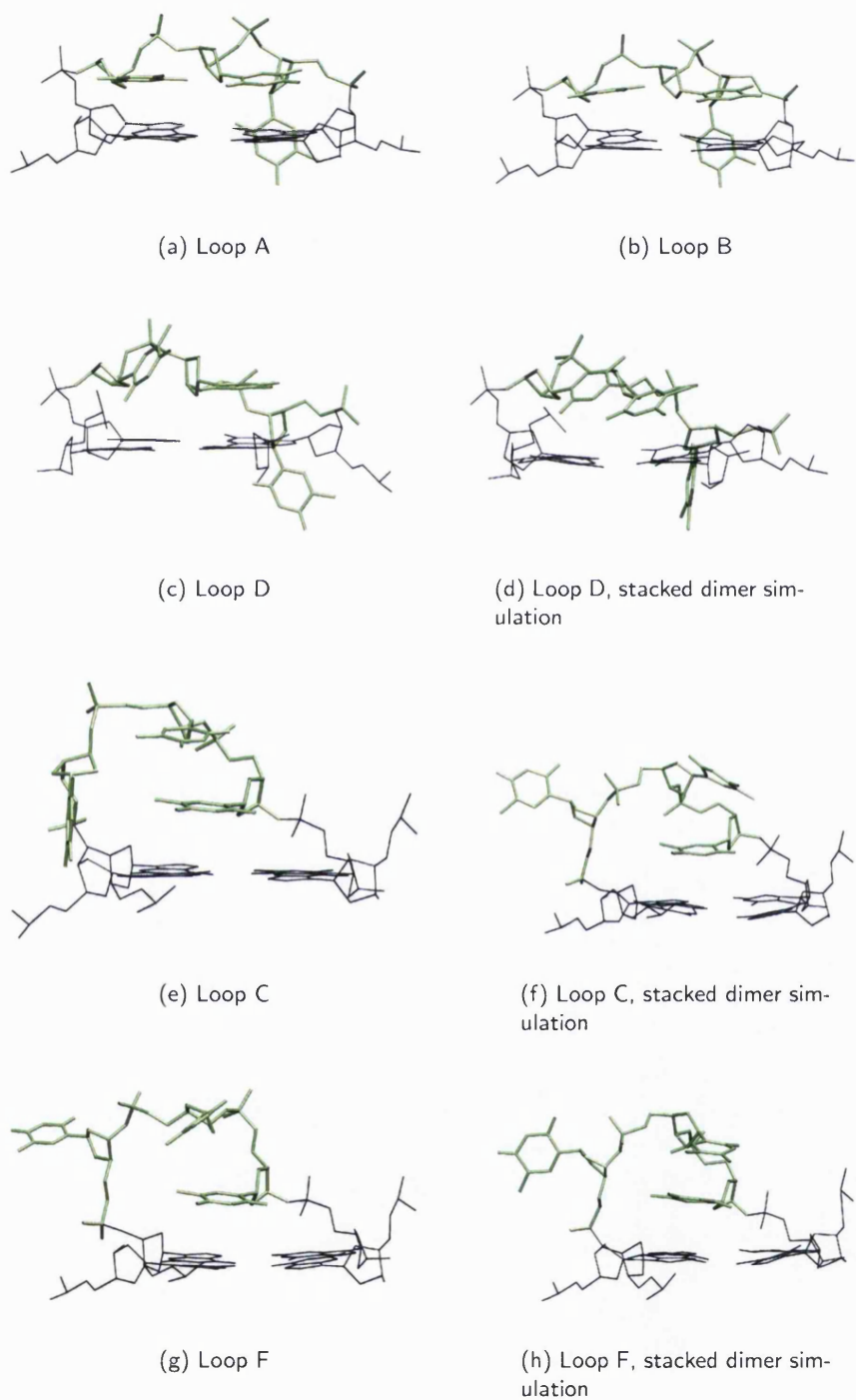


Figure 3.29: Average loop structures over the last 2 to 4 ns of dynamics for the $P2_1$ dimer simulations. (^{Br}U residues were replaced with T).

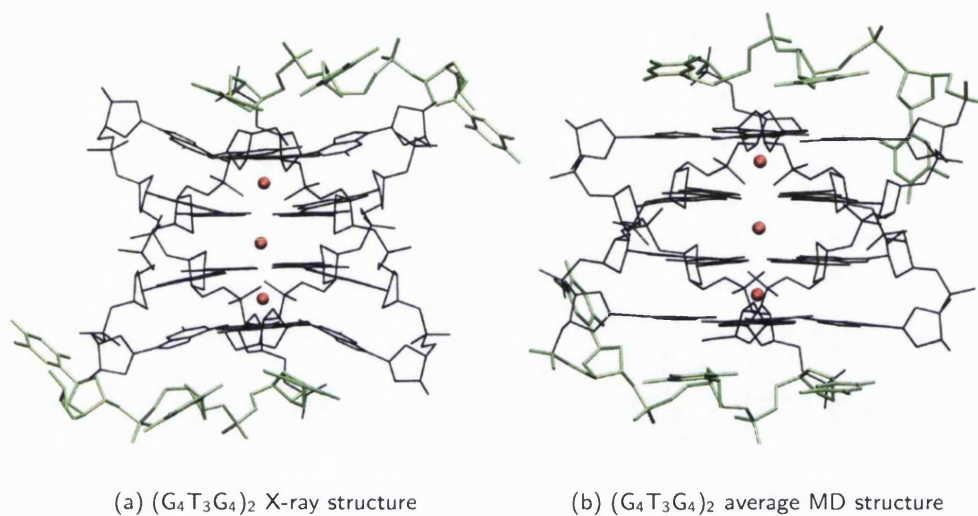


Figure 3.30: $(G_4T_3G_4)_2$ C222 X-ray structure and average structure over the final 2 to 4 ns of dynamics.

did not affect the stability of the loops, as stacking interactions with the G-quartets were conserved. Loops C and D also formed hydrogen bonding interactions (Figure 3.27(a)), and the direct hydrogen bonds were conserved during the simulations.

Energetics of the experimental structures were calculated from the MD simulations, in the same manner as described in Section 3.1.2. G_{total} and individual loop contributions G_{loop1} and G_{loop2} are shown in Table 3.7. Figure 3.31 shows the G_{total} values at each step of the simulations. The linear regression to the data shows that the structures are generally equilibrated, and that the starting (X-ray) structures were already in low energy conformations.

The stabilisation gained by the head-to-head dimers upon formation of an eight G-quartet stack can be estimated by comparing the free energies of the separate dimers, and the energy of the two stacked dimers. The two quadruplex stack is on average $112 \text{ kcal.mol}^{-1}$ more stable than the sum of the two individual dimers ($G_{\text{total}} = -9159 \text{ kcal.mol}^{-1}$ for the dimer stack, and individual dimer energies taken from Table 3.7). Most of the increased stabilisation comes from the extra K^+ ion located between the two G-quadruplex dimers, whereas only three K^+ ions were included in each individual G-quadruplex calculation. Overall, Table 3.7 shows that the head-to-head dimers tend to be less favourable structures than the head-to-tail dimers. The lowest energy head-to-tail dimer is 11 kcal.mol^{-1} more favourable than the lowest energy head-to-head dimer.

Table 3.7: G_{total} values (kcal.mol^{-1}) calculated using MM-PBSA for the crystal structure simulations. All values were calculated over the final 2 to 4 ns of dynamics. Values for the head-to-head dimers (strands C, D and strands E, F) were calculated both during the separate and the stacked dimer simulations. BrU residues in the $(\text{G}_4^{\text{BrU}}\text{UT}_2\text{G}_4)_2$ P2_1 structure were mutated to T residues prior to simulation. Type 1 loops are highlighted in magenta. Standard errors of the mean are in brackets.

Dimer	Strands	G_{total}^*	G_{stem}	G_{loop1}	G_{loop2}	$-\text{TS}_{\text{total}}$
C222 head-to-tail	A, A	-4556 (4)	-3639 (4)	-455 (4)	-462 (4)	-582 (1)
Separate simulations						
P2_1 head-to-tail	A, B	-4551 (4)	-3634 (4)	-457 (4)	-460 (4)	-581 (1)
P2_1 head-to-head	C, D	-4545 (4)	-3635 (4)	-447 (4)	-456 (4)	-579 (1)
P2_1 head-to-head	E, F	-4523 (5)	-3637 (4)	-445 (4)	-441 (4)	-585 (2)
Simulation of the stacked dimers						
P2_1 head-to-head	C, D	-4520 (4)	-3629 (4)	-440 (4)	-448 (4)	-579 (1)
P2_1 head-to-head	E, F	-4527 (4)	-3629 (4)	-447 (4)	-445 (4)	-583 (2)

*G values do not include the solute entropy

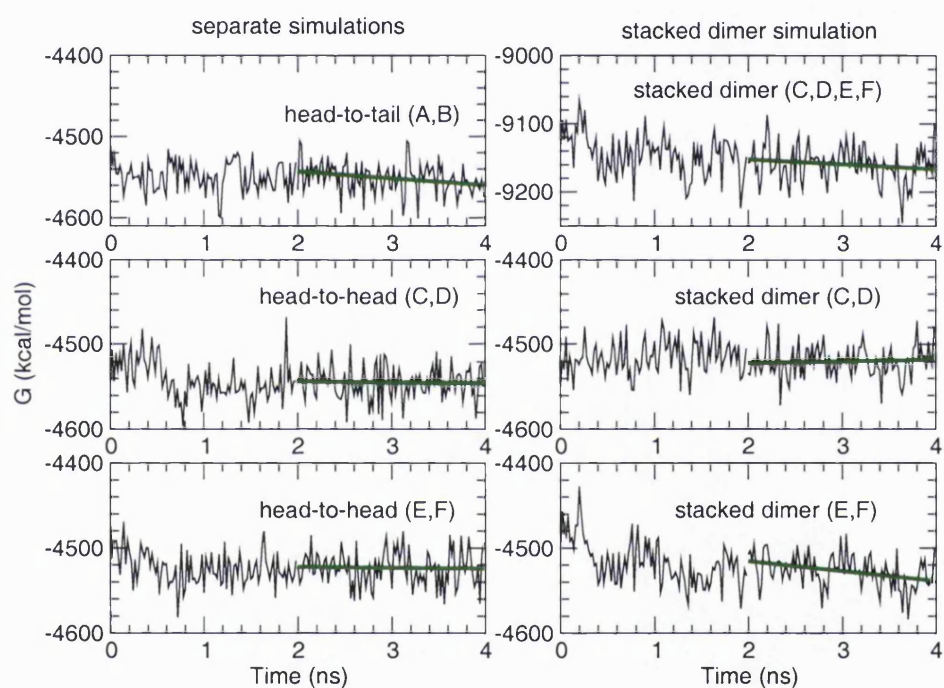


Figure 3.31: Plots of G_{total} ($\text{kcal}\cdot\text{mol}^{-1}$) calculated using MM-PBSA during simulations of the $(G_4^{\text{Br}}\text{UT}_2G_4)_2$ $P2_1$ structures. G_{total} values do not include the solute entropic contribution. Green lines are the linear regression to the data over the final 2 to 4 ns of simulation.

Table 3.7 also shows the free energy of the two head-to-head dimers simulated together as a stack of dimers. During this simulation, the head-to-head dimer G_{stem} energies were less favourable (by at least 6 kcal.mol^{-1}) than during their simulations alone in solution, although this energy difference is close to the MM-PBSA error margin. Lower G_{stem} values could be due to the increased rigidity caused by the greater number of G-quartet stacks, which does not allow the G-stems to equilibrate as well as when the dimers are separate in solution. However, no difference in the entropic contribution was calculated between the separate and stacked simulations.

According to the MM-PBSA calculations, the difference between the head-to-head and head-to-tail dimers is due mostly to the loop energies, as G_{stem} energies are within 5 kcal.mol^{-1} of each other in the separate simulations. This suggests that the energy difference between alternating *syn-anti-syn-anti* and *syn-syn-anti-anti* G glycosidic angles around the G-quartets is very small. Loops in the type 1 conformation over the wide quadruplex groove were the most favourable, as highlighted in magenta in Table 3.7, which is consistent with their high occurrence in the various crystal structures.

The predicted T_3 loop conformations had similar free energies to those calculated during the experimental structure simulations. The most favourable predicted T_3 loop free energy was $-461 \text{ kcal.mol}^{-1}$ (Table 3.3, page 57). However, this was for a loop with three residues stacking on top of the G-quadruplex (see Figure 3.4(d), page 52). The experimental structures are no more favourable than the predicted structures, which may be due to the inability of the force field to differentiate between loop conformations. T_3 loops over the narrow quadruplex grooves (formed by strands C and F) were similar, after simulation, to the $T_3\text{-L}_n\text{-2}$ predicted structure (Figure 3.6(b), page 55). The latter also had more favourable free energies, shown in Table 3.3, than any of the experimental narrow groove loops simulated (Table 3.7). The experimental narrow groove loops are however difficult to compare with predicted conformations, as they may adopt completely different structures in the absence of crystal packing constraints.

3.4.2 $(G_4T_2AG_4)_2$ modelling

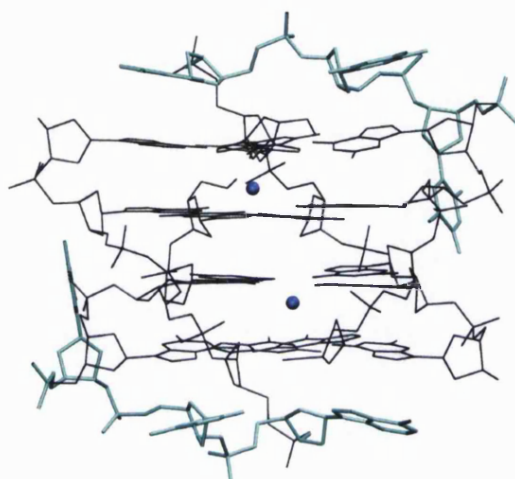
MD simulations of the $(G_4^{\text{Br}}UTAG_4)_2$ crystal structure were carried out after replacement of $^{\text{Br}}\text{U}$ with a T residue. The crystal structure was determined in Na^+ , hence the simulations were carried out with Na^+ ions within the quadruplex channel, rather than K^+ ions. The system charge was also neutralised with the addition of solution Na^+ ions. Only two channel Na^+ ions were located in the crystal structure and included in the simulations.

The simulation of the crystal structure with two Na^+ ions within the channel was

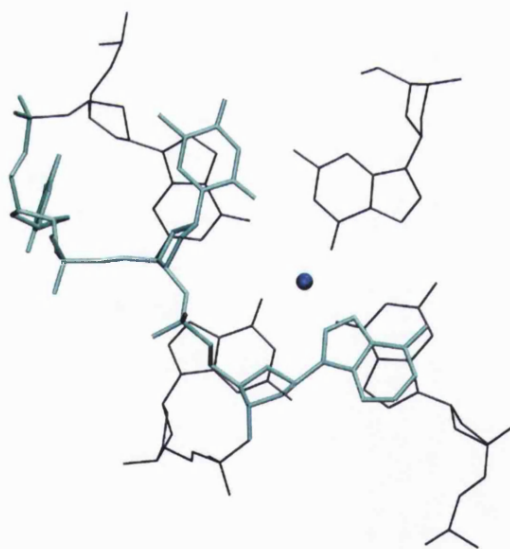
relatively stable over 4 ns, some distortion of the G-quadruplex core being observed, as shown in Figure 3.32. The G-quartets slipped with respect to each other, so that the central channel is no longer aligned. This is made apparent by the Na⁺ ion positions in Figure 3.32(a). On the other hand, the simulations were able to reproduce the loop residue stacking interactions. Figure 3.32(b) shows that the middle T and A residues deviated little from the crystal structure (shown in Figure 3.24(b), page 84).

In order to attempt to determine the cause of the G-quartet slippage, a simulation with three Na⁺ ions within the central channel was carried out. The third Na⁺ ion was added between the two central G-quartets, equidistant from the eight G residue O atoms. The three Na⁺ ions remained in the starting binding sites for the first 800 ps of simulation. After that time, one of the terminal Na⁺ ions moved out of the channel and into the loop region. This was followed 200 ps later by movement of the central ion into the binding site left vacant by the first ion. The three ions then remained stable for the rest of the simulation. The path of the three Na⁺ ions during the simulation is shown in Figure 3.33. The strand slippage which was observed during the simulation with two Na⁺ ions was also observed in this simulation, although this was less pronounced. MD simulations therefore suggest that only two Na⁺ ions are required for stability. The smaller radius of Na⁺ compared to K⁺ ions mean that the former are much more mobile during the simulations. The behaviour of the Na⁺ ions is unlikely to be due to insufficient screening of the charge-charge repulsion between two cations in the channel. During the final 3 ns of simulation, the two bottom ions shown in Figure 3.33 were separated by the same distance as if they were both within the quadruplex channel. The crystal structure of (TG₄T)₄, also determined in Na⁺ ions, clearly showed the presence of Na⁺ ions at each G-quartet step¹¹. The crystallographic data obtained here for (G₄^{Br}UTAG₄)₂ are not of sufficient quality to ascertain the absence of a Na⁺ ion within the two central G-quartets. MD simulations do suggest that the structure has a preference for two Na⁺ ions over three. However, results concerning ions obtained with MD simulations are not always reliable, due to limitations of the additive force field (see Chapter 6).

The experimental T₂A loop conformation was predicted with simulated annealing methods, as shown in Figure 3.9(b), page 60. The T₂A-L_w-3 loop has a similar general conformation, with the first loop residue in the quadruplex groove and two other residue stacking with the G-quartet plane. However, as previously, this loop conformation was not preferred compared to the other loops obtained. Moreover, the simulated annealing runs were unable to reproduce the experimental preference for *anti* glycosidic angles in the type 1 loop residues, as both *syn* and *anti* residues were predicted.



(a) Side view



(b) Top view

Figure 3.32: $(G_4T_2AG_4)_2$ structure averaged over the final 2 to 4 ns of simulation. Two Na^+ ions were included in the central quadruplex channel. Loops are shown in cyan, Na^+ ions in blue.

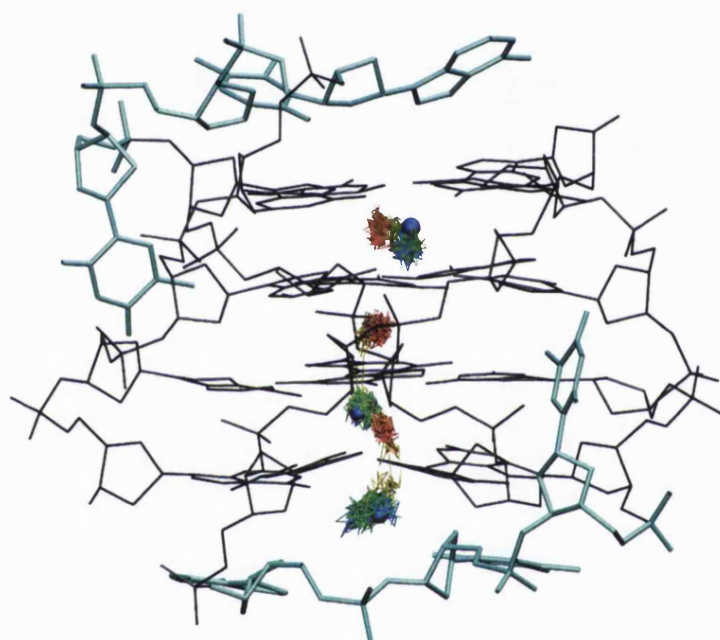


Figure 3.33: Final structure of $(G_4T_2AG_4)_2$ after 4 ns MD with three Na^+ ions included in the quadruplex channel. The loops are shown in cyan, and Na^+ ions as blue spheres. The path of each ion during the trajectory is shown. The colour scale corresponds to the trajectory frames: red from 0 to 1 ns, ochre from 1 to 2 ns, green from 2 to 3 ns and finally blue from 3 to 4 ns.

3.4.3 $(G_4T_2G_4)_2$ modelling

T_3 and T_2A loop conformations were difficult to predict with MD simulations because all structural types were found to be equally possible. This was not the case for T_2 loop quadruplexes, whose structures are constrained by the limited span of the two-nucleotide loop. In this case, the particular loop conformation is less important, as certain T_2 loop structures were found to be unstable due to strain in the loop backbone. In order to compare the T_3 and T_2A loop structures with potential T_2 loop structures, the $(G_4^{Br}UT_2G_4)_2$ X-ray structures were simulated, after replacing the $^{Br}UT_2$ loops with T_2 loops.

A $(G_4T_2G_4)_2$ head-to-head dimer, with loops over the narrow and wide grooves, was stable over a 4 ns simulation, and the loop residues were able to form stacking interactions with the G-quartets (Figure 3.34(a)). However, the head-to-tail dimer, which has both T_2 loops over wide quadruplex grooves, became distorted during the simulation (Figure 3.34(b)). It therefore appears that the strain caused by a T_2 loop bridging a wide groove can be compensated for by the second loop bridging a shorter distance. However, two T_2 loops bridging wide grooves are unstable. The stable head-to-head $(G_4T_2G_4)_2$ dimer simulated has the same topology as one of the $(TG_4T_2G_4T)_2$ solution NMR structures⁵⁰. The latter sequence also forms a head-to-tail dimer in solution, however this has a different topology to the $(G_4T_3G_4)_2$ head-to-tail dimer. As mentioned in Section 3.1.4, it is possible for the different head-to-tail topologies observed to be due to effects of the loop lengths, rather than G-quartet stem stability. The simulations carried out here tend to support this view. A head-to-tail dimer with both loops over the wide groove, as formed by the $G_4T_3G_4$ sequence, cannot support two T_2 loops without the G-quartets becoming distorted. However, dimerisation which enables T_2 loops to bridge a wide and a narrower groove can lead to stable structures.

Further information is difficult to obtain from MD simulations. The $(TG_4T_2G_4T)_2$ structures were solved in Na^+ containing solutions, however the simulations were carried out here with K^+ ions between the G-quartets. The simulations cannot reproduce experimental differences due to the counterions present, which means that ionic effects cannot be taken into account. However, $(G_4^{Br}UTAG_4)_2$ forms a head-to-tail dimer with the same topology as $(G_4T_3G_4)_2$, and the former structure was solved in Na^+ ions. It is therefore unlikely that replacing the K^+ ions with Na^+ ions would stabilise the G-quartet core of a $G_4T_2G_4$ head-to-tail dimer with two highly strained loops.

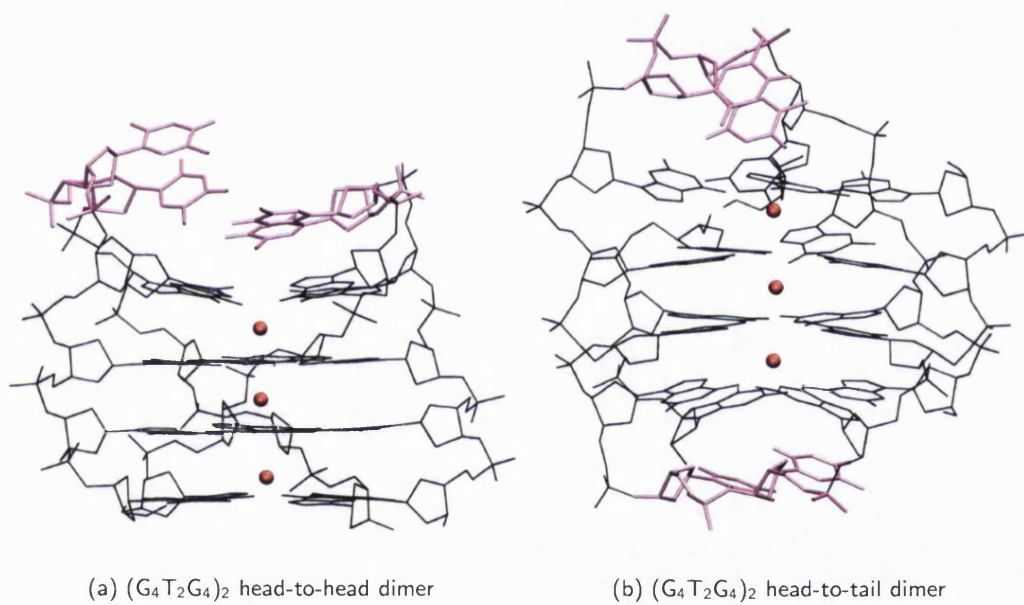


Figure 3.34: Average structures over the final 2 to 4 ns of dynamics of the head-to-head (strands C, D) and head-to-tail (strand A, B) $P2_1$ dimers after replacement of the $^{Br}UT_2$ loops with T_2 loops.

3.4.4 Conclusion

The development of molecular dynamics methods is ongoing, and force fields are continually being improved. The problems associated with the prediction of quadruplex loop conformations have already been pointed out when looking at T₄ loops⁹⁴. Not only were the authors unable to predict the experimental diagonal loop conformation, but the experimental loop structure was also unstable during MD simulations. T₃ and T₂A loops of the type 1 conformation were predicted in the simulated annealing runs, however they tended to be unstable during MD and LES simulations (see Figure 3.4, page 52). The experimental T₃ loop conformations were stable during the simulations, indicating that the problem with the predictive simulations might have been due to the starting structures. Initial structures were obtained from implicit solvent simulated annealing runs, and these may not have been energetically close enough to favourable conformations in explicit solvent.

As several NMR structures have shown, quadruplex loops can be flexible in solution, and different conformations are probably very similar in energy^{48,78}. In such cases, it is difficult to know whether the inability of the simulations to predict loop conformations is due to limitations in the force field, or because the loops are inherently flexible. T₃ and T₂A formed only lateral loops in the crystal structures solved, rather than diagonal loops. However, the simulations were unable to differentiate between lateral and diagonal loop structures. Simulations of the X-ray structures showed that the loop conformations were stable in solution, and hence not only in the context of crystal packing interactions. This suggests that the force field used is adequate to simulate these types of loop conformations, which are not stabilised by cations.


Chapter 4

Modelling of monomeric G-quadruplex folding

Modelling of the $(G_4T_3G_4)_2$, $(G_4T_2AG_4)_2$ and $(G_4T_2G_4)_2$ dimeric G-quadruplexes showed that theoretical methods could provide information about G-quadruplex folding, especially for sequences with shorter loops. In this Chapter, the study of loop length effects on G-quadruplex folding has been broadened by looking at a series of monomeric structures. Rather than attempt to determine specific loop conformations, simulations were used to determine whether certain loop lengths favoured parallel or antiparallel quadruplexes. Previous MD simulations were unable to differentiate between different conformations of a single loop (Section 3.1). However, simulations can reveal whether certain quadruplexes are unlikely to form in solution, due to strain caused by the loops. The particular loop conformation itself was less important, as long as a conformation could be found which would not destabilise the quadruplex. A greater range of loop lengths was considered, ranging from T to T₆.

Simulations were based on experimental structures of the human telomeric sequence, AG₃T₂AG₃T₂AG₃T₂AG₃. The latter can form parallel (x-ray structure)⁴⁷ and antiparallel (NMR structure)⁶² quadruplexes, both of which were used as templates for the simulations. The effects of systematically changing the loops from T to T₆ were studied using a combination of theoretical, and experimental methods. UV melting and circular dichroism (CD) data were obtained by Julian Huppert, and the materials and methods are detailed by Hazel *et al*¹⁴⁶.

Table 4.1: MD simulations using the three monomeric quadruplex templates. The length of each simulation is indicated in nanoseconds. Modified loops are highlighted in magenta.

Abbreviation	Sequence	Parallel	Antiparallel	Mixed
				
T ₂ A-G ₃	AG ₃ T ₂ AG ₃ T ₂ AG ₃ T ₂ AG ₃	6	6	
T ₂ A-G ₃ T	AG ₃ T ₂ AG ₃ T ₂ G ₃ T ₂ AG ₃	4	4	4
T ₂ A-G ₃ T ₂	AG ₃ T ₂ AG ₃ T ₂ G ₃ T ₂ AG ₃	4	4	4
T ₂ A-G ₃ T ₃	AG ₃ T ₂ AG ₃ T ₃ G ₃ T ₂ AG ₃	4	4	
T ₂ A-G ₃ T ₄	AG ₃ T ₂ AG ₃ T ₄ G ₃ T ₂ AG ₃	4	4	
T ₂ A-G ₃ T ₅	AG ₃ T ₂ AG ₃ T ₅ G ₃ T ₂ AG ₃	4	4	
T ₂ A-G ₃ T ₆	AG ₃ T ₂ AG ₃ T ₆ G ₃ T ₂ AG ₃	4	4	
G ₃ T	AG ₃ T ₂ G ₃ T ₂ G ₃ T ₂ G ₃	4		4
G ₃ T ₂	AG ₃ T ₂ G ₃ T ₂ G ₃ T ₂ G ₃	4	4	4

4.1 Computational methods

The templates used for the simulations were derived from the AG₃T₂AG₃T₂AG₃T₂AG₃ parallel crystal structure and antiparallel NMR structure (PDB codes 1KF1 and 143D, respectively). A mixed parallel/antiparallel template was also used, derived from the T₂G₄T₂G₄T₂G₄T₂G₄ solution structure (PDB code 186D). All three templates have three G-quartet stacks, and only the loop regions were modified. Structure manipulations were carried out in Insight II, as described in Section 3.1.1. Simulated annealing runs were used to generate loop conformations for each loop length, and the most energetically favourable conformation (as calculated by Discover, in Insight) was further subjected to 4 ns molecular dynamics using Amber. T₃ loops were generated by simply mutating the A residue from the native T₂A loops into T. All simulations followed the experimental protocol described in Section 3.1.1. Table 4.1 summarises the simulations which were carried out. For comparison purposes, the parallel and antiparallel templates were also simulated for 6 ns with the native T₂A loops. Single loop variants, in which only the single loop was modified, and multiple loop variants, in which all three loops were modified, were considered.

Simulated annealing and clustering methods were found to be less useful as the loop length increased. The increasing degrees of freedom in loops of four residues and above

lead to the generation of many different loop conformations. The number of simulated annealing runs carried out was found to be insufficient to generate highly populated clusters. For example, out of 100 parallel T_4 loop conformations generated, the most populated cluster when using an rms deviation threshold of 1.5 Å contained only seven structures. These structures are shown overlapped in Figures 4.1(a) and (b). There are significant deviations in loop conformation with a 1.5 Å threshold, however this reflects the overall disparity between all the conformations obtained. The low cluster populations meant that structure selection based on occurrence frequency was unrealistic. Instead, suitable loop starting conformations were selected depending on their Discover-calculated energies, and on the presence of loop-loop stabilising interactions, such as stacking and hydrogen bonding. Figure 4.1(c) shows a T_4 antiparallel loop cluster, obtained using a 1.5 Å threshold. The cluster only contained two structures, but three loop residues were involved in stacking interactions with the G-quartets, and this conformation was selected for further MD simulations.

Currently, no experimental structures have been determined in which there are three G-quartets with three lateral loops. In order to use only experimental templates, to reduce uncertainty due to the models used, the $T_2G_4T_2G_4T_2G_4T_2G_4$ structure was used as a template for quadruplexes with lateral loops (PDB code 186D). This structure contains two lateral loops and a third parallel loop, with three G-quartet stacks. The loops were modified as described above.

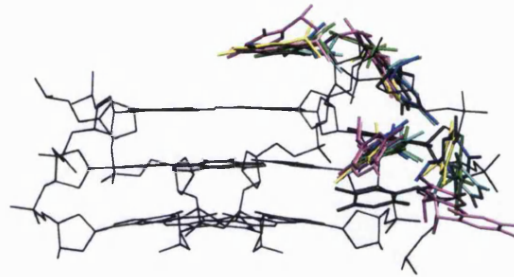
MM-PBSA calculations were carried out as described in Section 3.1.1. All calculations included two K^+ ions within the quadruplex channel. When fewer than two K^+ ions were present in the channel, the closest solution K^+ ion was included instead. Calculations of the electrostatic component of the solvation free energy, using Delphi II, were carried out using both BONDI and PARSE¹⁰⁸ radii, for comparison purposes.

4.2 Simulation results

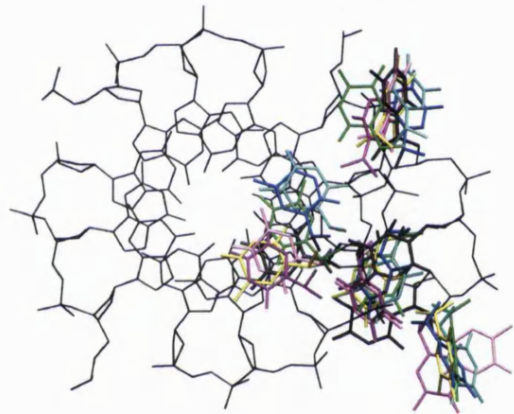
4.2.1 Single loop variants

Parallel quadruplexes

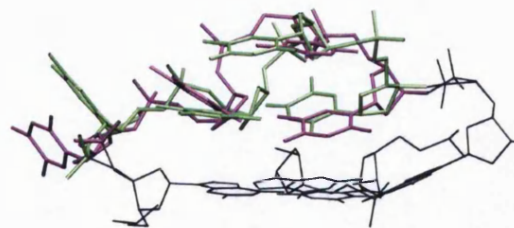
Figure 4.2 shows the parallel loop structures obtained during MD simulations. All T to T_6 loops gave rise to stable quadruplexes over the 4 ns simulations. No G-quartet distortion was observed, indicating that no loop length prohibits a parallel quadruplex fold. The parallel loops were in general much more flexible than the G-quartets (Table 4.2). The only exception was the single T loop, shown in Figure 4.2(a), which formed a stable



(a) Parallel quadruplex, side view



(b) Parallel quadruplex, top view



(c) Antiparallel quadruplex

Figure 4.1: (a) and (b) Overlap of the seven structures in the most populated cluster (1.5 Å threshold) obtained for the parallel T_4 loop using simulated annealing. The loop backbones of all but one loop have been omitted for clarity, leaving only the bases. The G-quartets are shown in black. (c) Overlap of two T_4 loop structures in an antiparallel cluster. Only the upper G-quartet is shown in black.

hydrogen bond with a G residue, and exhibited little fluctuation.

In the 6 ns control simulation of the parallel X-ray structure, rearrangements were observed in all three loops. In the crystal structure, the loop residues are in a compact, stacked conformation. During the simulation, the residues reached further into the solvent, forming more extended loop structures, as shown in Figure 4.3(b). Although the X-ray loop conformation was lost within the first few hundred picoseconds in the T₂A-G₃ simulation, stacking interactions between loop residues formed throughout the dynamics. The A residues generally stacked with either of the T residues. The T₂A loops in the single loop variant simulations also showed the same flexibility, with rms deviations ranging from 2.4 to 6.1 Å, with respect to the starting structure (Table 4.2).

The modified parallel loops also displayed a range of rms deviations, as shown in Table 4.2. The loop rms deviations generally increased with loop length, however this also depended on the loop conformation. The T₄ loop residues formed hydrogen bonding interactions with the G residue N2 atoms, and these were maintained during the dynamics. This is reflected in the relatively low rms deviation of the T₄ loop when compared to the shorter T₃ loop (2.3 Å and 3.4 Å, respectively). The T₅ and T₆ loops also retained some of the initial hydrogen bonds between T and G residues during the simulations. However, the latter had more residues not involved in hydrogen bonding, and these were more flexible, explaining the increasing rms deviations for the T₄, T₅ and T₆ loops (2.3, 3.6 and 5.6 Å, respectively).

The parallel quadruplex structures in this study were found to be relatively unaffected by the different loop lengths and conformations. Rms deviations of the G-stems were largely independent of the loop fluctuations. For example the T₆ loop in the T₂A-G₃T₆ structure had a high rms deviation during the simulation, but the G-quartets were amongst the most stable (rms deviations of 5.6 and 1.4 Å, respectively). The high flexibility of parallel loops during the simulations indicates that finding particularly favourable conformations would be more difficult with increasing loop length. It is likely that the longer loops can adopt many energetically closely related conformations. The simulations do show, however, that parallel G-quadruplexes are stable with loops up to six nucleotides in length, for the timescale of the simulations.

Antiparallel quadruplexes

Average structures of the antiparallel single loop variants are shown in Figure 4.4. Loops of lengths three and greater formed stable antiparallel quadruplexes. Single T loops were unable to span the diagonal of the G-quartets, causing distortion to the latter, as shown in Figure 4.4(a). This was also reflected in a higher than average G-quartet rms

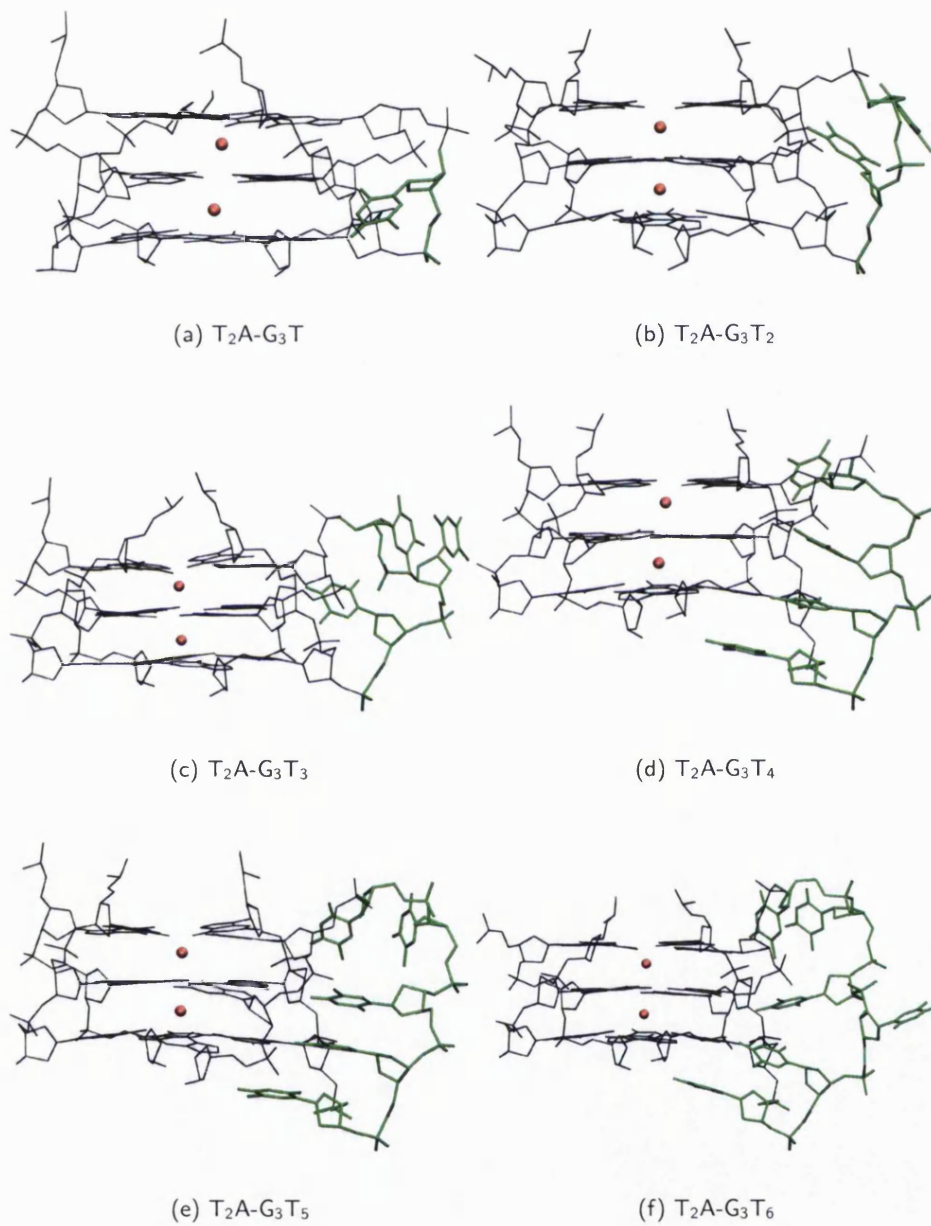


Figure 4.2: MD structures of the parallel single loop variants. The structures were averaged over the final 2 to 4 ns of simulation. G-quartets are in black, loops in green and K⁺ ions in red. The T₂A loops have been omitted for clarity.

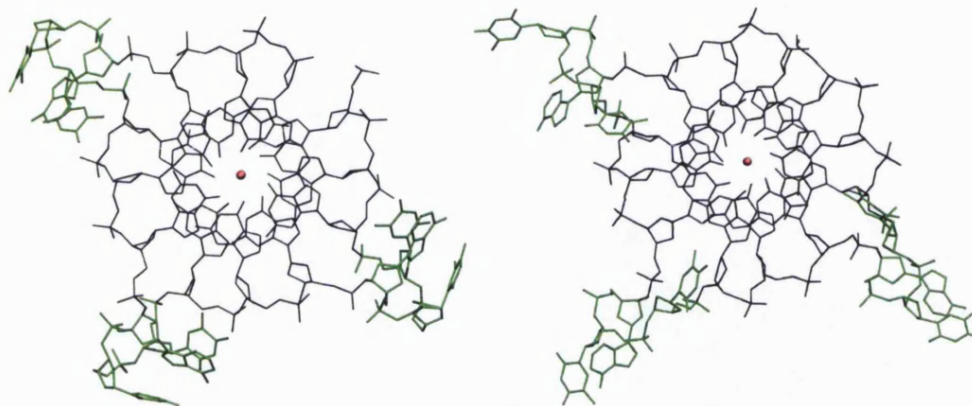
Table 4.2: Rms deviations (\AA) of the G-quartet and loop residues in the single loop variants. The modified loops are highlighted in the colours used in Figures 4.2, 4.4 and 4.5. All rms deviations were calculated after fitting of the G-quartets. Table reproduced from¹⁴⁶.

Simulation	G-quartet	Loop 1	Loop 2	Loop 3	Loop 2 (2 ns average)*
Parallel					
T ₂ A-G ₃ [†]	1.8	3.1	3.7	2.8	2.3 (2.0, 2.3) [‡]
T ₂ A-G ₃ T	1.3	6.1	1.2	3.6	0.5
T ₂ A-G ₃ T ₂	1.4	3.2	2.3	3.3	1.2
T ₂ A-G ₃ T ₃	1.4	3.2	3.5	3.7	2.2
T ₂ A-G ₃ T ₄	1.4	3.3	2.3	3.8	0.9
T ₂ A-G ₃ T ₅	1.5	2.4	3.6	3.1	2.0
T ₂ A-G ₃ T ₆	1.4	3.3	5.6	3.6	1.8
Antiparallel					
T ₂ A-G ₃ [†]	1.7	4.0	2.7	2.6	1.3 (0.8, 1.4) [‡]
T ₂ A-G ₃ T	2.5	2.5	1.9	1.3	0.9
T ₂ A-G ₃ T ₂	1.4	2.3	1.7	1.2	0.9
T ₂ A-G ₃ T ₃	1.4	4.1	2.7	2.0	0.9
T ₂ A-G ₃ T ₄	1.5	4.2	2.2	1.9	1.1
T ₂ A-G ₃ T ₅	1.5	3.7	2.3	1.8	0.9
T ₂ A-G ₃ T ₆	1.6	4.3	4.5	2.3	2.0
Mixed parallel/antiparallel					
T ₂ A-G ₃ T	1.1	4.7	1.8	4.1	0.7
T ₂ A-G ₃ T ₂	1.2	1.5	1.5	4.2	1.2

*Rms deviations calculated with respect to the 2 to 4 ns average structure. (Values over the final 2 to 4 ns only).

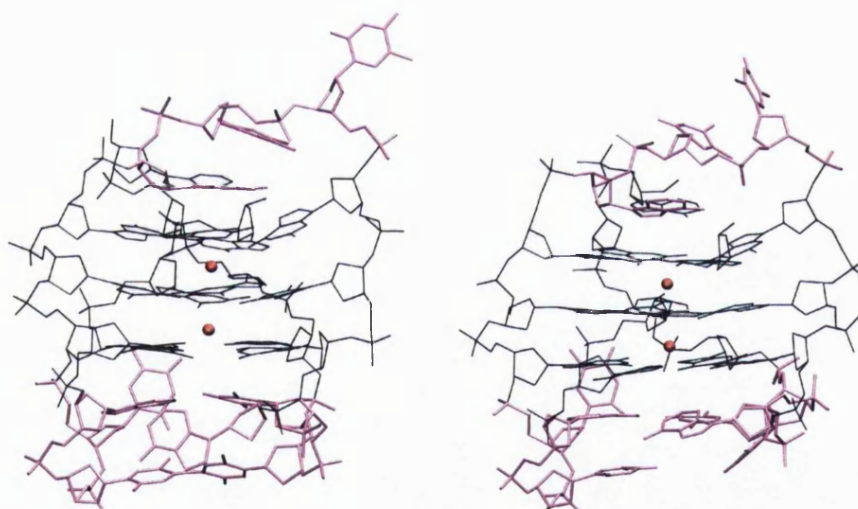
[†]6 ns simulation

[‡]loop 1 and loop 3 values are indicated in brackets. The average was calculated over the last 3 ns of MD



(a) T₂A-G₃ X-ray structure

(b) T₂A-G₃ after 6 ns MD



(c) T₂A-G₃ NMR structure

(d) T₂A-G₃ after 6 ns MD

Figure 4.3: T₂A-G₃ X-ray and NMR structures, and final 6 ns structures. The NMR structure was determined in Na⁺ ion containing solution, but for consistency with the other simulations, K⁺ ions were included between the G-quartet stacks, as found in the X-ray structure. (a) and (b), top view of the parallel structures, (c) and (d) side view of the antiparallel structures.

deviation, shown in Table 4.2. The diagonal T₂ loop was under some strain, causing fluctuations in the G-quartet immediately below the loop. However, the longer T₂A lateral loops compensated for this strain, and the bottom G-quartets remained planar, as shown in Figure 4.4(b). Mixed parallel/antiparallel quadruplexes with a lateral, rather than diagonal, central loop were simulated for loops of one and two nucleotide lengths (Figure 4.5). These structures were stable over the MD simulations, and the T and T₂ loops had relatively low rms deviations (Table 4.2).

At equal length, the antiparallel diagonal loops were more stable than the parallel loops during the simulations (Table 4.2). As this was found for all loop lengths considered, it is less likely to be due only to the particular starting structures selected. This was also the case for the control experimental structure simulations, where the diagonal T₂A loop had an rms deviation of 2.7 Å over the 6 ns simulation, whereas the parallel T₂A loops had rms deviations of between 2.8 and 3.7 Å. As in the parallel quadruplex simulations, the modified diagonal loops did not affect the two lateral loops in the antiparallel structures. Loop 1 in all the antiparallel quadruplexes had a high initial rms deviation, but stabilised into a conformation which was similar in all the quadruplexes simulated. This is also similar to the conformation adopted by the lateral loops during the simulation of the unmodified NMR structure, shown in Figure 4.3(d), where three residues form a triplet stacked with the G-quartets. The modified loops had rms deviations which depended more on the interactions formed between the loop residues than on the loop length. Only the T₆ diagonal loop had a much higher rms deviation than the other loops.

4.2.2 Multiple loop variants

Quadruplexes in which all three loops were modified were simulated for T and T₂ loops. Simulations using the three templates were carried out, except for G₃T, for which only the parallel and mixed parallel/antiparallel templates were used. The antiparallel G₃T₂ simulation resulted in G-quartet distortion, and a simulation of the antiparallel template with shorter T loops was therefore not considered necessary.

Figure 4.6 shows the average parallel, antiparallel and mixed parallel/antiparallel structures for one- and two-nucleotide loops. The G₃T parallel structure was stable over the dynamics, as shown in Figure 4.6(a). The loop rms deviations were relatively high (Table 4.3), however this did not affect the G-quartet structure. On the other hand, the G₃T mixed parallel/antiparallel quadruplex was unstable during the simulation. Figure 4.6(b) shows that the G-quartets were severely distorted, and hydrogen bonding between G residues was lost. Only one K⁺ ion remained within the quadruplex chan-

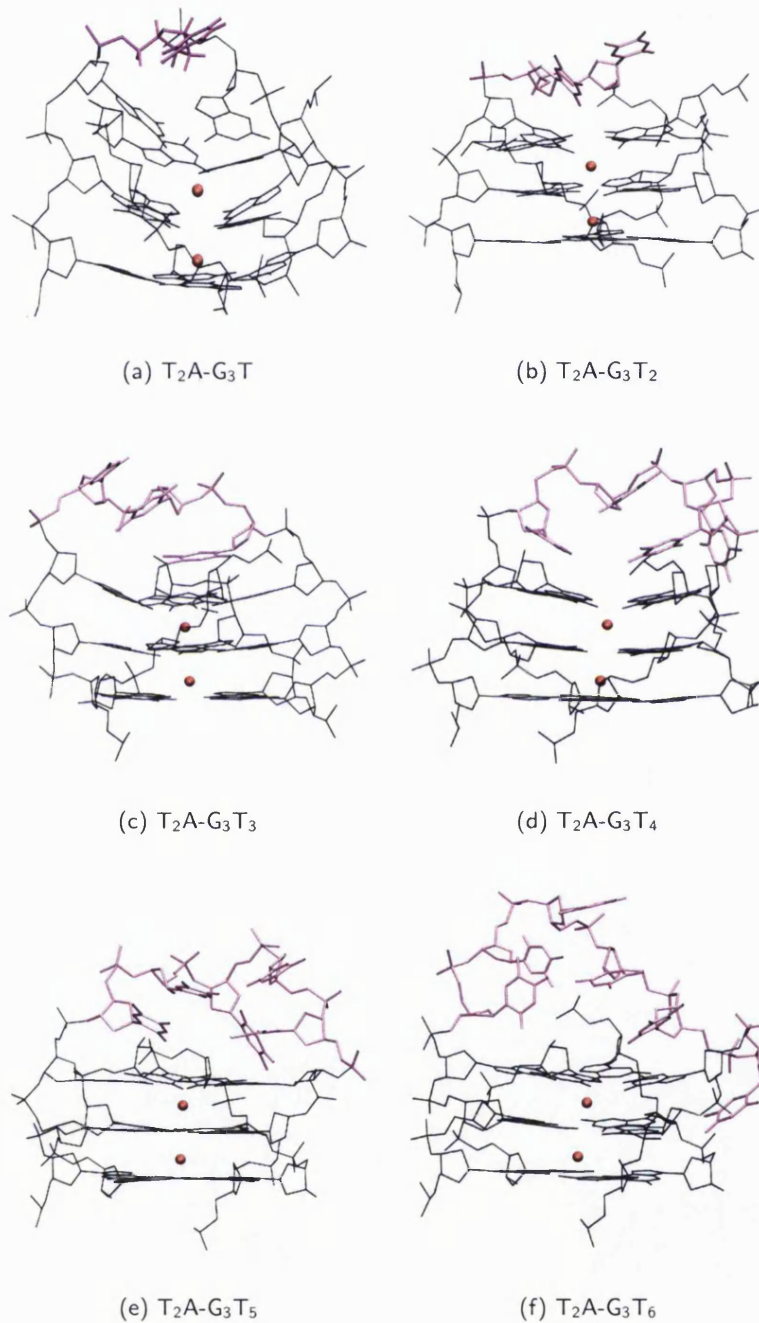


Figure 4.4: MD structures of the antiparallel single loop variants. The structures were averaged over the final 2 to 4 ns of simulation. G-quartets are in black, diagonal loops in purple and K⁺ ions in red. The T₂A loops have been omitted for clarity.

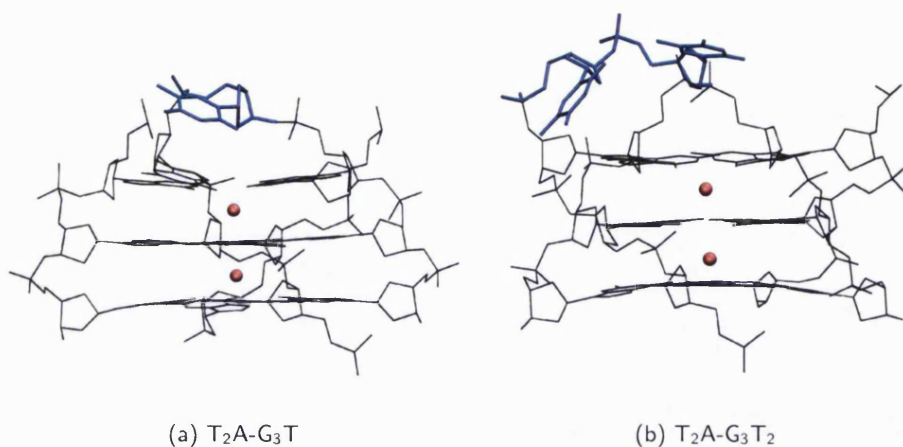


Figure 4.5: MD structures of the mixed parallel/antiparallel single loop variants. The structures were averaged over the final 2 to 4 ns of simulation. G-quartets are in black, lateral loops in blue and K⁺ ions in red. The T₂A loops have been omitted for clarity.

nel, the other moving out into the surrounding solution. As the three parallel T loop quadruplex was stable, the distortion is most likely due to the strain caused by the very short single residue lateral loops. These simulations suggest that G₃T is more likely to form a parallel, rather than antiparallel, structure in solution.

G₃T₂ was stable using both the parallel and the mixed parallel/antiparallel templates (Figures 4.6(c) and (d)). On the other hand, the antiparallel G₃T₂ quadruplex with a diagonal and two lateral loops was unstable, as shown by the G-quartet distortion (Figure 4.6(e)). This was much more pronounced than for the T₂A-G₃T₂ antiparallel simulation, in which the two longer lateral loops compensated in part for the strained diagonal loop. This suggests that G₃T₂ is as likely to form parallel or antiparallel structures in solution, as long as the latter have lateral, rather than diagonal loops. The rms deviations in Table 4.3 show that the loops were again flexible during the simulations.

4.3 Experimental data

UV melting curves and CD spectra of the single and multiple loop variants were obtained by Julian Huppert¹⁴⁶. The CD spectra are shown in Figure 4.7. Quadruplexes melt with a significant hyperchromic shift at 295 nm, and melting temperatures can be obtained by fitting to the resulting absorbance curves⁶⁹. UV melting temperatures for all the sequences are shown in Table 4.4. The sequences studied experimentally differ from the

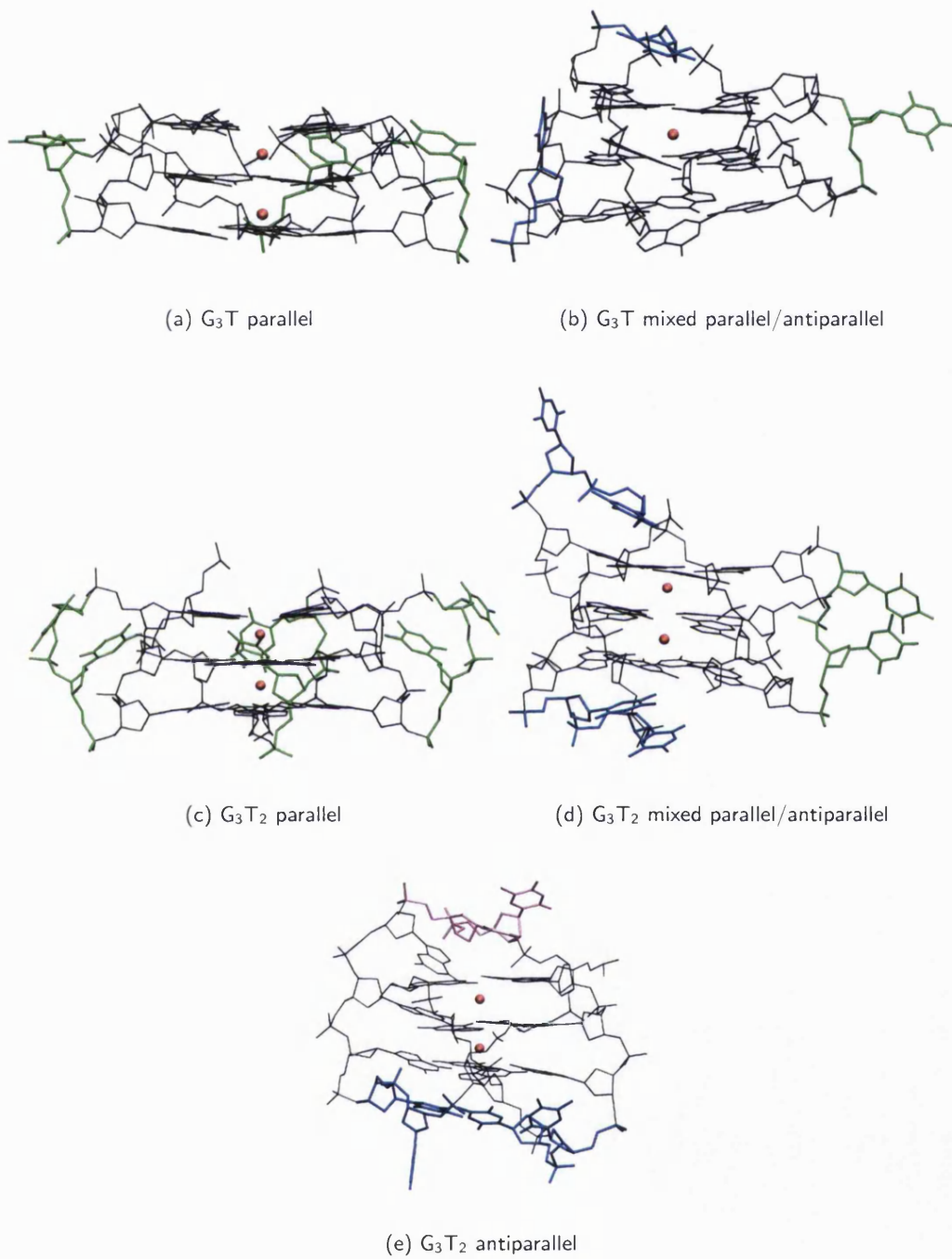


Figure 4.6: MD structures of the multiple loop variants with the parallel, antiparallel and mixed parallel/antiparallel templates. The structures were averaged over the final 2 to 4 ns of simulation. G-quartets are black, parallel loops are green, lateral loops are blue, diagonal loops purple and K⁺ ions are red.

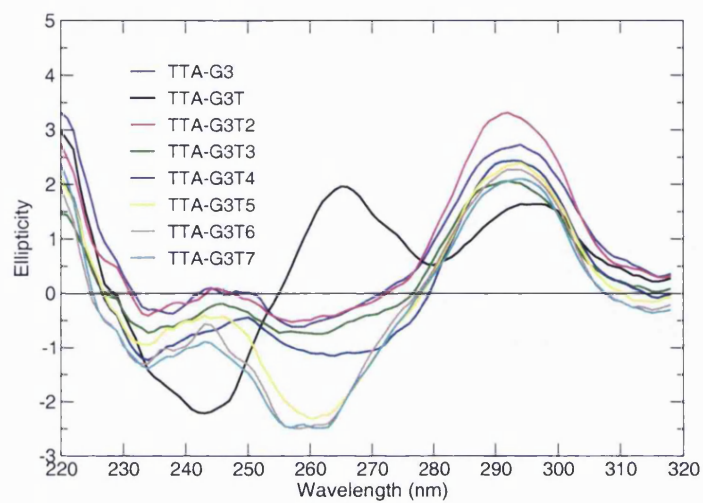
Table 4.3: Rms deviations (Å) for the G-quartets and loop residues in the multiple loop variants. The modified loops are highlighted in the colours used in Figure 4.6. All rms deviations were calculated after fitting of the G-quartets. Table reproduced in part from¹⁴⁶.

Simulation	G-quartet	Loop 1	Loop 2	Loop 3
Parallel				
G ₃ T	1.3	3.6	3.0	1.6
G ₃ T ₂	1.2	2.9	3.2	3.5
Antiparallel				
G ₃ T ₂	1.9	4.8	3.4	2.8
Mixed parallel/antiparallel				
G ₃ T	1.8	3.7	2.0	4.4
G ₃ T ₂	1.5	2.2	3.9	5.5

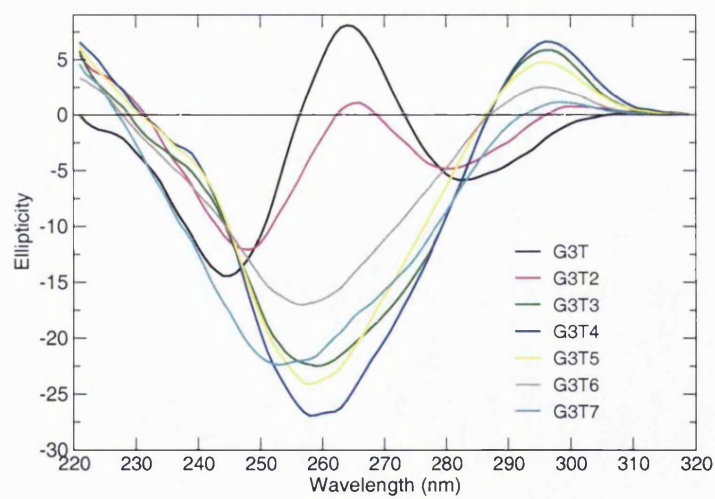
sequences modelled only in the 3' terminal T residue. The latter facilitates experimental measurements, but is not present in either the crystal or NMR structures used as templates, and was therefore not included in the simulations. For the simulations, the 5' terminal T residue in the multiple loop variants was replaced with A, for consistency with the single loop variant simulations.

UV melting data and CD spectra both confirm that all sequences form G-quadruplex structures in 100 mM K⁺ solutions. The importance of loop length is shown by the wide range of melting temperatures obtained for the various sequences. The most stable quadruplex, G₃T, has a melting temperature of over 80 °C in a 20 mM K⁺ solution. Increasing the loop length reduces the stability of the multiple loop variants compared to the unfolded state. The G₃T₇ sequence (which was not included in the simulations) did show quadruplex formation using CD spectroscopy, but is too unstable to allow the T_m to be measured. The single loop variant melting temperatures show less dependence on loop length. The maximum value is obtained for T₂A-G₃T₃ (T_m = 68 °C), while T₂A-G₃T₇ is the least stable (T_m = 56 °C). The change of a single loop within the sequences has much less influence on stability than the change of all three loops. Hence, a single T₇ loop can be accommodated with little change in quadruplex stability, whereas three T₇ loops are very destabilising.

CD spectra of G-quadruplexes are commonly used to determine whether these form parallel or antiparallel structures (Section 1.4.2). The CD spectrum of G₃T shows a parallel quadruplex characteristic maximum around 265 nm, and minimum around 240 nm. Sequences G₃T₃ to G₃T₇ have characteristic antiparallel quadruplex curves, with



(a) Single loop variants



(b) Multiple loop variants

Figure 4.7: CD spectra for the (a) single and (b) multiple loop variants. Figures reproduced from¹⁴⁶.

Table 4.4: UV melting temperatures ($^{\circ}\text{C}$) for the single and multiple loop variants in this study. Samples were measured in 100 mM K^+ containing solutions. Modified loops are highlighted in magenta. Table adapted from¹⁴⁶.

Abbreviation	Sequence	T_m
$\text{T}_2\text{A-G}_3$	$\text{AG}_3\text{T}_2\text{AG}_3\text{T}_2\text{AG}_3\text{T}_2\text{AG}_3\text{T}$	63
$\text{T}_2\text{A-G}_3\text{T}$	$\text{AG}_3\text{T}_2\text{AG}_3\text{T}_3\text{G}_3\text{T}_2\text{AG}_3\text{T}$	64
$\text{T}_2\text{A-G}_3\text{T}_2$	$\text{AG}_3\text{T}_2\text{AG}_3\text{T}_2\text{G}_3\text{T}_2\text{AG}_3\text{T}$	66
$\text{T}_2\text{A-G}_3\text{T}_3$	$\text{AG}_3\text{T}_2\text{AG}_3\text{T}_3\text{G}_3\text{T}_2\text{AG}_3\text{T}$	68
$\text{T}_2\text{A-G}_3\text{T}_4$	$\text{AG}_3\text{T}_2\text{AG}_3\text{T}_4\text{G}_3\text{T}_2\text{AG}_3\text{T}$	61
$\text{T}_2\text{A-G}_3\text{T}_5$	$\text{AG}_3\text{T}_2\text{AG}_3\text{T}_5\text{G}_3\text{T}_2\text{AG}_3\text{T}$	61
$\text{T}_2\text{A-G}_3\text{T}_6$	$\text{AG}_3\text{T}_2\text{AG}_3\text{T}_6\text{G}_3\text{T}_2\text{AG}_3\text{T}$	59
$\text{T}_2\text{A-G}_3\text{T}_7$	$\text{AG}_3\text{T}_2\text{AG}_3\text{T}_7\text{G}_3\text{T}_2\text{AG}_3\text{T}$	56
G_3T	$\text{TG}_3\text{TG}_3\text{TG}_3\text{TG}_3\text{T}$	too stable
G_3T_2	$\text{TG}_3\text{T}_2\text{G}_3\text{T}_2\text{G}_3\text{T}_2\text{G}_3\text{T}$	71
G_3T_3	$\text{TG}_3\text{T}_3\text{G}_3\text{T}_3\text{G}_3\text{T}_3\text{G}_3\text{T}$	58
G_3T_4	$\text{TG}_3\text{T}_4\text{G}_3\text{T}_4\text{G}_3\text{T}_4\text{G}_3\text{T}$	48
G_3T_5	$\text{TG}_3\text{T}_5\text{G}_3\text{T}_5\text{G}_3\text{T}_5\text{G}_3\text{T}$	34
G_3T_6	$\text{TG}_3\text{T}_6\text{G}_3\text{T}_6\text{G}_3\text{T}_6\text{G}_3\text{T}$	20
G_3T_7	$\text{TG}_3\text{T}_7\text{G}_3\text{T}_7\text{G}_3\text{T}_7\text{G}_3\text{T}$	too unstable

maxima around 295 nm, and minima around 260 nm. The G_3T_2 sequence could form a mixture of both parallel and antiparallel quadruplexes, as the curve displays a minimum around 240 nm, but also possible small maxima around both 265 and 295 nm. The $T_2A-G_3T_2$ to $T_2A-G_3T_7$ sequences display antiparallel quadruplex-type CD spectra. T_2A-G_3T has maxima both around 265 and 295 nm, which could also reflect the formation of both parallel and antiparallel structures.

The experimental results confirm that loop length does have an influence on the quadruplex structure formed, as well as on stability. Even in the case of the single loop variants, where the stability of the quadruplexes was not significantly affected, CD spectra suggest that different structures are formed by the different sequences.

4.4 Comparison of simulations and experiment

4.4.1 CD spectroscopy results

G_3T is the only sequence for which CD data indicates the formation of a parallel structure only. The MD simulations also suggested this. The mixed parallel/antiparallel quadruplex with three single T residue loops distorted during the course of the simulation, whereas the parallel quadruplex was stable (Figure 4.6). G_3T_2 was able to form both parallel and mixed parallel/antiparallel quadruplexes according to the simulations (Figure 4.6), and has a CD spectrum which may be due to several structures coexisting in solution. The same was found for T_2A-G_3T , where both parallel and mixed parallel/antiparallel quadruplexes were stable over the 4 ns simulations, whereas the antiparallel quadruplex was distorted (Figure 4.4(a)). The CD spectra for both these sequences are similar in that peaks characteristic of both parallel and antiparallel structures occur.

Sequences with all other loop lengths were able to form stable parallel and antiparallel quadruplexes during the simulations. On the other hand, CD spectra suggest that only antiparallel quadruplexes are forming in solution. As both structural types were stable during simulations of the longer loop sequences, other methods are required to differentiate them. For this purpose, free energy calculations were carried out to further characterise the different quadruplexes.

4.4.2 MM-PBSA free energy calculations

MM-PBSA calculations were carried out in an attempt to compare sequences for which both parallel and antiparallel structures were stable during the simulations. The abso-

lute free energy of each quadruplex was estimated, and $\Delta G_{\text{antiparallel}-\text{parallel}}$ values were calculated. Values calculated using BOND radii are shown in Table 4.5. Figure 4.8 shows that for the parallel and antiparallel single loop variant simulations, most values were stabilised over the final 2 ns, except for the $T_2A-G_3T_5$ parallel simulation. The values are necessarily approximate, due to the fact that no attempt was made to obtain the most favourable loop conformation for each of the loop lengths simulated. However, in cases where the structure formed is limited by strain caused by short loop lengths, the specific loop conformation is expected to be less important. As the free energies calculated are not free energy differences between folded and unfolded conformations, only sequences with the same number of atoms can be compared. This means that the effect of loop length on quadruplex stability cannot be compared using the MM-PBSA calculations. However, different conformations of the same sequence can be compared.

Overall, the antiparallel quadruplexes were enthalpically favoured (although ΔG in Table 4.5 is not strictly ΔH , as the solvent entropic contribution is implicitly included). On the other hand, the solute entropic contribution favoured the parallel quadruplexes. In most cases, the enthalpic difference between antiparallel and parallel quadruplexes was the dominant term, so that ΔG_{total} favoured the antiparallel quadruplexes. However, both enthalpic and entropic contributions were very similar in some cases, leading to small ΔG_{total} values of 2 or 3 kcal.mol⁻¹, which is within the error of MM-PBSA calculations. Due to the similar relative contributions of ΔH and $T\Delta S$, it is important to include some measure of the solute entropy, although it is very approximative.

T_2A-G_3T and G_3T_2 were the only two quadruplexes for which the parallel structures were enthalpically favoured compared to the antiparallel structures. Both T_2A-G_3T and G_3T_2 antiparallel structures were heavily distorted during the dynamics, whereas the parallel and mixed parallel/antiparallel structures were not (Figures 4.4(a) and 4.6(e)). The free energies calculated did not, however, always reflect the stabilities observed during the simulations. For example, the G_3T mixed parallel/antiparallel structure was unstable during the simulation, however ΔG_{total} favours the latter over the parallel structure by 9 kcal.mol⁻¹. This is despite the fact that the mixed quadruplex had only two hydrogen bonded G-quartets and a single K⁺ ion remaining within the central channel. This does cast some doubt as to the ability of MM-PBSA calculations to estimate the relative stabilities of quadruplex structures.

For loop lengths of three residues and more, the free energy differences were largely in favour of the antiparallel quadruplexes, which is in accordance with the CD spectra shown in Figure 4.7. This was not the case for the $T_2A-G_3T_5$ structure, where the positive $\Delta G_{\text{antiparallel}-\text{parallel}}$ value is probably due to the particular loop conformations

Table 4.5: MM-PBSA free energy calculations over the final 2 to 4 ns of each simulation. Two K^+ ions were included in all the calculations, and BONDI radii were used. All values are in kcal.mol^{-1} and $T = 300\text{K}$. The mixed parallel/antiparallel simulations are indicated with "mixed". In that case the antiparallel column refers to the mixed structures. G values do not include the solute entropic contribution, however, G_{total} includes the latter. Standard errors of the mean are in brackets.

Simulation	Antiparallel/mixed		Parallel		Antiparallel-parallel		
	G	-TS	G	-TS	ΔG	$-T\Delta S$	ΔG_{total}
$T_2A-G_3^*$	-4353 (4)	-579 (1)	-4329 (4)	-589 (2)	-24	+10	-14
T_2A-G_3T	-4021 (5)	-527 (1)	-4029 (4)	-536 (2)	+8	+9	+17
T_2A-G_3T (mixed)	-4030 (4)	-532 (1)			-1	+4	+3
$T_2A-G_3T_2$	-4169 (4)	-553 (1)	-4157 (4)	-559 (2)	-12	+6	-6
$T_2A-G_3T_2$ (mixed)	-4173 (4)	-554 (1)			-16	+5	-11
$T_2A-G_3T_3$	-4305 (5)	-578 (2)	-4279 (4)	-587 (2)	-26	+9	-17
$T_2A-G_3T_4$	-4448 (4)	-601 (1)	-4422 (5)	-608 (2)	-26	+7	-19
$T_2A-G_3T_5$	-4568 (4)	-626 (1)	-4561 (4)	-635 (2)	-7	+9	+2
$T_2A-G_3T_6$	-4707 (5)	-649 (1)	-4667 (4)	-662 (2)	-40	+13	-27
G_3T (mixed)	-3388 (4)	-431 (1)	-3373 (4)	-437 (1)	-15	+6	-9
G_3T_2	-3795 (4)	-507 (2)	-3806 (4)	-514 (1)	+11	+7	+18
G_3T_2 (mixed)	-3813 (4)	-509 (1)			-7	+5	-2

*Values were calculated over the final 2 to 6 ns of simulation

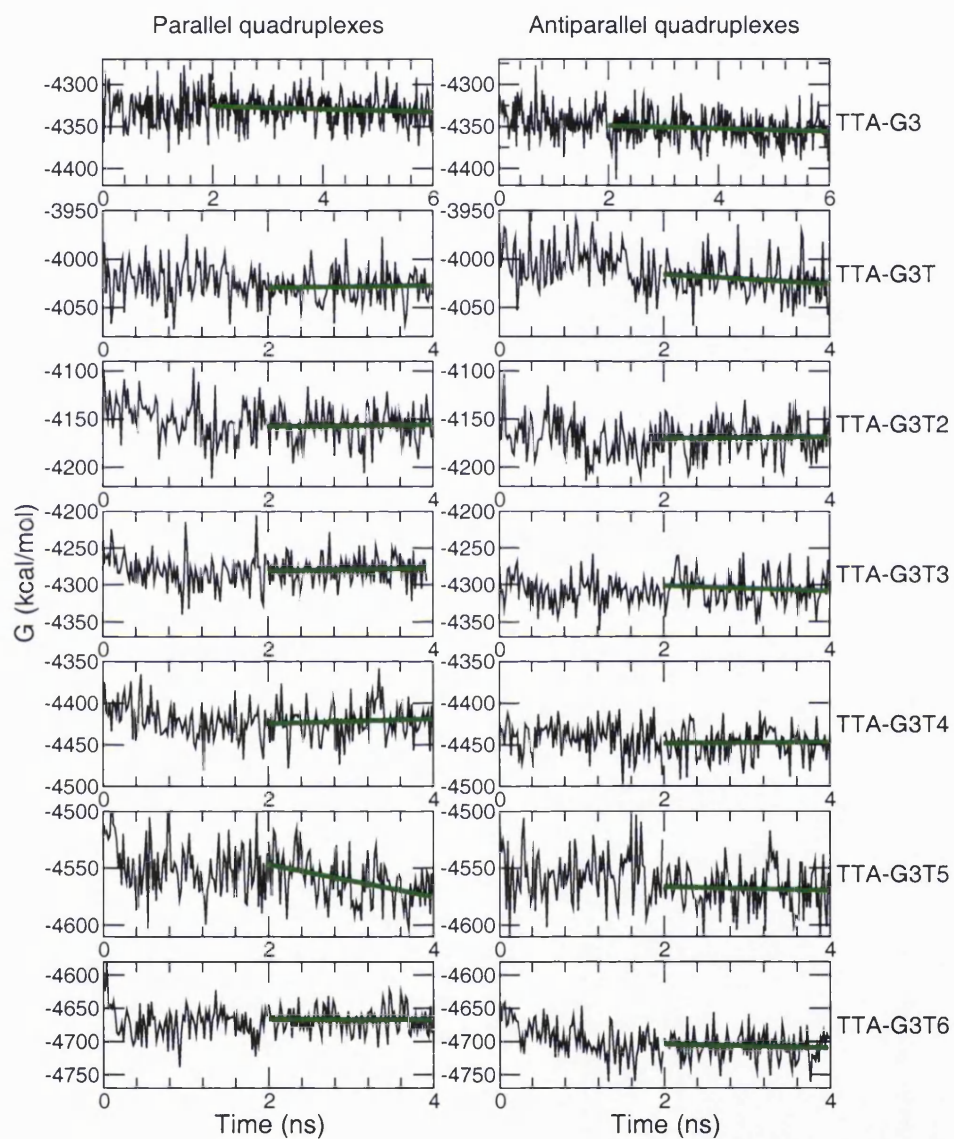


Figure 4.8: Variation of the free energy of the parallel and antiparallel single loop variants over the 4 ns simulations (6 ns for the native quadruplexes). The linear regression to the data over the final 2 to 4 ns is shown as a green line (2 to 6 ns for the T_2A-G_3 simulations).

which were chosen in the simulations. Moreover, Figure 4.8 shows that the free energy of the parallel $T_2A-G_3T_5$ structure was still decreasing after 4 ns of simulation, and the structure has clearly not yet equilibrated. This does show that despite the more favourable antiparallel structures, some parallel structures which have similar energies can also be found.

In order to study the effect of different parameters on the MM-PBSA calculations, the $G_{\text{solvation}}$ term was also calculated using PARSE radii¹⁰⁸ (Table 4.6). These were used in two recent G-quadruplex studies using MM-PBSA^{93,94}. Absolute G_{polar} values calculated with Delphi using PARSE radii are on average $200 \text{ kcal.mol}^{-1}$ less favourable than those calculated with BONDI radii. For most quadruplexes, PARSE radii favour the antiparallel conformations less than BONDI radii. The only exception to this is the T_2A-G_3T simulation, for which PARSE radii favoured the mixed parallel/antiparallel structure, whereas BONDI radii favoured the parallel structure. This probably reflects how close the free energies are, since CD spectra show peaks due to both parallel and antiparallel conformations in solution (Figure 4.7). The free energy difference between PARSE and BONDI radii calculations is not constant, and depends on the particular quadruplex structures. The largest difference is found for the $T_2A-G_3T_2$ structure, for which the enthalpic term calculated with PARSE radii favours the parallel, rather than antiparallel structure.

The different $G_{\text{solvation}}$ values are due to PARSE radii being on average smaller than BONDI radii. This has the effect of moving the dielectric boundary closer to the atoms, thus lowering the energy of polar atoms for PARSE radii (through stronger polarisation)¹⁴⁷. This effect is expected to be more important for atoms exposed to solvent, than for buried atoms. Parallel quadruplexes have larger solvent accessible surface areas than antiparallel structures, which in turn means that the former have a greater proportion of exposed polar atoms. The effect of using PARSE rather than BONDI radii is therefore more pronounced for the parallel structures, leading to generally less favourable $\Delta G_{\text{antiparallel-parallel}}$ values. Free energy differences between the parallel and antiparallel T_2A-G_3 structures are expected to be low, as both structures are thought to coexist in solution and readily interconvert¹⁴⁸. This provides some confidence in the energies calculated with PARSE, rather than BONDI radii, as $\Delta G_{\text{antiparallel-parallel}}$ is smaller for the former. The T_2A-G_3 quadruplexes were obtained directly from experimental structures, and hence do not suffer from errors due to possible bad initial loop conformations.

Table 4.6: MM-PBSA free energy calculation using PARSE radii (kcal.mol^{-1}). Two K^+ ions were included in all calculations, and values were averaged over the final 2 to 4 ns of simulation. $T\Delta S$ values are taken from Table 4.5. ΔG values do not include the solute entropic contribution, ΔG_{total} includes the latter.

Simulation	ΔG	ΔG_{total}
$\text{T}_2\text{A-G}_3^*$	-11	-1
$\text{T}_2\text{A-G}_3\text{T}$	+10	+19
$\text{T}_2\text{A-G}_3\text{T}$ (mixed)	-11	-7
$\text{T}_2\text{A-G}_3\text{T}_2$	+8	+14
$\text{T}_2\text{A-G}_3\text{T}_2$ (mixed)	-11	-6
$\text{T}_2\text{A-G}_3\text{T}_3$	-15	-6
$\text{T}_2\text{A-G}_3\text{T}_4$	-22	-15
$\text{T}_2\text{A-G}_3\text{T}_5$	-4	+5
$\text{T}_2\text{A-G}_3\text{T}_6$	-38	-25
G_3T (mixed)	-9	-3
G_3T_2	+19	+26
G_3T_2 (mixed)	-5	0

*Values were calculated over the final 2 to 6 ns of simulation

4.4.3 Importance of interactions within the loops

In general, free energy calculations showed that antiparallel quadruplexes were enthalpically favoured, whereas the entropic contribution favoured the parallel quadruplexes. The antiparallel loop residues were able to form many stabilising interactions, by stacking on the G-quartets and with each other, as shown by the examples in Figure 4.4. Hydrogen bonds stabilised the loops, which, at equal loop lengths, had lower rms deviations than the parallel loops (Table 4.2, page 110). Stacking interactions did occur between the residues in the parallel loops, but these were relatively short-lived. Parallel quadruplexes were entropically favoured due to greater loop flexibility, compared to the antiparallel structures.

Interactions between loop residues probably contribute to the preference for antiparallel compared to parallel structures, which is suggested by the CD spectra for loops of lengths two or three and above. During simulations of the antiparallel single loop variants, the two adjacent lateral T_2A loops consistently formed a hydrogen-bonded triplet, which stacked on the G-quartet plane. Such stabilising interactions were not formed in the parallel structures. Fox *et al* have shown that when replacing the loop residues with non-nucleosidic linkers, characteristic parallel conformation peaks are observed in

the CD spectra⁷⁷. The presence of loop residues appears to promote the formation of antiparallel quadruplexes, as shown by changes in the CD spectra for quadruplexes with nucleosidic loops (Figure 4.7). It also appears that, in the present spectra, parallel quadruplexes are only detected when sequences are unable to form stable antiparallel structures.

4.4.4 G₃T structure

G₃T is particularly interesting, as there are some uncertainties about the structure adopted by this sequence in solution. Although it exhibits a characteristic parallel quadruplex CD spectrum, NMR data suggests that the closely related G₃TG₃TG₃TG₃T sequence forms an antiparallel quadruplex in solution^{65,72,149}. The proposed structure contains two, rather than three G-quartet stacks, and is similar to the three-lateral loop thrombin binding aptamer structure shown in Table 1.4 (page 22). An antiparallel structure is also suggested by the smaller dependence on K⁺ ion concentration of a similar intramolecular quadruplex in which the T loops were replaced by non-nucleosidic linkers of an equivalent length⁷⁷. This suggests that fewer than three G-quartets are formed, by comparison with quadruplexes of longer linker lengths. The CD spectrum of the sequence with non-nucleosidic linkers was, however, also characteristic of parallel quadruplex formation. There is as yet no high resolution NMR or X-ray structure of G₃T, the NMR model having been derived from low resolution NMR data⁶⁵.

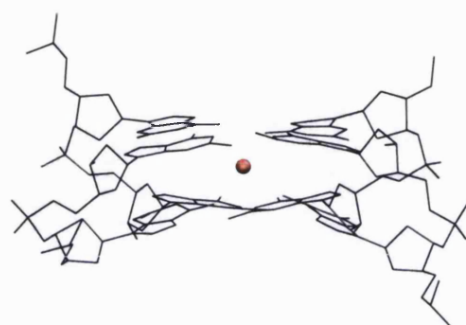
Figure 4.6(b) shows that during the simulation of the mixed parallel/antiparallel template with three T residue loops, only two G-quartets with one K⁺ ion were stable. The third G-quartet residues could, therefore, be counted as loop residues rather than belonging to the G-quartets. This did not affect the MM-PBSA estimated free energy of the quadruplex, which was found to be more stable than the parallel structure with three intact G-quartets (Tables 4.5 and 4.6). This type of structure with two G-quartets may, therefore, occur in solution, according to MD simulations. The antiparallel NMR structure suggested by Jing and Hogan for the G₃TG₃TG₃TG₃T sequence was simulated, using antiparallel G₂T₂G₂TGTG₂T₂G₂ thrombin binding aptamer structures as templates^{37,64}. The latter contains two G-quartets with three lateral loops, two T₂ and one TGT loop. A difficulty with this template is that no Na⁺ ion binding sites were identified in the crystal structure¹¹³. Ion positions have been suggested using a combination of NMR and MD simulations, revealing that there are favourable K⁺ binding sites within the G-quartets, and in the loop regions¹¹². Unfortunately, MD simulations are generally unable to reproduce stable ion binding within the loop regions⁹⁴. However, a stable 4 ns simulation of the thrombin binding aptamer was obtained with a single

K^+ ion located between the two G-quartet planes (Figure 4.9(a)). Even though the two G-quartets were slightly distorted, hydrogen bonding between the G residues was maintained during the simulations.

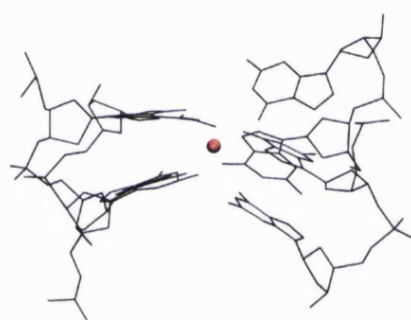
After mutation of the loop residues, antiparallel G_3T quadruplexes with three lateral loops and two G-quartets were simulated for 4 ns. Both the crystal (PDB code 1HUT) and NMR $G_2T_2G_2TGTG_2T_2G_2$ structures (PDB code 1C35) were used as templates, with three TG loops in the former and three GT loops in the latter, after modifications. However, the quadruplexes were unstable in both cases, with large distortions of the G-quartets occurring (Figures 4.9(c) and (d)). In order to verify that this was not due to unfavourable starting loop conformations, a simulation of the thrombin binding aptamer in which the TGT loop was changed to GT was carried out. This structure, with three two-nucleotide loops, was heavily distorted after 4 ns of dynamics, suggesting that the instabilities are due to the length of the loops, and not specifically unfavourable conformations (Figure 4.9(b)). The main difficulty when simulating the proposed G_3T antiparallel quadruplex is that three K^+ ions are expected to stabilise the structure⁷². However, during the simulations only a single ion was bound between the two G-quartet planes, and there were no stable binding sites within the loop region. This could explain why an unstable G-quadruplex is obtained during the simulations, corresponding to what Jing and Hogan refer to as a distorted “open” form, which would be the first of a two-step folding process in presence of K^+ ions⁶⁵.

CD spectra of both the G_3T sequence in this study and $G_3TG_3TG_3TG_3T$ show, however, no evidence of antiparallel quadruplex formation (Figure 4.7)⁷². This is in contrast to the thrombin binding aptamer structure, which displays a characteristic antiparallel quadruplex CD spectrum⁷⁰. Limitations of both CD spectroscopy and molecular dynamics techniques mean that differentiating between the structural forms is difficult (see Section 4.4.5). G_3T was the most stable quadruplex included in this study (Table 4.4). This stability is explained by Jing and Hogan as being due to the binding of two extra K^+ ions, presumably within the loop regions⁷². The observed stoichiometry of the $G_3TG_3TG_3TG_3T-K^+$ ion complex⁷² cannot be explained with the simulated parallel quadruplex, which contains only two K^+ binding sites, although binding sites within the loop regions cannot be discounted.

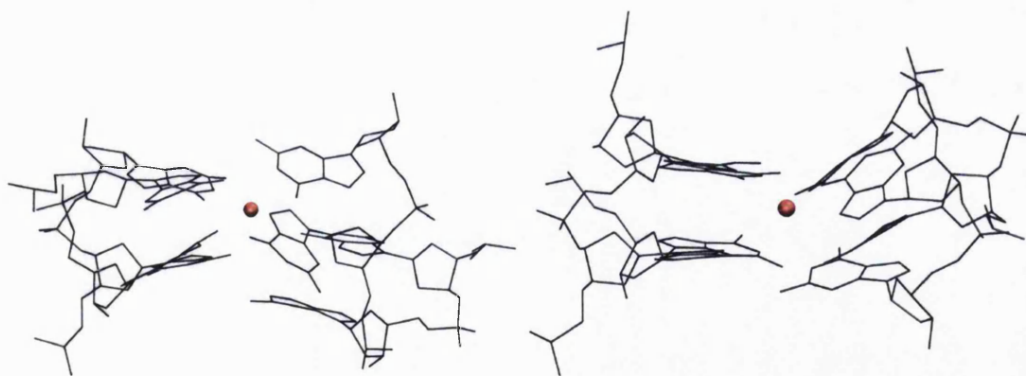
Some evidence suggesting the formation of parallel, rather than antiparallel G_3T quadruplexes comes from the NMR-determined topologies of $TGAG_3TG_3GAG_3TG_3GAA$ and $TGG_3AG_3T_5AG_3TG_3GA$ ⁶¹. The latter both form parallel quadruplexes, each with two single-nucleotide loops. The number of G-quartet stacks formed in each structure is maximised through the formation of parallel single-residue loops.



(a) $G_2T_2G_2TGTG_2T_2G_2$



(b) $G_2T_2G_2GTG_2T_2G_2$



(c) G_3T NMR structure

(d) G_3T X-ray structure

Figure 4.9: Final structures obtained after 4 ns simulations of (a) the thrombin binding aptamer, $G_2T_2G_2TGTG_2T_2G_2$, (b) the thrombin binding aptamer with the central loop changed to GT, (c) and (d) G_3T simulations using the NMR and X-ray thrombin binding aptamer templates, respectively. In all cases only the two G-quartets are shown, and loops have been omitted for clarity. K^+ ions are shown as red spheres.

4.4.5 Limitations of the methods used

Some limitations of both CD spectroscopy and MD simulations have been described previously (Sections 1.4.2 and 2.5, respectively). Although the use of CD spectroscopy to classify quadruplexes as parallel or antiparallel is widespread, it has appeared that this method is not entirely reliable. The presence of both parallel and antiparallel conformations can give rise to a purely antiparallel-type spectrum. This is, for example, the case for $G_3T_2AG_3T_2AG_3T_2AG_3$, which has been shown to form both parallel and antiparallel structures in solution¹⁴⁸. The CD spectrum of the latter, shown in Figure 4.7(a) in K^+ ion containing solution, only shows the presence of antiparallel-type structures. The interpretation of CD spectra is only based on the observation of previous G-quadruplex spectra, and hence no new structural information can be obtained. CD spectroscopy is however useful in determining the presence of G-quadruplex molecules, although topological assignments should be made with care.

CD spectra reflect the G glycosidic angles of quadruplex structures. Parallel quadruplexes have all-*anti* glycosidic angles, whereas antiparallel quadruplexes generally have alternating *syn-anti* glycosidic angles down the strands, giving rise to the different spectra. The $G_4T_2G_4T_2G_4T_2G_4$ mixed parallel/antiparallel structure has *syn-syn-anti* glycosidic angles down each strand, and its CD spectrum has peaks characteristic of both parallel and antiparallel structures⁶⁸. It is therefore difficult to distinguish between the formation of two separate structures, and the formation of a single mixed structure using CD. Characteristic parallel quadruplex spectra could also be due to a tetrameric association of strands, rather than folding of a parallel intramolecular quadruplex.

The reliability of free energy calculations depends, amongst others, on sufficient sampling of the conformational space of each molecule. Although this may be realistic for short loops, the longer loops have many degrees of freedom, and simulations much longer than 4 ns would be required to achieve this. The T_2A loops, especially in parallel quadruplexes, had high rms deviations, and did not equilibrate to stable conformations over the 6 ns simulation of the crystal structure. Although this probably reflects the large flexibility of these loops in solution, longer simulations are required to sample all the conformational states. Moreover, the coexistence of quadruplex structures with similar energies in solution requires very accurate methods to evaluate ΔG . The values calculated here are only qualitative ranking values, which depend on the chosen initial loop conformations.

4.5 Loop-dependent folding

The combination of molecular dynamics and experimental (UV and CD spectroscopy) data has shown that loops can have a determinant effect on G-quadruplex folding. The loops affect not only the stability, but also the type of structure formed. The effects of loop size can be summarised as:

- Structures with three T residue loops can only form parallel structures.
- Structures with three T₂ loops can form both parallel and antiparallel structures, but parallel structures may be preferred.
- Structures with one T residue loop and two longer loops can form both parallel and mixed parallel/antiparallel structures.
- Structures with T₂ to T₆ loops can form both parallel and antiparallel structures, with the antiparallel conformation likely to be preferred.

Since the effects of shorter loop lengths on stability were due to steric factors, these results are likely to be applicable to loops of other residues than T. The influence of each loop is dependent upon the other loops within the quadruplex. Hence, a single T residue loop can adopt a lateral conformation, as long as the strain caused by this loop is compensated for by two other longer loops. In the same manner, although increasing all three loop lengths destabilises the quadruplexes, changing only one out of the three loops has little impact on stability (Table 4.4).

Free energy calculations showed that different structures often had similar energies, indicating that both parallel and antiparallel quadruplexes might coexist in solution. Several structures have been shown to form both parallel and antiparallel quadruplexes in solution. (TAG₃T₂AG₃T)⁴⁸ and G₃T₂AG₃T₂AG₃T₂AG₃¹⁴⁸ adopt both parallel and antiparallel conformations, which are similar in energy.

The stability trends measured for the varying loop lengths in this study cannot be generalised to all quadruplexes. Indeed, Balagurumorthy *et al* have shown that for dimeric G-quadruplexes, (G₄T₃G₄)₂ is less stable than (G₄T₄G₄)₂, which is the opposite to the trend observed here⁶⁷. Moreover, the simulation results suggest that the individual effect of each loop is dependent upon the other loops in the quadruplex structure. The results might also be dependent, in some cases, on the number of G-stacks which are formed, especially for parallel loop formation.

Chapter 5

Modelling of G-quadruplex-ligand interactions

The main motivation behind solving new G-quadruplex structures is to enable the design of ligands specific to these molecules. Specific targeting of G-rich sequences throughout the genome requires a knowledge of the receptor structure. Previous results, and the structural and modelling data in this work shows that this is rarely straightforward. G-rich sequences can fold into a variety of G-quadruplex structures, offering different binding sites for small molecules. In the present work, the binding of a set of quindoline analogues to both parallel and antiparallel quadruplexes was studied. Molecular dynamics simulations were used to compare the binding of the various ligands to the two receptors.

5.1 Ligand binding to G-quadruplexes

Direct structural data on ligand-quadruplex complexes is limited, compared to the number of small molecules which are known, or thought, to stabilise G-quadruplexes. There are two X-ray structures of small molecules bound to different quadruplexes. The structure of an acridine derivative complexed with an antiparallel $(G_4T_4G_4)_2$ quadruplex showed the ligand binding between a terminal G-quartet and a T_4 loop¹⁵⁰. Daunomycin was found to bind to a parallel $(TG_4T)_4$ quadruplex, by forming two layers of three ligands each, stacked between the G-quadruplexes¹⁵¹. In both cases, ligands have been shown to stack on the G-quadruplex surface, rather than intercalate between the G-quartets. For most ligand complexes, however, structures are not available, and lower resolution methods are used to obtain binding information. NMR data is frequently used to generate models of ligand binding¹⁵²⁻¹⁵⁴. For example, porphyrin binding to quadru-

plexes has been investigated using NMR in combination with fluorescence methods¹⁵⁵. Structure-activity relationships have been determined for a set of acridine derivatives, using surface plasmon resonance and telomerase assays⁴³.

Structural data and quadruplex binding assays have enabled the determination of three important G-quadruplex ligand features:

- All ligands have an aromatic core, which is capable of π -stacking with the G-quadruplex surface. A larger aromatic system can moreover increase the selectivity for quadruplexes over duplex or triplex DNA.
- The aromatic core generally has substituents capable of forming interactions with the quadruplex grooves.
- Positive charges on the ligands, both in the aromatic core and the side-chains, appear to increase binding efficiency.

All the ligands are thought to bind by stacking of the aromatic core with the G-quartets, and by the formation of interactions between the substituents and the G-quadruplex grooves. Some ligands have also been shown to be specific for certain quadruplex structures¹⁵⁶, or to favour the formation of specific quadruplex folds¹⁵⁵.

Theoretical methods such as molecular dynamics are frequently used to complement low resolution experimental data, and suggest ligand binding modes^{157–159}. Frequently, only simple and approximate binding calculations are carried out, in order to allow the comparison of a large set of ligands at a reduced computational cost^{43,157}. More thorough simulations have also been used to look at different binding modes of ligands, such as porphyrins¹⁶⁰. In this work, the binding of a set of quindoline derivatives to G-quadruplexes has been studied using long timescale MD simulations. Binding free energies were estimated using the MM-PBSA method, described in Section 2.3. The MM-PBSA method is a less computationally demanding alternative to more exact binding free energy calculation methods such as free energy perturbation (FEP) and thermodynamic integration (TI). The latter two methods, although reliable, only enable the comparison of very similar ligands, in identical binding sites. MM-PBSA methods have previously been used to study the binding of various ligands to duplex DNA^{161–163}, and have been successful in ranking and identifying correct binding modes of ligands to proteins^{164–167}. This method is attractive in that, theoretically, a much larger set of ligands and different binding modes can be compared.

The quindoline derivatives studied are shown in Figures 5.1 and 5.2. These have been shown to bind to quadruplex DNA, and inhibit telomerase¹⁶⁸. Other quindoline

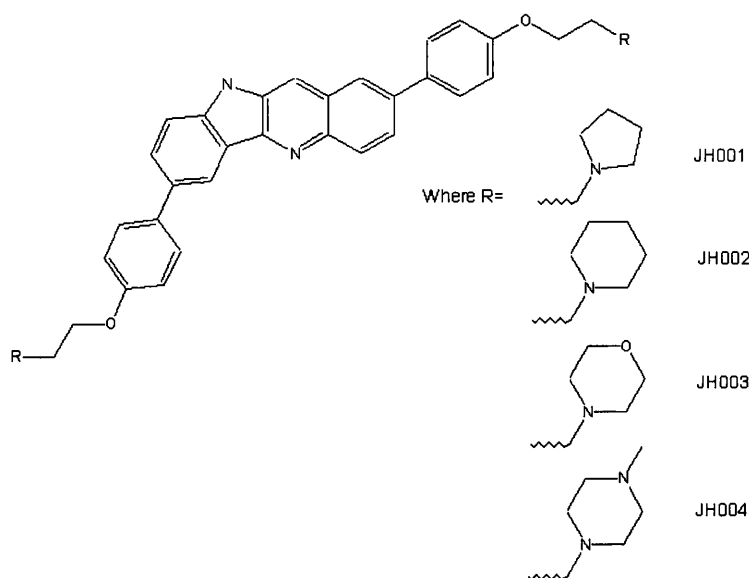


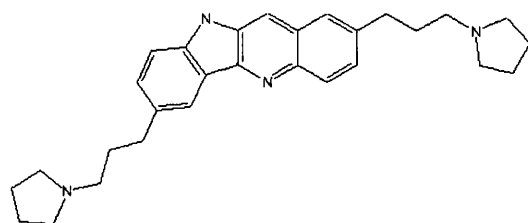
Figure 5.1: JH ligand structures.

Table 5.1: G-quadruplex ΔT_m ($^{\circ}\text{C}$) values for the quindoline ligands in this study. T_m values were determined at $10\ \mu\text{M}$ ligand concentrations. B ligand values were obtained from¹⁶⁸, JH ligand values were measured by Francisco Cuenca.

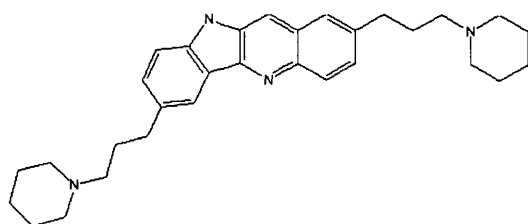
Ligand	ΔT_m	Ligand	ΔT_m
B124B	30	JH001	37
B128B	26	JH002	32
B123B	3	JH003	1
B120B	9	JH004	28

derivatives have also been shown to cause senescence and telomere shortening in human cancer cell lines¹⁶⁹. The JH ligands are related to the B ligands, but have an extended aromatic region and longer side-chains. Quadruplex stabilisation, as measured by ΔT_m values, has been previously determined for both ligand sets, using fluorescence resonance energy transfer (FRET). ΔT_m values are shown in Table 5.1.

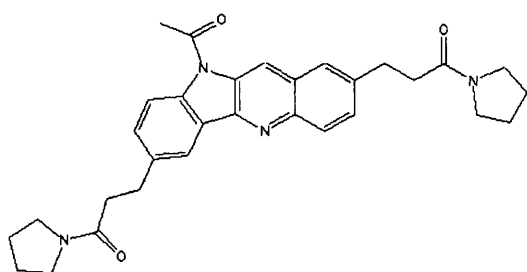
Studying ligand binding to G-quadruplexes is complicated by the inherent flexibility of the DNA. The initial choice of receptor structure is by no means straightforward, and can have a significant impact on the results. The human telomeric sequence, for example, can fold into parallel or antiparallel structures, which present radically different binding surfaces (Figure 4.3, page 111). Since its publication, the parallel structure



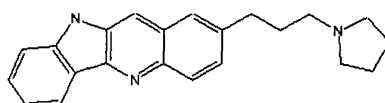
(a) B124B



(b) B128B



(c) B123B



(d) B120B

Figure 5.2: B ligand structures.

of $AG_3T_2AG_3T_2AG_3T_2AG_3$ has been regarded as an attractive target for ligand binding⁴⁷. The exposed G-quartet surfaces can accommodate ligands without any structural changes, and the latter do not compete with stabilising G-quartet-loop interactions. The parallel structure is also appealing from a computational standpoint, as docking ligands onto the surface can be achieved without requiring any modification of the receptor. However, all indications from experimental and theoretical studies are that G-quadruplexes are highly flexible structures. In order to account for this in the ligand binding studies, two receptor structures were investigated, a parallel and an antiparallel structure. The X-ray $AG_3T_2AG_3T_2AG_3T_2AG_3$ structure was used as a parallel quadruplex model. The $(G_4T_4G_4)_2$ quadruplex was used as an antiparallel model. The latter was chosen as its structure has been solved in complex with an acridine molecule¹⁵⁰.

5.2 Computational methods

5.2.1 Docking

Three different receptor structures were included in the binding studies. The crystal structure of $AG_3T_2AG_3T_2AG_3T_2AG_3$ (PDB code 1KF1) was used directly, after removal of the terminal A residue. During MD simulations, the parallel quadruplex structure above was shown to adopt a more extended loop geometry (Figure 4.3(b), page 111). In order to study the effect of the different loop conformations, the structure of $AG_3T_2AG_3T_2AG_3T_2AG_3$ obtained after a 4 ns simulation was also used as a receptor in the docking simulations, after removal of the terminal A residue. Binding to an antiparallel quadruplex was investigated using the crystal structure of $(G_4T_4G_4)_2$ bound to an acridine ligand (PDB code 1L1H). The acridine molecule was removed prior to docking. The Affinity module of Insight II was used for all the docking simulations. Ligand atom types and charges were assigned automatically using CVFF. Docking was carried out in implicit solvent, with a distance-dependent dielectric used to mimic solvent effects.

During all docking simulations, the ligands were fully flexible, allowing the side-chains to form stabilising interactions with the quadruplex grooves. Flexibility of the receptors was accounted for in two ways. In the antiparallel quadruplex, the loop residues forming the binding site were allowed to move with the ligand, and only the G-quartets were held fixed. For the parallel quadruplex, two structures with different loop conformations were used, although each was held fixed during the docking runs.

A two-step docking procedure was followed for the parallel receptors. Starting structures were generated by randomly positioning the ligand some distance (4 to 6 Å) above

the G-quadruplex plane. Initial searching for ligand binding positions was carried out by setting the electrostatic interactions to zero, and using a purely repulsive van der Waals potential. This avoids numerical instabilities which could occur due to the random placement of the ligand at the beginning of the search¹⁷⁰. 200 structures were generated using a Monte Carlo minimisation method, in which the ligand was first moved randomly, and then minimised for 300 steps. New structures were accepted if they had a lower energy than the previous structures, or if their Boltzmann factor was greater than a random number between 0 and 1. An rms deviation threshold of 1 Å was also used to ensure all structures were different. In the second stage, all 200 structures were recalled, minimised and subjected to simulated annealing for further refinement using more realistic van der Waals and electrostatic energy terms.

Docking to the antiparallel quadruplex was carried out differently. In this case, the binding site is well defined, and ligands were substituted to the acridine in the crystal structure. Minimisations and simulated annealing were carried out with the ligands in different starting orientations. The loop residues were kept flexible, and minimised with the ligands.

5.2.2 Molecular dynamics

The docked structures with the lowest energies were further simulated in explicit solvent. Prior to molecular dynamics, parameters for the ligand molecules were obtained. Bond, angle and torsional parameters were derived by analogy to parameters for equivalent atom types in gaff, the Amber general atom force field. Van der Waals parameters were also derived in this manner. Atomic charges were assigned for each of the ligands using RESP fitting to the electrostatic potential calculated with Gamess¹⁷¹. Ligands were first minimised at the HF level of theory with a 6-31G* basis set, for consistency with the gaff-derived parameters. The electrostatic potential was obtained from a single point energy calculation, at the same level of theory. Charges were calculated using a two-step procedure. In the first stage, all charges were fit independently, although charges of the aromatic rings on each side-chain were constrained to be identical in the JH ligands. In the second stage, all methyl and methylene groups were refitted, again constraining side-chains to identical charges, in both the B and JH ligands. The charges determined for all ligands are tabulated in Appendix B. The total molecular charge was specified as 0 for B123B and JH003, +1 for B120B, and +2 for B124B, B128B, JH001, JH002 and JH004. New ligand parameters and charges were incorporated into gaff by creating a "prep" file with the Antechamber module of Amber.

The equilibration procedure followed for each complex was identical to that described

previously for quadruplexes (Section 3.1.1). During the equilibration steps, restraints were applied to both the DNA and ligand. Simulations of 4 ns were carried out, again as described previously, and the ligands were compared by calculating binding free energies with the MM-PBSA method. For each calculation, BONDI radii were used, and free energies were averaged over the final 2 to 4 ns of simulation. MM-PBSA calculations were also carried out on separate trajectories of the complexes, ligands and receptors (Section 5.5.2). For this purpose, all the ligands were simulated alone in solution, with Cl^- ions neutralising the charge, where required. The parallel and antiparallel receptors were also simulated alone using the $\text{AG}_3\text{T}_2\text{AG}_3\text{T}_2\text{AG}_3\text{T}_2\text{AG}_3$ X-ray structure, and the $(\text{G}_4\text{T}_4\text{G}_4)_2$ X-ray structure with no ligand bound (PDB code 1JPQ).

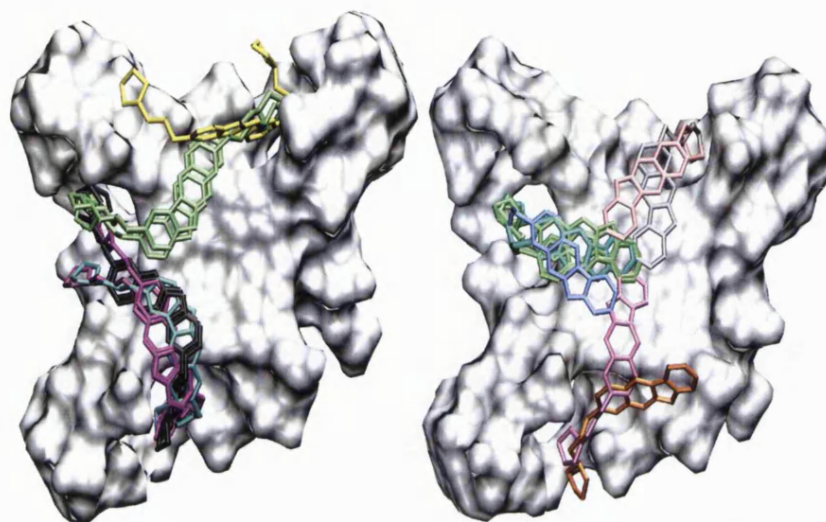
5.3 Docking results

5.3.1 $\text{G}_3\text{T}_2\text{AG}_3\text{T}_2\text{AG}_3\text{T}_2\text{AG}_3$ parallel receptor

Initially, both faces of the parallel G-quadruplex receptor were used for ligand docking. However, there was a clear preference for ligand binding to the 3' G-quadruplex face, rather than the 5' face. The two faces are non-equivalent, and the 3' face is more hydrophilic⁴⁷. Interaction energies between the ligand and the G-quadruplex computed using Discover were around $100 \text{ kcal.mol}^{-1}$ more favourable for the 3' face for the highest ranked structures. The 3' face was therefore used in all subsequent binding studies. This binding preference has already been suggested, and is in accord with surface plasmon resonance results⁴³.

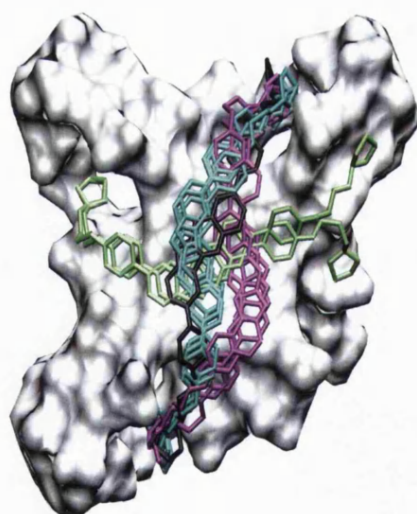
The ten most favourable docked structures of the B124B, B120B and JH001 ligands are shown in Figure 5.3. The effect of substituent length on binding is immediately apparent. The B124B ligand chromophore tends to stack with a single G residue, with each substituent lying in an adjacent quadruplex groove (Figure 5.3(a)). The only exception to this is shown in yellow in Figure 5.3(a), in which case the whole ligand is within the quadruplex groove. As the ligands are expected to interact via stacking of the chromophore with the G-quartets, this structure was not selected for further MD. Discrete binding sites for the aromatic core appear to be favoured, as shown by the close overlap of ligands with side-chains in the same grooves (coloured identically in Figure 5.3(a)). This could, however, be an effect of the side-chain-groove interactions governing the position of the aromatic core. Indeed, the B120B ligand, which possesses only a single side-chain, is much less positionally constrained on the G-quadruplex surface (Figure 5.3(b)).

In contrast to the B ligands, the JH ligands stretch across the quadruplex, with



(a) B124B ligand

(b) B120B ligand



(c) JH001 ligand

Figure 5.3: Top view of the overlap of the ten most favourable docked structures of (a) B124B, (b) B120B and (c) JH001 using the parallel quadruplex as a receptor. Closely related structures are drawn in the same colour. The DNA is represented as a solvent-excluded surface.

the substituents occupying opposite, rather than adjacent grooves (Figure 5.3(c)). The increased substituent length enables interactions with the quadruplex groove backbone, while maximising the stacking of the aromatic core with the G-quartets. The JH001 ligand chromophore occupies varied positions on the G-quartet plane, although one ligand orientation on the surface appears to be preferred (second to fourth groove rather than first to third groove).

In all the structures obtained for each ligands, the side-chains form interactions with the quadruplex backbone. Most frequently, the protonated substituent ring N atom forms interactions with the DNA phosphate backbone. The non-charged ligand substituents are also located in the quadruplex grooves, although further away from the DNA backbone.

Figure 5.4 shows examples of the most favourable structures obtained using the extended form of the parallel receptor (after 4 ns of MD). Top solutions for the JH ligands are similar to those obtained when using the crystal structure as a receptor (Figure 5.4(a)). In many cases, however, due to the different loop geometry, the side-chains are not buried as deep within the grooves. Binding of the B124B ligand appears to be more dependent on the loop conformation. Figure 5.4(b) shows that the side-chains of the top nine B124B structures occupy only two of the grooves, indicating a preference for a reduced set of binding sites. The fourteenth lowest energy structure obtained for the B124B ligand is also shown in Figure 5.4(b), as this was the most favourable structure obtained in which the ligand stretches across the surface of the quadruplex, similar to the JH ligands.

The docking procedure identified many potential binding sites on the G-quadruplex surface for the ligands investigated. Despite this, however, similar top solutions were consistently obtained for the JH ligands, increasing the confidence in the relevance of binding sites found. Figure 5.5 shows the similarities of the most favourable JH ligand binding sites. The most favourable B ligand structures showed more variations.

5.3.2 $(G_4T_4G_4)_2$ antiparallel receptor

Docking using the antiparallel receptor was carried out with the B124B, B128B and B120B ligands. Compared to the parallel quadruplex surface, the antiparallel structure has a much more confined binding site, and hence the diversity of structures generated was much lower. A subset of these are shown for the B124B ligand in Figure 5.6. The side-chains are slightly too short to penetrate the quadruplex grooves, and tend to form interactions with the loop backbone instead. This was also the case in the $(G_4T_4G_4)_2$ -acridine complex crystal structure¹⁵⁰. A major difference between the ligands

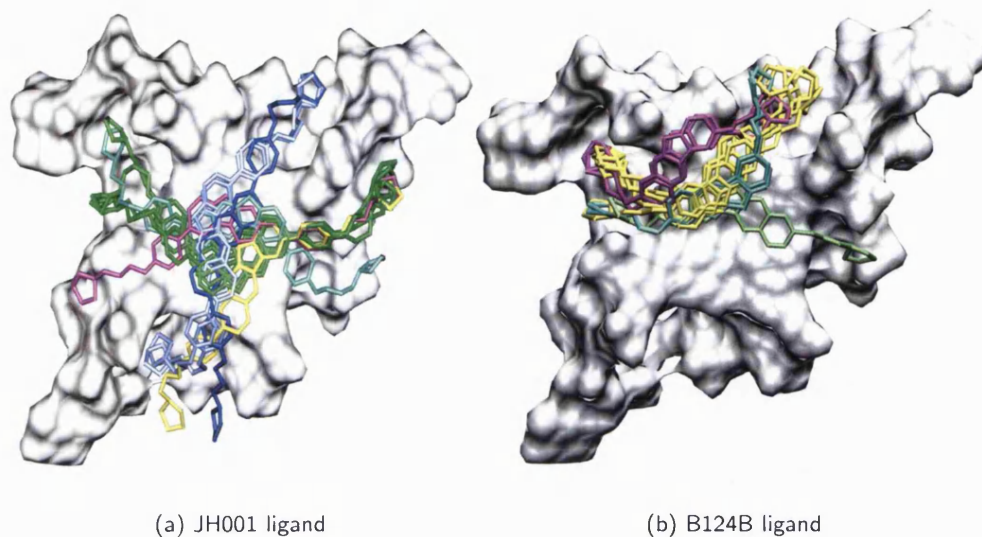


Figure 5.4: (a) Top view of the overlap of the ten most favourable docked structures of JH001. (b) Overlap of the nine most favourable docked structures of B124B and the fourteenth most favourable (in green). In both cases the 4 ns extended parallel quadruplex was used as a receptor. Closely related structures are drawn in the same colour.

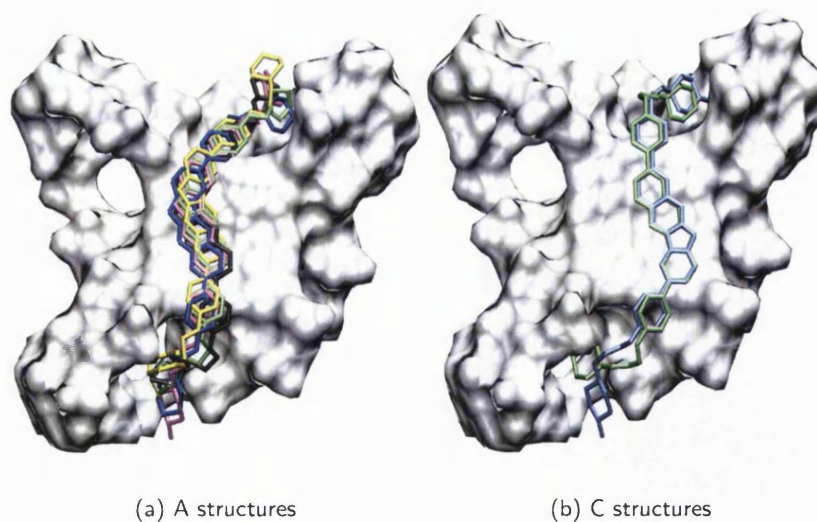


Figure 5.5: Top view of the overlap of the lowest energy structures obtained from the docking runs in Insight. (a) JH001-A green, JH001-A_{4ns} black, JH002-A blue, JH003-A yellow and JH004-A pink (b) JH002-C green and JH004-C mauve. Notation is as described in Table 5.2.

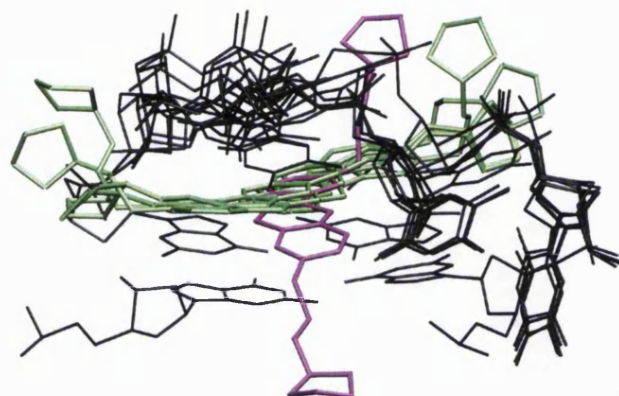


Figure 5.6: Overlap of the six most favourable docked structures of B124B using the antiparallel quadruplex as a receptor. Closely related structures are drawn in the same colour. The upper G-quartet of the DNA is shown in black, along with the T₄ loop which was allowed to move during the search.

investigated here and the acridine ligand is that the latter has a positive charge on the chromophore, which is not present here. This allows the acridine ligand to form hydrogen-bonding interactions with the loop residues. This type of interaction does not form here, which displaces the loop T residues slightly further away from the ligand molecule, compared to the crystal structure.

5.3.3 Energetic analysis

Table 5.2 lists the most favourable binding sites obtained with Insight for each ligand. In accordance with the experimental results in Table 5.1, the neutral ligands, JH003 and B123B, form the least stable complexes, with the lowest interaction energies. Moreover, the correct JH ligand trend is reproduced, with JH001 being the most favourable quadruplex binder, followed by JH002, JH004 and finally JH003. However, JH001 binding to the extended 4 ns receptor structure is much less favourable than to the X-ray structure. Although the experimental trends are reproduced within the ligand sets, the Discover calculations rank the B124B ligand as being the most favourable of all ligands, which is not found experimentally. Discover energy calculations include only a crude description of solvent effects and neglect the entropic contribution to binding. In order to obtain more accurate binding free energy estimates, MD simulations, followed by MM-PBSA calculations, were carried out on the complexes listed in Table 5.2.

Table 5.2: Summary of the most favourable complexes obtained with Insight for each ligand. These were further subjected to 4 ns MD simulations. Docking carried out with the parallel receptor after 4 ns of simulation is indicated with the 4 ns subscript. Similar binding sites of the JH ligands are indicated by the same letter. Interaction energies were calculated using the Discover module of Insight II and are in kcal.mol⁻¹.

Ligand	MD simulation	Interaction energy
Parallel quadruplex receptor		
JH001	JH001-A	-576
	JH001-A _{4ns}	-516
	JH001-B _{4ns}	-530
JH002	JH002-A	-567
	JH002-C	-567
JH003	JH003-A	-62
JH004	JH004-A	-556
	JH004-C	-556
B124B	B124B-A	-636
	B124B-B	-633
B128B	B128B-C	-539
B120B	B120B-D	-296
B123B	B123B-E	-54
Antiparallel quadruplex receptor		
B124B	B124B-1	-576
	B124B-2	-556
B128B	B128B-2*	
B120B	B120B-3	-208

*No energy was calculated as the B128B ligand was substituted for B124B in the B124B-2 structure

5.4 Molecular dynamics simulations

5.4.1 Parallel quadruplex complexes

All parallel quadruplex complexes were stable over 4 ns simulations. Although large ligand movement was observed in some cases, the ligands remained bound to the quadruplex surface. No effect of binding was observed on the G-quadruplexes. The hydrogen bonds between G residues were preserved, and loops behaved in the same manner as during the simulations of parallel quadruplexes alone. Both K^+ ions remained within the quadruplex channel during all of the simulations.

JH ligand simulations

Figure 5.7 shows the initial and final structures of some of the parallel quadruplex complexes simulated. Despite the ligand movement which occurred during the dynamics, the general binding site in each JH001 ligand simulation was maintained. Most of the ligand movement occurred in the side-chains, which formed ephemeral interactions with the quadruplex backbone. Fluctuations of the quadruplex loops caused the ligand side-chains to move. An example of this is shown in Figure 5.7(b), where in addition to the solvent accessible surface of the initial G-quadruplex structure, the surface of the final 4 ns receptor structure is shown as a blue mesh. Although the ligand side-chain in that region appears to be displaced from its initial position, it is not far removed from its original position relative to the DNA. The ligand chromophore of the JH ligands was also mobile on the quadruplex surface, although much less so than the side-chains. In the initial structures from the docking runs, the chromophore was generally positioned in a central position over the G-quartets, such that the benzene rings of each substituent partially stacked on the G-quartets. During the simulations, the ligands tended to slide over the surface, in order for one of the benzene rings to stack fully with the G-quartets, while the other benzene ring adopted an in-groove position. This is shown in the most extreme case for the JH004-C simulation in Figures 5.8(a) and (b), where the aromatic core of the ligand was displaced over the quadruplex groove. The sliding of the ligand across the G-quartet plane caused one of the side-chains to become deeply buried within the quadruplex groove (Figure 5.8(d)), while the other was pulled up towards the surface of the G-quartets (Figure 5.8(c)).

Figure 5.7(b) shows that the JH001- B_{4ns} simulation, starting from the 4 ns equilibrated structure of the receptor, conserved the initial position of the ligand better than the other simulations. The ligand chromophore did not exhibit as much movement during the simulation, as shown by the initial and final structure overlap. Using an ex-

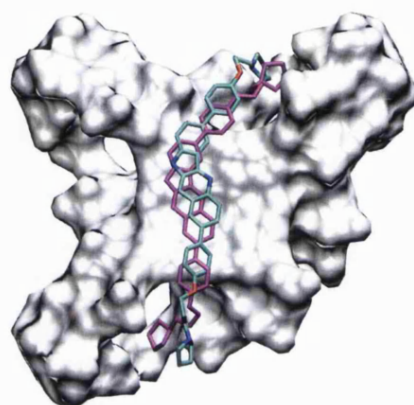
tended conformation of the receptor in the early stages of the docking procedure might, therefore, provide a more “realistic” receptor structure, at least when compared to the structure adopted during the simulations.

B ligand simulations

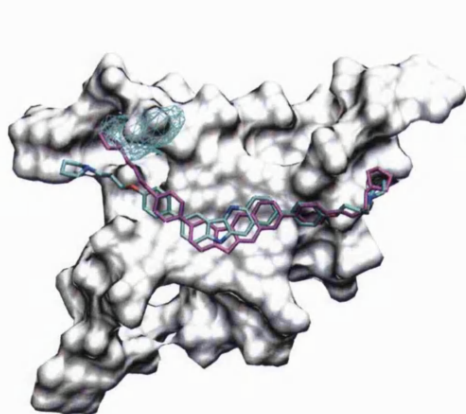
The flexibility of the B ligands during the simulations was much greater than the JH ligands. Initial and final structures of the B ligands are shown in Figure 5.9. Despite their flexibility, the JH ligand side-chains always occupied the same grooves during the simulations, and the ligand aromatic core moved only slightly on the quadruplex surface. The B ligands were much more mobile during the simulations, mostly due to the inability of the side-chains to maintain interactions with the quadruplex backbone. In each of the starting structures shown in Figure 5.9, the ligand side-chains were folded down into the quadruplex grooves, as predicted by the docking simulations. However, in the early stages of each simulation, at least one ligand substituent moved out of the quadruplex groove, and pointed into the solvent. The loss of this stabilising interaction meant that the chromophore was no longer “anchored” on the quadruplex surface, and was free to move around the G-quartet plane. In the B124B-A and B128B-C complexes, the ligands, originally stacking with a single G residue, favoured a position stacking over the centre of the G-quartet plane (Figures 5.9(a) and (c), respectively). The B124B-B complex, in which the ligand side-chains were originally located in opposite grooves, was the most flexible (Figure 5.9(b)). Neither of the side-chains were able to maintain favourable interactions with the DNA backbone, and the ligand was free to move on the quadruplex surface. The latter adopted a position normal to the initial structure.

The B ligands have shorter side-chains compared to the JH ligands, and this has an important impact on the simulations. Although the JH ligands can stretch across the G-quadruplex surface, and still form interactions with the backbone, this is impossible for the B ligands. Stacking between the B ligand chromophores and the G-quartets is favoured over the formation of interactions between the side-chains and the quadruplex grooves. Movement of the ligand was always at the expense of side-chain interactions, but enabled a better stacking of the ligand with the G-quartets. In both the JH and B ligand simulations, the chromophores favoured stacking over the centre of the G-quartet delocalised π surface, rather than discrete G residues.

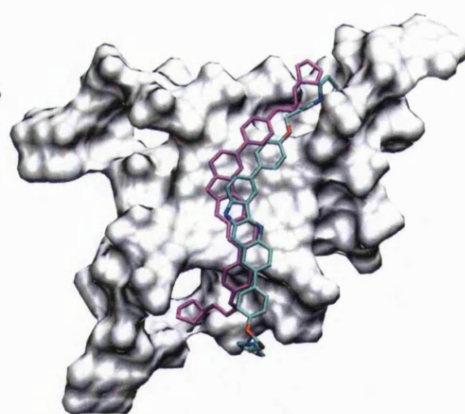
Surprisingly, the B ligand which showed the least movement on the G-quartet surface was B120B, which has a single substituent (Figure 5.9(d)). In the initial structure, the ligand side-chain interacted with a phosphate backbone O atom, deep in the quadruplex groove, but this changed after about 2 ns of simulation to an interaction with a phosphate



(a) JH001-A



(b) JH001-B_{4ns}



(c) JH001-A_{4ns}

Figure 5.7: Top view of the initial (purple) and final 4 ns (colour by atom) structures of the JH001-parallel quadruplex complexes. (a) JH001-A, (b) JH001-B_{4ns} and (c) JH001-A_{4ns}. The solvent-excluded surface of the DNA in the starting structure is shown in grey. (b) The solvent-excluded surface of the DNA in the final 4 ns structure around the ligand binding site is shown as a blue mesh.

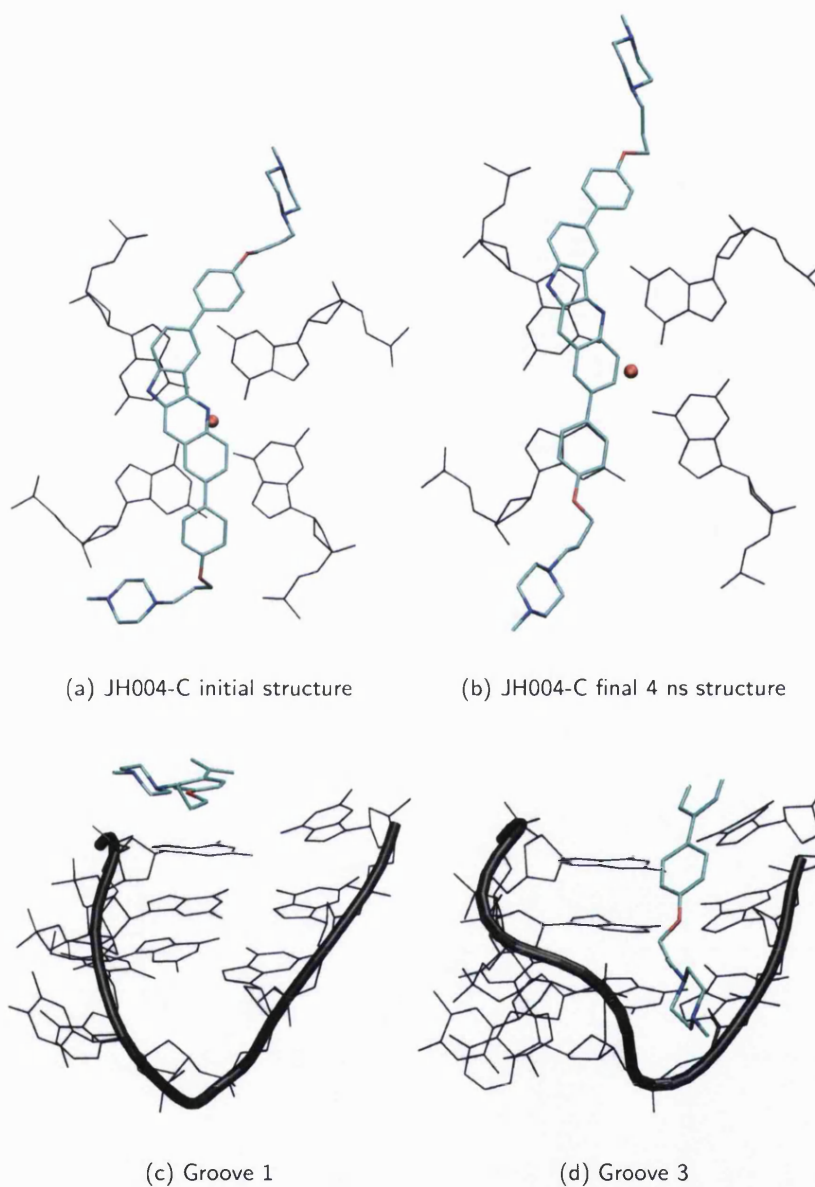


Figure 5.8: Top view of the (a) initial and (b) final structures of the JH004-C simulation. A single G-quartet is shown in black, with a channel K^+ ion shown in red. The ligand is coloured by atom. (c) Side view of the ligand interactions within the first and (d) third quadruplex grooves in the JH004-C final 4 ns structure. The groove backbone is shown as a thick black tube.

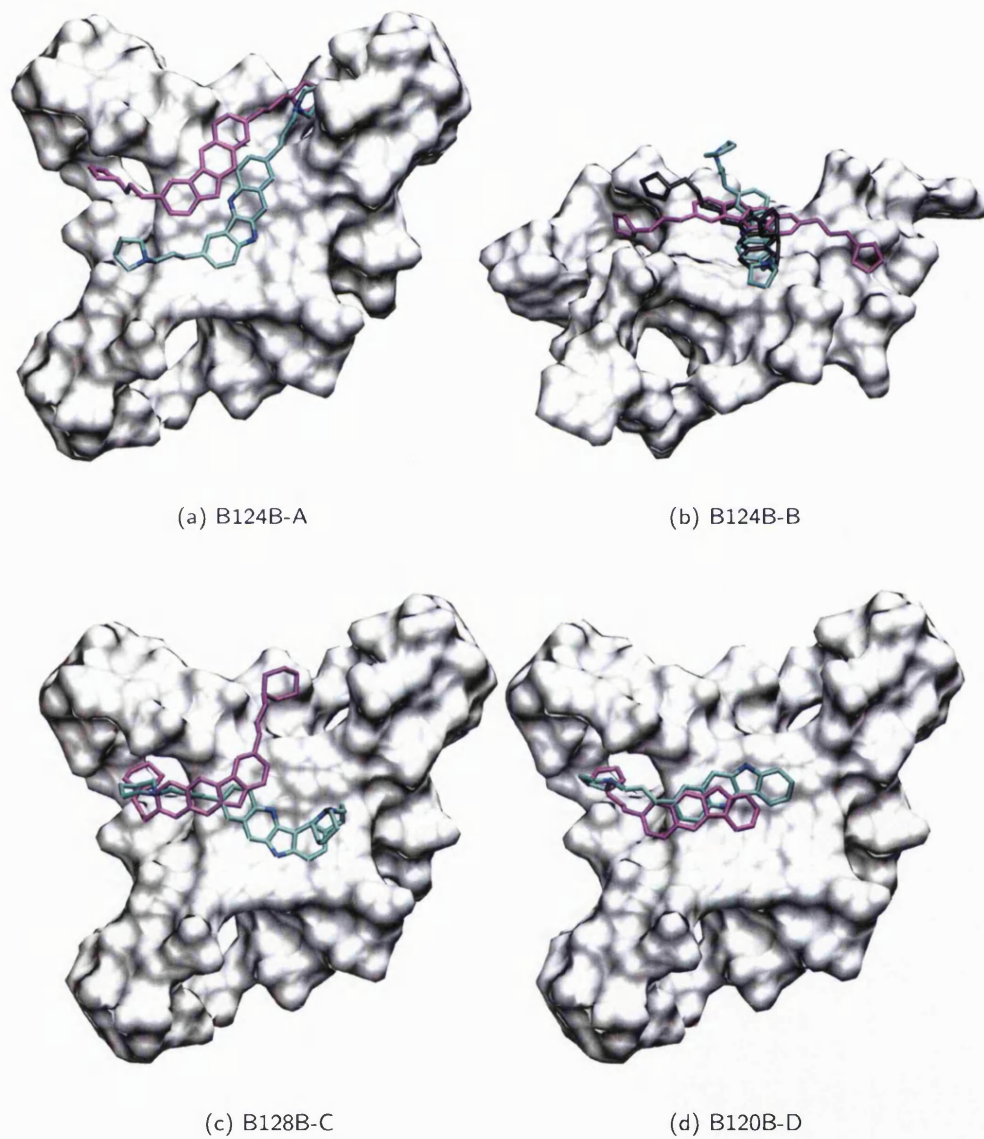


Figure 5.9: Top view of the starting (purple) and final 4 ns structures (colour by atom) of the B ligand-parallel quadruplex complexes. (b) 2 ns structure is also shown in black.

group O atom towards the top of the groove. This interaction was stable during the remaining 2 ns of simulation. This change was accommodated by a movement of the chromophore on the G-quartet surface, but this was much less pronounced than in the other B ligand simulations.

5.4.2 Antiparallel quadruplex complexes

Initial and final structures of the B ligand-antiparallel quadruplex complexes are shown in Figures 5.10 and 5.11. The confined binding site, between the G-quartet and loop, restricted ligand movement during the simulations. The side-chains, however, were flexible, and able to form varying interactions with the T₄ loop backbone. The hydrogen bonds formed between the protonated N atom of the side-chains and the DNA are shown in purple in Figures 5.10 and 5.11. Stacking interactions were formed between the ligand and loop residues. The side-chains were also favourably solvated by water molecules in the B124B-1 and B128B-2 complexes. As for the parallel quadruplexes, ligand binding to the antiparallel structures did not cause any G-quartet distortion. A slight tilt of the G-quartets around the ligands was observed, however hydrogen bonding and stacking of the G residues was unaffected.

5.5 Free energy analyses

The MM-PBSA method was used to calculate binding free energies according to the scheme detailed in Section 2.3. The method is approximate, but should enable the comparison of the energetics of similar ligands bound to similar receptors. Absolute binding free energies can be calculated if the entropy is taken into account, although this is the least reliably estimated contribution. This is required, however, if binding energy differences between the parallel and antiparallel complexes are to be estimated. The MM-PBSA method, however, is generally most successful in reproducing relative, rather than absolute $\Delta G_{\text{binding}}$ values¹⁶².

$\Delta G_{\text{binding}}$ can be calculated with the MM-PBSA method in two different ways. A single trajectory can be used for the calculation of the complex, receptor and ligand contributions. The complex is simulated in solution, and trajectories for the receptor and ligand are extracted from the former. This approximation is only valid if there is no perturbation of either the receptor or the ligand upon binding, as any reorganisation energy is ignored. When structural changes occur in either the ligand or the receptor upon complex formation, a three trajectory approach can be used, where the complex, receptor and ligands are each simulated separately in solution.

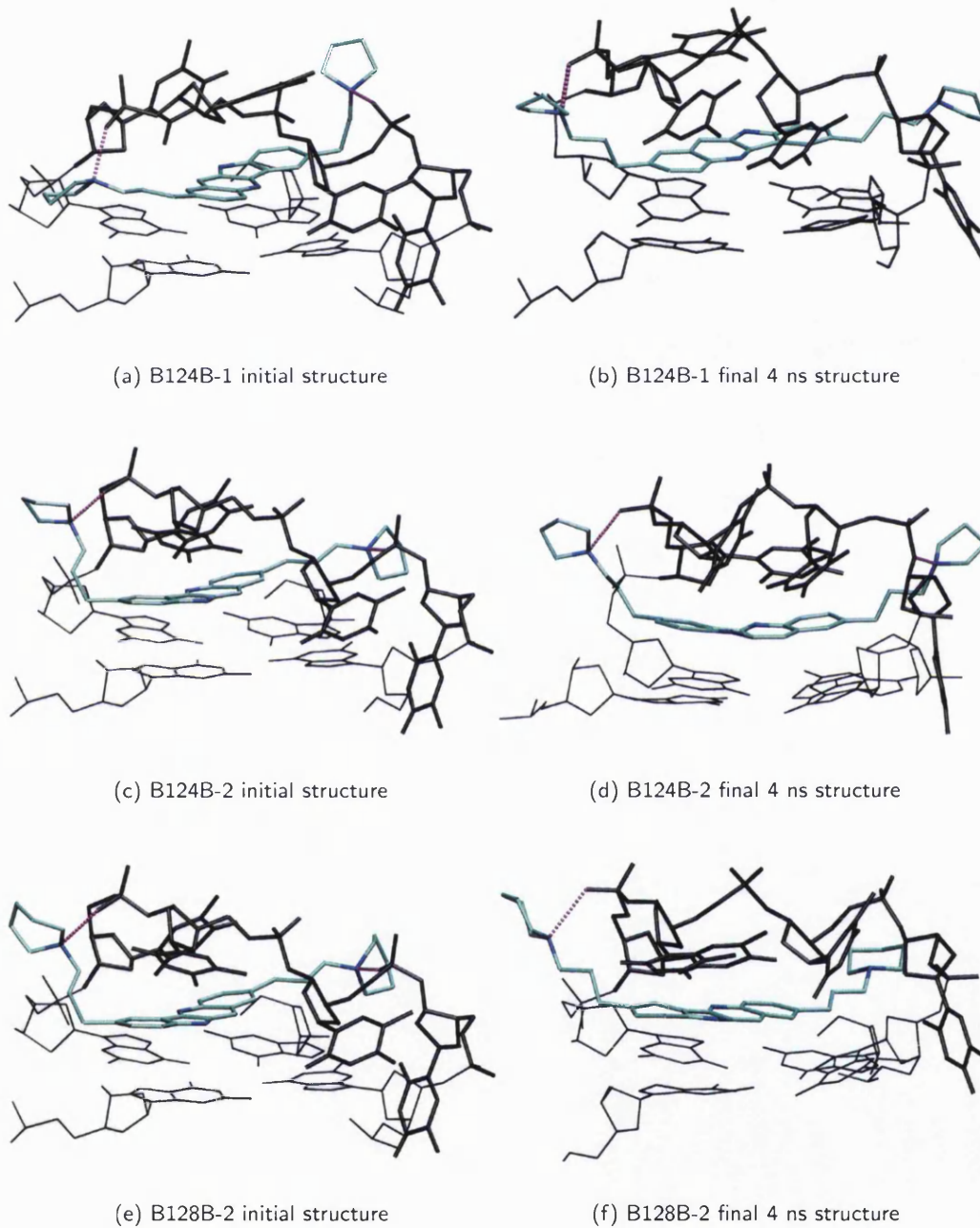


Figure 5.10: Initial and final 4 ns structures of the B124B-1, B124B-2 and B128B-2 antiparallel quadruplex complexes. One G-quartet and T₄ loop are shown in black, and ligand-quadruplex hydrogen bonds are shown as purple dotted lines.

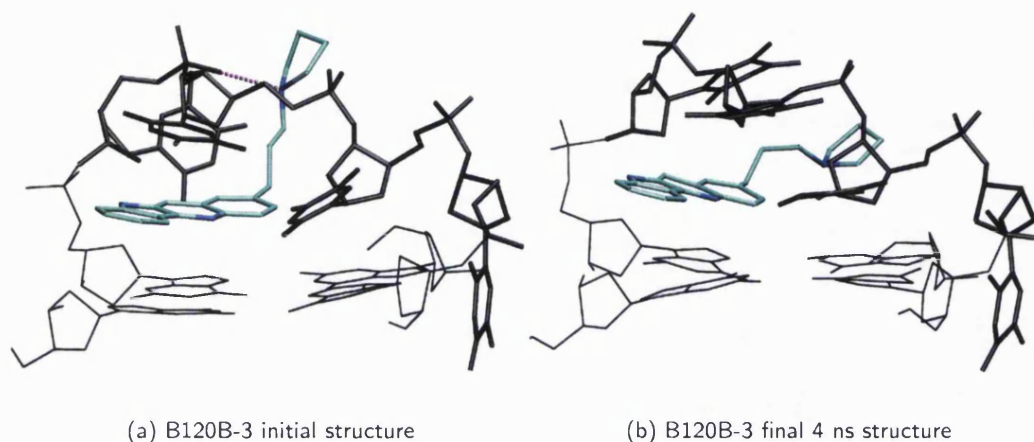


Figure 5.11: Initial and final 4 ns structures of the B120B-3 antiparallel quadruplex complex. One G-quartet and T_4 loop are shown in black, and ligand-quadruplex hydrogen bonds are shown as purple dotted lines.

Ligand binding to the parallel quadruplex structure is by stacking on the G-quartet surface. MD simulations of the complexes have shown that this does not cause any significant change in the receptor structure. The quadruplex loops interact with the ligand side-chains, however the former are already very flexible during trajectories of the receptor alone. It therefore appears that neglecting conformational change upon binding is a reasonable approximation, enabling the computational cost to be more than halved by using a single trajectory for the calculations. In the antiparallel quadruplex simulations the ligands intercalate between the loops and the G-quartets. In this case, ligand binding does cause a significant change to the receptor structure, and the binding energy measurement should preferably attempt to take this into account. This is possible when using three separate trajectories. The energy difference between the receptor when simulated alone, and the receptor as part of the complex is then an approximation of the energetic cost of binding site formation.

5.5.1 Single trajectory analysis

The binding free energies calculated for all the parallel quadruplex complexes using a single trajectory are shown in Table 5.3. The JH001, JH002 and JH004 ligands all have similar $\Delta G_{\text{binding}}$ values, which are within the calculation errors. Variations in $\Delta G_{\text{binding}}$ between different trajectories of the same ligand complexes are as large as

Table 5.3: Binding free energies of the parallel G-quadruplex-ligand complexes in this study. ΔG , $T\Delta S$ and $\Delta G_{\text{binding}}$ were calculated from a single simulation. All values are in kcal.mol^{-1} and $T=300$ K. Standard errors of the mean are in brackets. $\Delta G_{\text{binding}}^{3\text{trajectories}}$ values were calculated using separate trajectories for the complex, receptor and ligand.

Simulation	ΔG^*	$-T\Delta S$	$\Delta G_{\text{binding}}$	$\Delta G_{\text{binding}}^{3\text{trajectories}}$
JH001-A	-41 (5)	+30 (2)	-11	-10
JH001-A _{4ns}	-38 (5)	+28 (2)	-10	-5
JH001-B _{4ns}	-43 (2)	+28 (2)	-15	-13
JH002-A	-39 (6)	+30 (2)	-9	+3
JH002-C	-41 (5)	+30 (2)	-11	-11
JH003-A	-28 (5)	+28 (2)	0	+1
JH004-A	-38 (6)	+30 (2)	-8	-12
JH004-C	-43 (5)	+29 (2)	-14	-11
B124B-A	-30 (2)	+25 (3)	-5	+7
B124B-B	-26 (1)	+23 (2)	-3	-7
B128B-C	-28 (2)	+25 (2)	-3	+2
B120B-D	-24 (2)	+19 (2)	-5	-2
B123B-E	-32 (2)	+25 (2)	-7	-2

*Does not include the solute entropy

between different ligand complexes. This implies that the MM-PBSA method is not sufficiently accurate to discriminate between such similar ligands. The JH003 ligand, however, which has an overall zero charge, forms a much less favourable complex, with a calculated $\Delta G_{\text{binding}}$ of zero, in accordance with the experimental data in Table 5.1.

The B ligands formed less favourable complexes than the JH ligands. This is in accordance with observations during the simulations, in which the shorter side-chains were unable to form lasting contacts with the quadruplex backbone. All the B ligand complexes had similar $\Delta G_{\text{binding}}$ values, within 4 kcal.mol^{-1} , independent of the ligand charge. Both B123B, which has an overall charge of zero, and B120B, which has a charge of +1, formed complexes which were as stable as the B124B and B128B ligands. These results are not in accordance with experimental data obtained on the B ligand set. Both B124B and B128B were shown to stabilise G-quadruplexes, whereas B120B and B123B showed little, or no, effect (Table 5.1). The incorrect ranking of charged and neutral ligands using MM-PBSA has been observed previously, suggesting that the continuum approximation may not be adequate when describing the comparative desolvation costs of neutral and charged ligands¹⁶⁴.

$\Delta\Delta G_{\text{binding}}$ between similar ligands is expected to be the most reliably calculated value. However, there are significant structural differences between some of the ligands compared. The B120B ligand only has a single side-chain, while the B123B ligand has a different aromatic core, with an added COCH_3 group (Figure 5.2). Structural differences between the JH ligands are much smaller, and the JH003 neutral ligand is not significantly different from the other JH ligands. The results could suggest that the B ligands are too different to allow direct comparisons of $\Delta G_{\text{binding}}$. If this is the case, however, it is a significant setback for MM-PBSA calculations, since the latter were chosen specifically to enable the comparison of such structurally different ligands. This can also explain why significant $\Delta G_{\text{binding}}$ differences are obtained between the B and JH ligands, even though experimentally they stabilise G-quadruplexes by similar amounts (Table 5.1).

Decomposition of $\Delta G_{\text{binding}}$ into electrostatic and van der Waals components can explain some of the differences between ligands, and these terms are shown in Table 5.4. ($\Delta E_{\text{electrostatic}} + \Delta G_{\text{polar}}$) accounts for both the electrostatic interaction energy between ligand and receptor, and the electrostatic contribution to the solvation free energy. The ($\Delta E_{\text{vdW}} + \Delta G_{\text{nonpolar}}$) term accounts for the van der Waals interaction energy, and the non-polar component of the solvation free energy. For all complexes, the van der Waals component dominates the interaction energy, compensating for the unfavourable electrostatic contribution. Generally, the JH ligands have much more favourable van der Waals interactions with the quadruplexes than the B ligands. However, B123B forms very favourable van der Waals interactions with the DNA, leading to $\Delta G_{\text{binding}} = -7$ kcal.mol⁻¹ in Table 5.3. This could be due to the extra π -delocalisation on the aromatic core of this ligand.

5.5.2 Separate trajectory analysis

$\Delta G_{\text{binding}}^{3\text{trajectories}}$, calculated from three separate trajectories, is shown in Table 5.3 for the parallel quadruplex complexes. G_{complex} was calculated from the same trajectory as the single trajectory calculations, whereas G_{receptor} and G_{ligand} were calculated from separate simulations. MD simulations of 4 ns for each complex were generally found to be sufficient to obtain converged free energy values. However, the free energy of the parallel quadruplex receptor alone in solution did not equilibrate over an initial 4 ns simulation. Figure 5.12 shows the evolution of the parallel receptor free energy over the simulation. The linear regression to the free energy shows a decreasing trend over the 2 to 6 ns period. The energy did equilibrate over the following 6 to 20 ns, although it gradually increased by about 5 kcal.mol⁻¹ over this time period. When using this

Table 5.4: Decomposition of $\Delta G_{\text{binding}}$ of the parallel G-quadruplex-ligand complexes into electrostatic and van der Waals components. Free energies of the ligand, receptor and complex were calculated from a single simulation. All values are in kcal.mol⁻¹. Standard errors of the mean are in brackets.

Simulation	$\Delta E_{\text{electrostatic}} + \Delta G_{\text{polar}}$	$\Delta E_{\text{vdW}} + \Delta G_{\text{nonpolar}}$
JH001-A	+19 (4)	-61 (4)
JH001-A _{4ns}	+14 (4)	-52 (4)
JH001-B _{4ns}	+20 (2)	-63 (3)
JH002-A	+26 (4)	-66 (4)
JH002-C	+16 (4)	-57 (4)
JH003-A	+22 (4)	-50 (4)
JH004-A	+23 (4)	-61 (4)
JH004-C	+19 (4)	-63 (4)
B124B-A	+12 (2)	-42 (2)
B124B-B	+8 (1)	-35 (1)
B128B-C	+17 (2)	-45 (2)
B120B-D	+16 (2)	-40 (2)
B123B-E	+25 (2)	-57 (2)

simulation to calculate G_{receptor} , the results depend strongly on which portion of the trajectory is averaged. G_{receptor} is -4134 kcal.mol⁻¹ on average over the 2 to 4 ns trajectory portion, but decreases to -4144 kcal.mol⁻¹ on average over the 6 to 10 ns portion. In order to use equilibrated free energy values in the calculations, the 8 to 10 ns portion of the trajectory was used when computing $\Delta G_{\text{binding}}$. This corresponds to the most favourable receptor conformations obtained during the simulation.

The use of a single or three separate trajectories is not expected to have a significant impact on the free energy of the parallel quadruplex complexes. No change of the receptor was observed upon ligand binding, suggesting that reorganisation costs are low. Generally, $\Delta G_{\text{binding}}^{3\text{trajectories}}$ free energies in Table 5.3 were within error of the single trajectory values. The use of separate trajectories adds uncertainties in the calculations, due to fluctuations in receptor free energy, which are not necessarily due to ligand binding. This is exemplified by the JH002-A complex, which has a favourable $\Delta G_{\text{binding}} = -9$ kcal.mol⁻¹, but an unfavourable $\Delta G_{\text{binding}}^{3\text{trajectories}} = +3$ kcal.mol⁻¹. This difference is due mostly to the G_{receptor} term, which was less favourable during the complex simulation, compared to the simulation of the parallel quadruplex alone, and the other complex simulations. The binding free energy decomposition into ligand, receptor

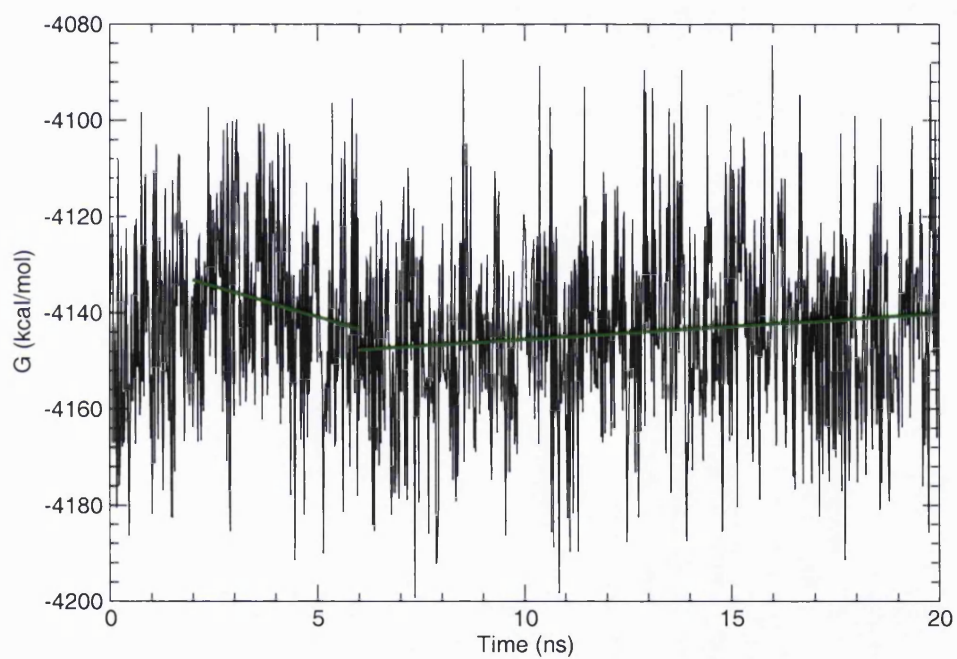


Figure 5.12: Evolution of the free energy of the parallel quadruplex receptor, with no terminal A residue, alone in solution, over the 20 ns simulation. The linear regression to the data is shown in green over 2 to 6 ns and 6 to 20 ns periods.

and complex contributions is shown in Figure 5.13 for the JH002-A simulation. G_{receptor} in the separate trajectory calculation is approximately 10 kcal.mol^{-1} more favourable than in the single trajectory calculation. This is not, however, necessarily due to the presence of a ligand, since the receptor free energy can fluctuate significantly depending on the simulation, even in the absence of a ligand (Figure 5.12). The effect of the unfavourable receptor is cancelled out in the single trajectory calculation, since both the complex and the receptor are affected by the same free energy difference. This was also the case for the B124B-A simulation, as shown by the $+12 \text{ kcal.mol}^{-1}$ difference between $\Delta G_{\text{binding}}$ and $\Delta G_{\text{binding}}^{3\text{trajectories}}$. The similarity of $\Delta G_{\text{binding}}$ and $\Delta G_{\text{binding}}^{3\text{trajectories}}$ for all the other trajectories indicates that the cost of binding site formation in the parallel quadruplex complexes can be omitted in the free energy calculations.

Although the use of separate trajectories should provide more information about ligand binding than the single trajectory approximation, it does not necessarily yield more accurate energies. Figure 5.13 shows that the fluctuation in $\Delta G_{\text{binding}}$ in the separate trajectory calculation is much greater than for the single trajectory calculation. Using a single trajectory for the receptor and complex allows cancellation of the fluctuations in energy which occur at each step during the simulations. Therefore, when binding is assumed not to involve any large reorganisation, it is perhaps more useful to calculate energies from a single trajectory, in order to minimise errors.

5.5.3 Antiparallel quadruplex complexes

Both single and separate trajectory binding free energies of the B ligand-antiparallel quadruplex complexes are shown in Table 5.5. As expected, the energetic cost of binding site formation in the antiparallel quadruplexes is non-negligible. $\Delta G_{\text{binding}}$ considerably overestimates the binding energy, and $\Delta G_{\text{binding}}^{3\text{trajectories}}$ is on average 28 kcal.mol^{-1} less favourable than the single trajectory value. $\Delta G_{\text{binding}}^{3\text{trajectories}}$ reproduces experimental data best, with B124B forming the most favourable complex, and B120B the least favourable. B124B-2 is more stable than B128B-2, however the 2 kcal.mol^{-1} energy difference is within the error of the MM-PBSA calculation. Although relative free energies between the ligands are in accordance with experimental data, the absolute binding energies are not. ΔT_m values for B120B and B128B indicate that these ligands do bind to and stabilise quadruplexes (albeit very little for the B120B ligand), however absolute $\Delta G_{\text{binding}}^{3\text{trajectories}}$ values are positive. Absolute binding free energies require the calculation of the entropic contribution to the free energy, which is the least reliably calculated term.

B ligand $\Delta G_{\text{binding}}^{3\text{trajectories}}$ values for the antiparallel quadruplexes are similar to binding energies obtained for the parallel quadruplex complexes. The MM-PBSA calculations

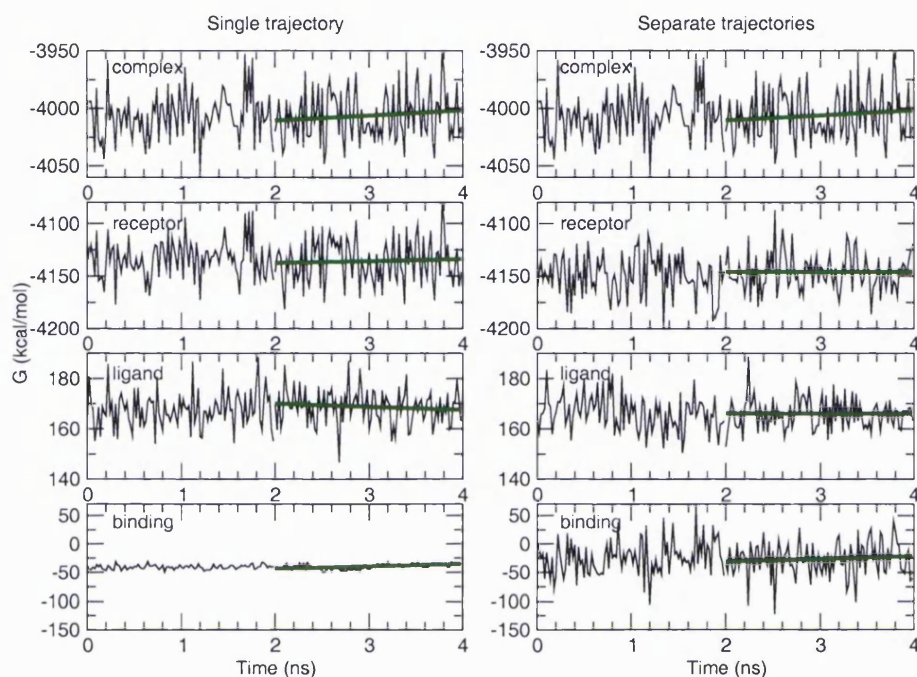


Figure 5.13: Free energies of the JH002-A complex calculated using a single and three separate trajectories. The green line shows the linear regression to the data over the 2 to 4 ns period used in the $\Delta G_{\text{binding}}$ calculations. G_{complex} values are identical in both calculations. G_{receptor} for the separate trajectory calculation corresponds to the free energy of the parallel quadruplex simulation over the 6 to 10 ns period (also shown in Figure 5.12).

Table 5.5: $\Delta G_{\text{binding}}$ for the B ligand-antiparallel quadruplex complexes. Calculations include three K^+ ions within the quadruplex channel. All values are in kcal.mol^{-1} and $T=300$ K.

Simulation	ΔG^*	$-T\Delta S$	$\Delta G_{\text{binding}}$	$\Delta G_{\text{binding}}^{\text{3trajectories}}$
B124B-1	-48 (2)	+23 (2)	-25	-5
B124B-2	-50 (2)	+21 (1)	-29	+3
B128B-2	-51 (2)	+22 (2)	-29	+5
B120B-3	-25 (2)	+19 (2)	-6	+19

*Does not include the solute entropy

thus suggest that B124B and B128B bind equally to both parallel and antiparallel quadruplexes. B120B, however, binds significantly less favourably to the antiparallel structure, due to the overestimation of the B120B ligand binding energy to the parallel quadruplex. MM-PBSA results suggest that the antiparallel quadruplex is a more appropriate receptor for the study of quindoline ligand binding. Free energy calculations were able to reproduce the B ligand experimental ranking using the antiparallel receptor only.

5.6 Conclusions

MM-PBSA calculations are computationally expensive since MD simulations have to be carried out to average the energies over the conformational space of the complexes. On the other hand, the Insight energy calculations were computationally cheap, as only a minimisation of the complexes was carried out. Moreover, these were able to reproduce the experimental JH ligand ranking using the X-ray structure of the parallel quadruplex receptor. However, the B128B complex was calculated to be much less favourable than the B124B complexes, when this is not found to be the case experimentally. Insight calculations using the antiparallel receptor structure did not take the cost of binding site formation into account, however this was shown to be non-negligible using MM-PBSA. Despite the limitations associated with MM-PBSA calculations, the major contributing factors to binding were identified. Solvation effects, and in some cases deformation energy, are important considerations in $\Delta G_{\text{binding}}$ calculations. Moreover, MD simulations showed how ligand binding is affected by the flexible nature of the quadruplex receptors. The shorter substituents of the B ligands were unable to maintain favourable interactions with the DNA backbone, in contrast to the JH ligands.

Unfortunately, the MM-PBSA method was unsuccessful in several respects. Experimental ranking of the JH and B ligands was not reproduced using the parallel quadruplex receptor. Although JH003 formed the least favourable complex, no significant difference was calculated between the JH001, JH002 and JH004 ligands. More concerning, however, was the inability to rank the B120B and B123B ligands compared to B124B and B128B, when using the parallel receptor. MM-PBSA methods suffer from limitations, and one of these is the requirement to accurately sample all of the ligand and receptor conformational space, in order to obtain meaningful results. This is especially relevant when considering the flexible parallel quadruplex molecules. Moreover, the B ligands diffused across the quadruplex surface during the simulations, and adopted many different binding sites. The 4 ns simulations cannot sample all the relevant conformational substates of each complex. However, free energies did equilibrate over the simulations,

suggesting that all the binding sites sampled during a simulation were relatively close in energy. The fact that ligand side-chains were often solvated, rather than forming interactions with the quadruplex backbone, could indicate inadequacies in the force field. Stacking interactions were found to be dominant in the simulations, however charge-charge interactions have been shown to be important experimentally¹⁶⁸.

Experimental ranking of the B ligands was reproduced when using the antiparallel quadruplex receptor. In this case, the binding site is clearly defined, and the ligands were much less flexible during the simulations. Conformational sampling is expected to be less of an issue than during the parallel quadruplex simulations, as is the initial choice of ligand position. MM-PBSA free energies were trajectory-dependent, relying on the docking simulations to provide sufficiently accurate ligand binding sites. How well this is achieved by the Insight simulations is difficult to establish in the absence of experimental structures.

Comparing free energies of structurally different ligands is challenging, and even more so when the binding sites are ill-defined, and both receptor and ligand are very flexible. In the absence of structural data on the binding of similar ligands to parallel quadruplexes, using the known structure of the antiparallel quadruplex-acridine complex as a model appears to be more reliable when comparing ligand binding energies. The calculation of smaller free energy differences, such as the difference between five and six membered rings in the ligand substituents, might require more exact methods such as FEP or TI.

Chapter 6

Ion binding within the G-quadruplex channel

The crystal structure of a $(TG_4T)_4$ quadruplex with both Na^+ and Ca^{2+} ions within the central channel has been determined by Michael Lee in this laboratory, and is shown in Figure 6.1. Although it is known that a number of cations can stabilise G-quadruplexes¹⁷², MD simulations have generally been limited to Na^+ or K^+ -containing structures. Divalent cations are notoriously difficult to simulate using MD, and parameterisation problems are more pronounced than for monovalent ions¹⁷³. The $(TG_4T)_4$ crystal structure in the presence of Ca^{2+} ions has provided the opportunity to assess the ability of MD simulations to model such a system.

Binding of ions within the quadruplex channel is accompanied by a large charge transfer from the G residues onto the cations¹³⁰. QM minimisations and Mulliken charge analysis have been used to quantify this charge transfer. However, Mulliken charges do not necessarily reproduce the electrostatic potential of a molecule. For consistency with the Amber force field used in the MD simulations, RESP fitting was used to calculate the charge of the G-quartet-ion systems in the present study, after QM minimisations using the Hartree-Fock (HF) 6-31G* level of theory. Density functional theory (DFT) methods have also been shown to give encouraging results for G-quartet-ion calculations¹³⁰, and were therefore also used for comparative purposes.

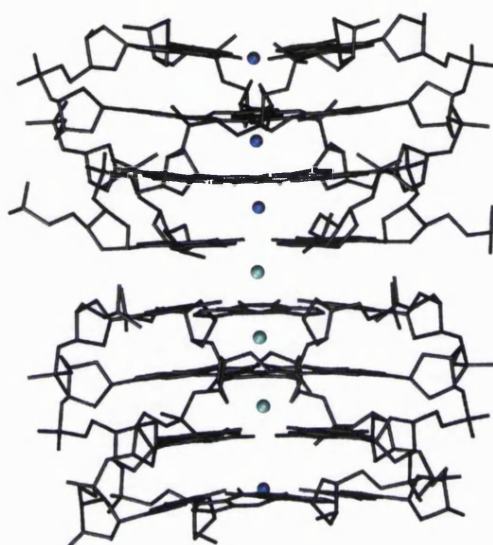


Figure 6.1: X-ray structure of the $(TG_4T)_4$ dimer with Na^+ and Ca^{2+} ions, solved by M. Lee. Ca^{2+} ions are shown as cyan spheres, Na^+ ions as blue spheres. The terminal T residues have been removed.

6.1 Computational methods

6.1.1 Quantum mechanical calculations

In order to limit the system size for computationally expensive QM calculations, only the G bases and ions were considered. The initial structures consisted of a single G base, capped with a methyl group, with an ion placed at a distance of 2.3 Å from the O6 atom, in the G base plane. C_{4h} symmetry was applied within Gamess to generate the remaining three G bases, forming a quartet. Structures with out-of-plane ions were built by placing the ion 1.5 Å above the quartet plane.

The Gamess package was used for all QM calculations. Delocalised coordinates were found to provide the fastest convergence. Both HF and DFT B3LYP levels of theory were employed, using a 6-31G* basis set. Symmetry was specified as C_4 , in order to allow ions to move in and out of the G-quartet plane. As the primary purpose of the QM calculations was to derive charges for the ions and G bases, Hessian calculations were not carried out on the minimised structures, as these are too computationally demanding.

Charges on the ion and G bases were determined in two ways. Mulliken charges were obtained from the Gamess single point calculations after minimisation. RESP charges were calculated using a two step procedure, as previously described (Section 5.2.2). All

Table 6.1: Summary of the G-quartet-ion QM calculations carried out. The energy is in Hartrees.

Ion	Method	Initial geometry	Minimised geometry	Energy
K ⁺	HF	in-plane	in-plane	-2912.916
K ⁺	HF	out-of-plane	out-of-plane	-2912.921
K ⁺	DFT	in-plane	out-of-plane	-2926.096
Na ⁺	HF	in-plane	in-plane	-2475.656
Na ⁺	HF	out-of-plane	in-plane	-2475.656
Ca ²⁺	HF	in-plane	in-plane	-2990.366
Ca ²⁺	DFT	in-plane	slightly out-of-plane	-3003.556

charges were fit independently in the first stage, and G bases were constrained to have identical charges in the second stage.

6.1.2 Molecular dynamics simulations

The crystal structure of two stacked (TG₄T)₄ quadruplexes with four Na⁺ ions and three Ca²⁺ ions was subjected to MD simulations. Prior to simulation, the terminal T residues were removed, and additional solution Na⁺ ions were added by the Leap module of Amber to neutralise the system. The quadruplex was solvated in a box of TIP3P water molecules. Equilibration and dynamics were carried out as previously described (Section 3.1.1). Two 4 ns simulations were carried out. In the first simulation, a charge of +2 was maintained on all Ca²⁺ ions within the G-quartet channel. A second simulation with a +1.5 charge on the Ca²⁺ ions attempted to account for some charge transfer from the G bases onto the ions. In both cases the G residue charges were unmodified from Amber ff99 parameterised values.

6.2 Quantum mechanical calculations

Table 6.1 summarises the QM calculations carried out. The initial and final minimised geometries are indicated, as well as the final minimised energy. Final structures depended not only on the nature of the ion, but also on the minimisation method. The different final structures obtained are shown in Figure 6.2.

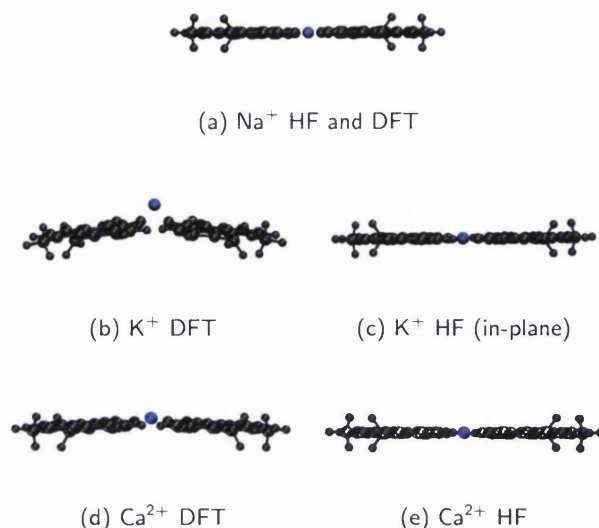


Figure 6.2: Ion-G-quartet structures minimised with QM methods. The G bases are shown in black and the ions in blue.

Na⁺ ions always adopted in-plane positions, independently of the method or starting structure (Figure 6.2(a)). Large K⁺ ions have been shown to preferentially adopt out-of-plane positions using DFT calculations¹⁷⁴. The DFT calculation starting from an in-plane K⁺ ion did not minimise to an out-of-plane structure (Figure 6.2(b)). However, a HF minimisation with an in-plane K⁺ ion remained in the initial conformation (Figure 6.2(c)). The in-plane K⁺ ion was accommodated through increased distances between the G bases (diametrical O6-O6 distance of 5.81 Å compared to 4.99 Å for the out-of-plane structure). The HF out-of-plane K⁺ structure was however about 3.1 kcal.mol⁻¹ more favourable than the in-plane structure (Table 6.1). These different G-quartet-ion structures enabled the comparison of the amount of charge transfer depending on the ion location with respect to the G quartet plane. G-quartet-Ca²⁺ structures were also level of theory dependent. An in-plane structure was obtained using HF, however the DFT minimised structure positioned the ion slightly out-of-plane (Figures 6.2(e) and (d), respectively).

Mulliken and RESP charges for the cations and O6 G atoms are shown in Table 6.2. There is a significant charge transfer from the G bases onto the cations, as shown by the reduced ionic charges. The amount of charge transfer was dependent upon the ion, but also on its position with respect to the G-quartet. Hence, the out-of-plane K⁺ ion had a RESP charge of 0.75, however this was further reduced to 0.60 when

Table 6.2: Ion and G base O6 atom charges for the QM minimised G-quartets.

Ion	Method	Minimised geometry	Mulliken charges		RESP charges	
			Ion charge	O6 charge	Ion charge	O6 charge
K ⁺	HF	in-plane	0.76	-0.76	0.60	-0.27
K ⁺	HF	out-of-plane	0.79	-0.75	0.75	-0.47
K ⁺	DFT	out-of-plane	0.66	-0.63	0.72	-0.41
Na ⁺	HF	in-plane	0.71	-0.75	0.49	-0.28
Ca ²⁺	HF	in-plane	1.51	-0.85	1.27	-0.39
Ca ²⁺	DFT	slightly out-of-plane	1.24	-0.69	1.26	-0.39

the K⁺ ion adopted an in-plane position. Charge transfer onto the Na⁺ ion was even greater, and can be explained by closer G bases, the O6-O6 distance across the quartet being 4.65 Å. As expected, there are significant differences between Mulliken and RESP charges, especially for the O6 atom charge. The importance of using consistent methods for charge derivation is highlighted by the very negative Mulliken O6 charge. The O6 G atom has a -0.5699 charge in the Amber force field, and in the presence of an ion the charge is expected to be less, rather than more, negative. This is indeed observed with RESP fitting of charges to the electrostatic potential.

QM minimisations and RESP charge fitting enabled the quantification of charge transfer between ion and G bases. Charge transfer depends not only on the nature of the ions, but also on the position of the ion with respect to the G-quartet. Although this may not affect K⁺ ions, which tend to occupy positions equidistant from the G-quartets, Na⁺ ions can occupy diverse binding sites. Na⁺ ions in the (TG₄T)₄ structure in Figure 6.1 are located both between and within G-quartets, resulting in different charge distributions along the quadruplex channel. Ca²⁺ ions in the (TG₄T)₄ structure are located between G-quartet planes, and therefore their ionic charge is expected to be slightly more positive than that calculated for the in-plane structures above.

6.3 Molecular dynamics simulations

The additive force field used in the simulations cannot account for the dynamic charge distribution between ions and G-quartets. Without the use of polarisable force fields, it is difficult to take charge transfer into account. Only single ion-quartet systems were

used in the QM calculations, however charge transfer is likely to depend on the location of the G-quartet (external or within a G-quartet stack), and the nature of the ions surrounding each quartet (which could be either Na^+ or Ca^{2+} in the $(\text{TG}_4\text{T})_4$ dimer considered). In a non-polarisable force field, different parameters would be required for nearly every G-quartet, and this would still not account for the dynamics of ions during the simulation.

MD simulations of the $(\text{TG}_4\text{T})_4$ dimer with a full +2 charge on the Ca^{2+} ions resulted in strong repulsion between the channel ions. Figure 6.3(a) shows that the Ca^{2+} ions were unable to occupy adjacent binding sites. Rapid diffusion of several Na^+ ions out of the quadruplex channel occurred during the heating stage of the equilibration. In an effort to take some charge transfer into account, but retain force field simplicity, the charge on Ca^{2+} ions was reduced to +1.5. This charge was chosen in order to maintain an integral overall charge and thus not require modification of other charges. This reduced Ca^{2+} charge enabled more ions to reside within the quadruplex channel (Figure 6.3(b)) during the simulation. Only one Na^+ ion diffused into the surrounding solution, and the other six ions remained either in the channel, or at the mouth of the quadruplex channel. The terminal Na^+ ions did not remain within the G-quartet plane, as observed in the crystal structure, but positioned themselves just above the channel. After initial rearrangement of the ions during the initial heating stage, ion location varied little during the simulation. A Ca^{2+} ion did move from one G-quartet step to an adjacent one during the dynamics, however this occurred over a few picoseconds only and the ion did not spend any time within the G-quartet plane.

6.4 Conclusions

Charge transfer from the G O6 atoms to the channel cations enables ions with a +2 charge to occupy adjacent G-quartet steps in the $(\text{TG}_4\text{T})_4$ crystal structure. Although this charge transfer can be quantified using QM calculations, MD simulations, even with reduced Ca^{2+} charges, could not reproduce the crystal structure. Strong repulsion between the cations lead to ion movement during the initial equilibration period. Although charge transfer also occurs between the Na^+ and K^+ ions and G-quartets, simulations of these ions are stable without taking this into account. However, this is not the case when larger charges are present.

Computational cost prohibits systems of greater than one or two G-quartets from being minimised using QM methods, necessitating the use of force field methods to simulate large structures. Present additive force fields, however, are inadequate when

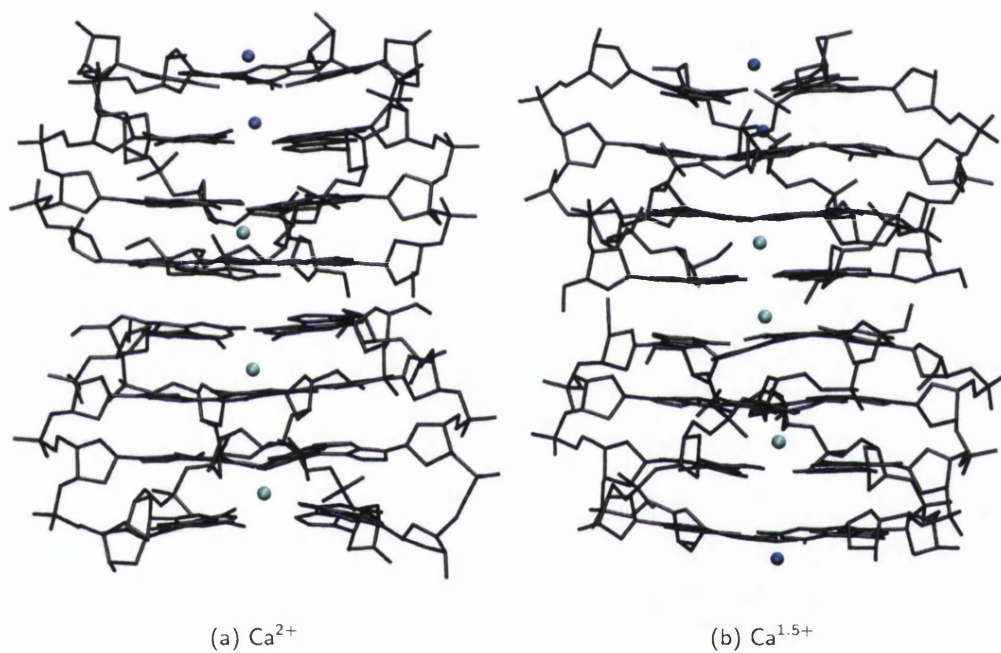


Figure 6.3: $(\text{TG}_4\text{T})_4$ dimer structures after 4 ns simulations with (a) a +2.0 and (b) a +1.5 charge on the Ca^{2+} ions. Ca^{2+} ions are shown in cyan, Na^+ ions in blue.

simulating quadruplexes with highly charged ions. Polarisable force fields are promising methods, which should permit such systems to be studied. These would allow the non-electrostatic terms of ion-G-quartet interactions, polarisability and charge transfer, to be included. Promising results have been obtained, for example when comparing Zn^{2+} and Mg^{2+} binding to DNA bases¹⁷³. However, polarisable force fields are still not commonly used, as many parameters have yet to be developed, especially for ions.

Chapter 7

Conclusions

The number of G-quadruplex structures, with different strand polarities and loop conformations, is continually increasing. Moreover, it is becoming more and more apparent that single sequences can form multiple structures, which often coexist under the same conditions. Not only has the coexistence of different folds been shown in solution, but the structure of $(G_4^{Br}UT_2G_4)_2$ shows that this can also occur in the crystal. The formation of two different dimers in a single crystal suggests that they are both equally stable.

In contrast to the previously determined $(G_4T_4G_4)_2$ structures in which diagonal loops were formed, $(G_4T_3G_4)_2$ and $(G_4T_2AG_4)_2$ have been found in this study to form antiparallel dimers with lateral loops. This confirms that loop length can be a determining factor in structures adopted by a quadruplex. Both T_3 and T_2A loops can adopt identical conformations, showing that length, rather than composition of the loops, is important. The structural preference for lateral, rather than diagonal, loops of three nucleotide length was not reproduced in MD simulations. Although the predicted most favourable T_3 loop is a lateral loop, diagonal loops have similar free energies, within the error margin of the free energy calculations. The crystal structures obtained do, however, show a preference, not only for lateral loops, but also for a particular lateral loop conformation (type 1 loop, described in Chapter 3). Free energy calculations suggest that the type 1 loop conformation is more stable than the other experimental loop conformations, however some predicted loops in different conformations are energetically similar. The loop conformation adopted in solution could be due to very small free energy differences, which cannot be reliably estimated with the free energy calculation methods used in this work.

MD simulations were more successful in determining the influence of shorter, one-

and two-nucleotide loops on G-quadruplex structure. $(G_4T_2G_4)_2$ was shown to be able to form only a subset of the structures which could be formed by $(G_4T_3G_4)_2$, due to restrictions on the distance which can be spanned by the two-nucleotide loops. Simulations of monomeric G-quadruplexes, in combination with CD data, suggest that sequences with three single-nucleotide loops can only form parallel quadruplexes, whereas sequences with a single-nucleotide loop and two longer loops can form parallel and mixed parallel/antiparallel quadruplexes. Moreover, sequences with three T_2 loops were shown to be able to form both parallel and antiparallel quadruplexes, as long as the latter contain lateral, rather than diagonal, loops. These simulations, as well as the crystal structures determined, highlight the importance of loop length on G-quadruplex folding. MD simulations could be extended to assess the effect of the number of G-quartet stacks on folding, which is likely to have an impact, especially on the formation of parallel quadruplex structures.

The structural plurality of G-quadruplexes makes the design of selective ligands challenging. Although the general ligand features outlined in Chapter 5 are quadruplex-specific, they do not enable the targeting of specific G-rich sequences. This is not straightforward, as in many cases the quadruplex structure is unknown, and even when it is known, there is likely to be an equilibrium between several conformations. Ligand binding is likely to favour a particular quadruplex fold, enabling some selectivity. Ligand binding to different G-quadruplex structures was studied using MD simulations and free energy calculations (Chapter 5). The simulation results were dependent on the particular receptors used, and experimental ligand ranking was reproduced only when using an antiparallel quadruplex structure. Failure to do so using the parallel quadruplex receptor is in part due to the poor experimental characterisation of these complexes. Parallel quadruplexes have large potential binding site areas, making the modelling work difficult. Contrary to stacking interactions between the ligands and quadruplexes, the ligand side-chain-DNA backbone interactions were generally not stable during the simulations, suggesting problems with the force fields when simulating these complexes. The development of computational methods to screen large libraries of ligands is challenging at present, due to an ever increasing number of possible receptor structures, and to the large number of possible binding sites on some of these structures.

Current MD force fields, such as the Amber force field used in this work, were parameterised to reproduce structures of proteins and duplex DNA. The fact that quadruplexes have been simulated with success testifies to the general applicability of the force fields. On the other hand, the characterisation of some important contributions to G-quadruplex folding and structure, such as cation binding, are beyond the scope of non-polarisable

force fields. MD simulations do require the presence of cations within the quadruplex channel for structural integrity, however more specific cation effects cannot be studied. Simulations of G-quadruplexes with Ca^{2+} ions were stable, however they were unable to reproduce the experimentally determined structure. The use of polarisable force fields could provide more information, and more accurately describe quadruplex-ion interactions. Moreover, increasing computational power will enable larger systems to be studied using QM methods, which are adapted to looking at local ion effects. However, additive force fields are commonly used because they are less computationally demanding. Perhaps a more appropriate parameterisation of ions, and more accurate free energy calculations, could enable ion-independent structure and loop conformations to be studied at no extra computational cost, using improved additive force fields.

Bibliography

- [1] M. Gellert, M. N. Lipsett & D. R. Davies (1962) Helix formation by guanylic acid. *Proc. Natl. Acad. Sci. USA* **48**, 2013–2018.
- [2] J. R. Fresco & J. Massoulié (1963) Polynucleotides v. helix-coil transition of polyribo-*g*uanylic acid. *J. Am. Chem. Soc.* **85**, 1352–1353.
- [3] C. Kang, X. Zhang, R. Ratliff, R. Moyzis & A. Rich (1992) Crystal structure of four-stranded *Oxytricha* telomeric DNA. *Nature* **356**, 126–131.
- [4] F. W. Smith & J. Feigon (1992) Quadruplex structure of *Oxytricha* telomeric DNA oligonucleotides. *Nature* **356**, 164–168.
- [5] S. Haider, G. N. Parkinson & S. Neidle (2002) Crystal structure of the potassium form of an *Oxytricha nova* G-quadruplex. *J. Mol. Biol.* **320**, 189–200.
- [6] F. Aboul-ela, A. I. H. Murchie & D. M. J. Lilley (1992) NMR study of parallel-stranded tetraplex formation by the hexadeoxynucleotide d(TG₄T). *Nature* **360**, 280–282.
- [7] F. Aboul-ela, A. I. H. Murchie, D. G. Norman & D. M. J. Lilley (1994) Solution structure of a parallel-stranded tetraplex formed by d(TG₄T) in the presence of sodium ions by nuclear magnetic resonance spectroscopy. *J. Mol. Biol.* **243**, 458–471.
- [8] G. Laughlan, A. I. Murchie, D. G. Norman, M. H. Moore, P. C. Moody, D. M. Lilley & B. Luisi (1994) The high-resolution crystal structure of a parallel-stranded guanine tetraplex. *Science* **265**, 520–524.
- [9] T. J. Pinnavaia, H. T. Miles & E. D. Becker (1975) Self-assembled 5'-guanosine monophosphate, nuclear magnetic resonance evidence for a regular, ordered structure and slow chemical exchange. *J. Am. Chem. Soc.* **97**, 7198–7200.

- [10] M. A. Keniry (2001) Quadruplex structures in nucleic acids. *Biopolymers* **56**, 123–146.
- [11] K. Phillips, Z. Dauter, A. I. Murchie, D. M. Lilley & B. Luisi (1997) The crystal structure of a parallel-stranded guanine tetraplex at 0.95 Å resolution. *J. Mol. Biol.* **273**, 171–182.
- [12] R. McElligott & R. J. Wellinger (1997) The terminal DNA structure of mammalian chromosomes. *EMBO J.* **16**, 3705–3714.
- [13] B. van Steensel, A. Smogorzewska & T. de Lange (1998) TRF2 protects human telomeres from end-to-end fusions. *Cell* **92**, 401–413.
- [14] P. Baumann & T. R. Cech (2001) Pot1, the putative telomere end-binding protein in fission yeast and humans. *Science* **292**, 1171–1175.
- [15] N. W. Kim, M. A. Piatyszek, K. R. Prowse, C. B. Harley, M. D. West, P. L. Ho, G. M. Coviello, W. E. Wright, S. L. Weinrich & J. W. Shay (1994) Specific association of human telomerase activity with immortal cells and cancer. *Science* **266**, 2011–2015.
- [16] S. Neidle & G. Parkinson (2002) Telomere maintenance as a target for anticancer drug discovery. *Nat. Rev. Drug Discov.* **1**, 383–393.
- [17] J.-L. Mergny & C. Hélène (1998) G-quadruplex DNA: A target for drug design. *Nat. Med.* **4**, 1366–1367.
- [18] A. J. Zaug, E. R. Podell & T. R. Cech (2005) Human POT1 disrupts telomeric G-quadruplexes allowing telomerase extension *in vitro*. *Proc. Natl. Acad. Sci. USA* **102**, 10864–10869.
- [19] A. K. Todd, M. Johnston & S. Neidle (2005) Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res.* **33**, 2901–2907.
- [20] J. L. Huppert & S. Balasubramanian (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.* **33**, 2908–2916.
- [21] T. Simonsson, P. Pecinka & M. Kubista (1998) DNA tetraplex formation in the control region of *c-myc*. *Nucleic Acids Res.* **26**, 1167–1172.
- [22] A. Siddiqui-Jain, C. L. Grand, D. J. Bearss & L. H. Hurley (2002) 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* **99**, 11593–11598.

- [23] P. Catasti, X. Chen, R. K. Moyzis, E. M. Bradbury & G. Gupta (1996) Structure-function correlations of the insulin-linked polymorphic region. *J. Mol. Biol.* **264**, 534–545.
- [24] M. Fry & L. A. Loeb (1994) The fragile X syndrome d(CGG)_n nucleotide repeats form a stable tetrahelical structure. *Proc. Natl. Acad. Sci. USA* **91**, 4950–4954.
- [25] A. Yafe, S. Etzioni, P. Weisman-Shomer & M. Fry (2005) Formation and properties of hairpin and tetraplex structures of guanine-rich regulatory sequences of muscle-specific genes. *Nucleic Acids Res.* **33**, 2887–2900.
- [26] R. M. Howell, K. J. Woodford, M. N. Weitzmann & K. Usdin (1996) The chicken β -globin gene promoter forms a novel "cinched" tetrahelical structure. *J. Biol. Chem.* **271**, 5208–5214.
- [27] A. Risitano & K. R. Fox (2003) Stability of intramolecular DNA quadruplexes: Comparison with DNA duplexes. *Biochemistry* **42**, 6507–6513.
- [28] A. T. Phan & J.-L. Mergny (2002) Human telomeric DNA: G-quadruplex, i-motif and Watson-Crick double helix. *Nucleic Acids Res.* **30**, 4618–4625.
- [29] K. Gehring, J.-L. Leroy & M. Guéron (1993) A tetrameric DNA structure with protonated cytosine-cytosine base pairs. *Nature* **363**, 561–565.
- [30] K. Paeschke, T. Simonsson, J. Postberg, D. Rhodes & H. J. Lipps (2005) Telomere end-binding proteins control the formation of G-quadruplex DNA structures *in vivo*. *Nat. Struct. Mol. Biol.* **12**, 847–854.
- [31] R. Giraldo & D. Rhodes (1994) The yeast telomere-binding protein RAP1 binds to and promotes the formation of DNA quadruplexes in telomeric DNA. *EMBO J.* **13**, 2411–2420.
- [32] R. Giraldo, M. Suzuki, L. Chapman & D. Rhodes (1994) Promotion of parallel DNA quadruplexes by a yeast telomere binding protein: A circular dichroism study. *Proc. Natl. Acad. Sci. USA* **91**, 7658–7662.
- [33] G. Fang & T. R. Cech (1993) Characterization of a G-quartet formation reaction promoted by the β -subunit of the *Oxytricha* telomere-binding protein. *Biochemistry* **32**, 11646–11657.
- [34] G. Fang & T. R. Cech (1993) The β subunit of *Oxytricha* telomere-binding protein promotes G-quartet formation by telomeric DNA. *Cell* **74**, 875–885.

- [35] H. Sun, J. K. Karow, I. D. Hickson & N. Maizels (1998) The Bloom's syndrome helicase unwinds G4 DNA. *J. Biol. Chem.* **273**, 27587–27592.
- [36] M. Fry & L. A. Loeb (1999) Human werner syndrome DNA helicase unwinds tetrahelical structures of the fragile X syndrome repeat sequence d(CGG)_n. *J. Biol. Chem.* **274**, 12797–12802.
- [37] K. Padmanabhan, K. P. Padmanabhan, J. D. Ferrara, J. E. Sadler & A. Tulinsky (1993) The structure of alpha-thrombin inhibited by a 15-mer single-stranded DNA aptamer. *J. Biol. Chem.* **268**, 17651–17654.
- [38] A. T. Phan, V. Kuryavyi, J.-B. Ma, A. Faure, M.-L. Andreola & D. J. Patel (2005) An interlocked dimeric parallel-stranded DNA quadruplex: A potent inhibitor of HIV-1 integrase. *Proc. Natl. Acad. Sci. USA* **102**, 634–639.
- [39] W. Li, P. Wu, T. Ohmichi & N. Sugimoto (2002) Characterization and thermodynamic properties of quadruplex/duplex competition. *FEBS Lett.* **526**, 77–81.
- [40] D. Sun, B. Thompson, B. E. Cathers, M. Salazar, S. M. Kerwin, J. O. Trent, C. Jenkins, T. S. Neidle & L. H. Hurley (1997) Inhibition of human telomerase by a G-quadruplex-interactive compound. *J. Med. Chem.* **40**, 2113–2116.
- [41] E. Izbicka, R. T. Wheelhouse, E. Raymond, K. K. Davidson, R. A. Lawrence, D. Sun, B. E. Windle, L. H. Hurley & D. D. Von Hoff (1999) Effects of cationic porphyrins as G-quadruplex interactive agents in human tumor cells. *Cancer Res.* **59**, 639–644.
- [42] J. F. Riou, L. Guittat, P. Mailliet, A. Laoui, E. Renou, O. Petitgenet, F. Mégnin-Chanet, C. Hélène & J.-L. Mergny (2002) Cell senescence and telomere shortening induced by a new series of specific G-quadruplex DNA ligands. *Proc. Natl. Acad. Sci. USA* **99**, 2672–2677.
- [43] R. Harrison, J. Cuesta, G. Chessari, M. Read, S. Basra, A. Reszka, J. Morrell, S. Gowan, C. Incles, F. Tanious, W. Wilson, L. Kelland & S. Neidle (2003) Trisubstituted acridine derivatives as potent and selective telomerase inhibitors. *J. Med. Chem.* **46**, 4463–4476.
- [44] M.-Y. Kim, H. Vankayalapati, K. Shin-ya, K. Wierzba & L. H. Hurley (2002) Telomestatin, a potent telomerase inhibitor that interacts quite specifically with the human telomeric intramolecular G-quadruplex. *J. Am. Chem. Soc.* **124**, 2098–2099.

- [45] Y. Krishnan-Ghosh, D. Liu & S. Balasubramanian (2004) Formation of an interlocked quadruplex dimer by d(GGGT). *J. Am. Chem. Soc.* **126**, 11009–11016.
- [46] N. Zhang, A. T. Phan & D. J. Patel (2005) (3+1) assembly of three human telomeric repeats into an asymmetric dimeric G-quadruplex. *J. Am. Chem. Soc.* advance online publication.
- [47] G. N. Parkinson, M. P. H. Lee & S. Neidle (2002) Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature* **417**, 876–880.
- [48] A. T. Phan & D. J. Patel (2003) Two-repeat human telomeric d(TAGGGTTAGGGT) sequence forms an interconverting parallel and antiparallel G-quadruplexes in solution: Distinct topologies, thermodynamic properties, and folding/unfolding kinetics. *J. Am. Chem. Soc.* **125**, 15021–15027.
- [49] P. Hazel, G. N. Parkinson & S. Neidle (2005) Topology variation and loop structural homology in crystal structures of a bimolecular DNA quadruplex. Manuscript in preparation.
- [50] A. T. Phan, Y. S. Modi & D. J. Patel (2004) Two-repeat *Tetrahymena* telomeric d(TGGGGTTGGGGT) sequence interconverts between asymmetric dimeric G-quadruplexes in solution. *J. Mol. Biol.* **338**, 93–102.
- [51] M. P. Horvath & S. C. Schultz (2001) DNA G-quartets in a 1.86 Å resolution structure of an *Oxytricha nova* telomeric protein-DNA complex. *J. Mol. Biol.* **310**, 367–377.
- [52] P. Schultze, N. V. Hud, F. W. Smith & J. Feigon (1999) The effect of sodium, potassium and ammonium ions on the conformation of the dimeric quadruplex formed by the *Oxytricha nova* telomere repeat oligonucleotide d(G₄T₄G₄). *Nucleic Acids Res.* **27**, 3018–3028.
- [53] P. Schultze, F. W. Smith & J. Feigon (1994) Refined solution structure of the dimeric quadruplex formed from the *Oxytricha* telomeric oligonucleotide d(GGGGTTTTGGGG). *Structure* **2**, 221–233.
- [54] M. Črnogelj, N. V. Hud & J. Plavec (2002) The solution structure of d(G₄T₄G₃)₂: a bimolecular G-quadruplex with a novel fold. *J. Mol. Biol.* **320**, 911–924.
- [55] G. D. Strahan, M. A. Keniry & R. H. Shafer (1998) NMR structure refinement and dynamics of the K⁺-[d(G₃T₄G₃)₂] quadruplex via particle mesh Ewald molecular dynamics simulations. *Biophys. J.* **75**, 968–981.

- [56] F. W. Smith, F. W. Lau & J. Feigon (1994) $d(G_3T_4G_3)$ forms an asymmetric diagonally looped dimeric quadruplex with guanosine 5'-syn-syn-anti and 5'-syn-anti-anti N-glycosidic conformations. *Proc. Natl. Acad. Sci. USA* **91**, 10546–10550.
- [57] M. Črnugelj, P. Šket & J. Plavec (2003) Small change in a G-rich sequence, a dramatic change in topology: new dimeric G-quadruplex folding motif with unique loop orientations. *J. Am. Chem. Soc.* **125**, 7866–7871.
- [58] Y. Wang & D. J. Patel (1994) Solution structure of the *Tetrahymena* telomeric repeat $d(T_2G_4)_4$ G-tetraplex. *Structure* **2**, 1141–1156.
- [59] V. Kuryavii, A. Majumdar, A. Shallop, N. Chernichenko, E. Skripkin, R. Jones & D. J. Patel (2001) A double chain reversal loop and two diagonal loops define the architecture of a unimolecular DNA quadruplex containing a pair of stacked $G(syn) \cdot G(syn) \cdot G(anti) \cdot G(anti)$ tetrads flanked by a $G \cdot (T-T)$ triad and a $T \cdot T \cdot T$ triple. *J. Mol. Biol.* **310**, 181–194.
- [60] A. Ambrus, D. Chen, J. Dai, R. A. Jones & D. Yang (2005) Solution structure of the biologically relevant G-quadruplex element in the human c-MYC promoter. Implications for G-quadruplex stabilization. *Biochemistry* **44**, 2048–2058.
- [61] A. T. Phan, Y. S. Modi & D. J. Patel (2004) Propeller-type parallel-stranded G-quadruplexes in the human *c-myc* promoter. *J. Am. Chem. Soc.* **126**, 8710–8716.
- [62] Y. Wang & D. J. Patel (1993) Solution structure of the human telomeric repeat $d(AG_3(T_2AG_3)_3)$ G-tetraplex. *Structure* **1**, 263–282.
- [63] Y. Wang & D. J. Patel (1995) Solution structure of the *Oxytricha* telomeric repeat $d[G_4(T_4G_4)_3]$ G-tetraplex. *J. Mol. Biol.* **251**, 76–94.
- [64] R. F. Macaya, P. Schultze, F. W. Smith, J. A. Roe & J. Feigon (1993) Thrombin-binding DNA aptamer forms a unimolecular quadruplex structure in solution. *Proc. Natl. Acad. Sci. USA* **90**, 3745–3749.
- [65] N. Jing & M. E. Hogan (1998) Structure-activity of tetrad-forming oligonucleotides as a potent anti-HIV therapeutic drug. *J. Biol. Chem.* **273**, 34992–34999.
- [66] I. N. Rujan, J. C. Meleney & P. H. Bolton (2005) Vertebrate telomere repeat DNAs favor external loop propeller quadruplex structures in the presence of high concentrations of potassium. *Nucleic Acids Res.* **33**, 2022–2031.

- [67] P. Balagurumoorthy, S. K. Brahmachari, D. Mohanty, M. Bansal & V. Sasisekharan (1992) Hairpin and parallel quartet structures for telomeric sequences. *Nucleic Acids Res.* **20**, 4061–4067.
- [68] V. Dapic, V. Abdomerovic, R. Marrington, J. Peberdy, A. Rodger, J. O. Trent & P. J. Bates (2003) Biophysical and biological properties of quadruplex oligodeoxyribonucleotides. *Nucleic Acids Res.* **31**, 2097–2107.
- [69] J.-L. Mergny, A.-T. Phan & L. Lacroix (1998) Following G-quartet formation by UV-spectroscopy. *FEBS Lett.* **435**, 74–78.
- [70] I. Smirnov & R. H. Shafer (2000) Effect of loop sequence and size on DNA aptamer stability. *Biochemistry* **39**, 1462–1468.
- [71] X. Zhang, E. Cao, X. Sun & C. Bai (2000) Circular dichroism spectroscopic studies on structures formed by telomeric DNA sequences *in vitro*. *Chin. Sci. Bull.* **45**, 1959–1963.
- [72] N. Jing, R. F. Rando, Y. Pommier & M. E. Hogan (1997) Ion selective folding of loop domains in a potent anti-HIV oligonucleotide. *Biochemistry* **36**, 12498–12505.
- [73] R. Jin, B. L. Gaffney, C. Wang, R. A. Jones & K. J. Breslauer (1992) Thermodynamics and structure of a DNA tetraplex: A spectroscopic and calorimetric study of the tetramolecular complexes of d(TG₃T) and d(TG₃T₂G₃T). *Proc. Natl. Acad. Sci. USA* **89**, 8832–8836.
- [74] P. Balagurumoorthy & S. K. Brahmachari (1994) Structure and stability of human telomeric sequence. *J. Biol. Chem.* **269**, 21858–21869.
- [75] M. A. Keniry, E. A. Owen & R. H. Shafer (1997) The contribution of thymine-thymine interactions to the stability of folded dimeric quadruplexes. *Nucleic Acids Res.* **25**, 4389–4392.
- [76] M. Cevc & J. Plavec (2005) Role of loop residues and cations on the formation and stability of dimeric DNA G-quadruplexes. *Biochemistry* **44**, 15238–15246.
- [77] A. Risitano & K. R. Fox (2004) Influence of loop size on the stability of intramolecular DNA quadruplexes. *Nucleic Acids Res.* **32**, 2598–2606.

- [78] I. Ourliac-Garnier, M.-A. Elizondo-Riojas, S. Redon, N. P. Farrell & S. Bombard (2005) Cross-links of quadruplex structures from human telomeric DNA by dinuclear platinum complexes show the flexibility of both structures. *Biochemistry* **44**, 10620–10634.
- [79] T. E. Cheatham III, J. L. Miller, T. Fox, T. A. Darden & P. A. Kollman (1995) Molecular dynamics simulations on solvated biomolecular systems: the Particle Mesh Ewald method leads to stable trajectories of DNA, RNA, and proteins. *J. Am. Chem. Soc.* **117**, 4193–4194.
- [80] T. Darden, D. York & L. Pedersen (1993) Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092.
- [81] T. Darden, L. Perera, L. Li & L. Pedersen (1999) New tricks for modelers from the crystallography toolkit: the particle mesh Ewald algorithm and its use in nucleic acid simulations. *Structure* **7**, R55–R60.
- [82] T. E. Cheatham III (2004) Simulation and modeling of nucleic acid structure, dynamics and interactions. *Curr. Opin. Struct. Biol.* **14**, 360–367.
- [83] J. Wang, P. Cieplak & P. A. Kollman (2000) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **21**, 1049–1074.
- [84] P. Dauber-Osguthorpe, V. A. Roberts, D. J. Osguthorpe, J. Wolff, M. Genest & A. T. Hagler (1988) Structure and energetics of ligand binding to proteins: Escherichia coli dihydrofolate reductase-trimethoprim, a drug-receptor system. *Proteins* **4**, 31–47.
- [85] J. R. Maple, M.-J. Hwang, T. P. Stockfisch, U. Dinur, M. Waldman, C. S. Ewig & A. T. Hagler (1994) Derivation of class II force fields. I. Methodology and quantum force field for the alkyl functional group and alkane molecules. *J. Comput. Chem.* **15**, 162–182.
- [86] D. A. Case, T. A. Darden, T. E. Cheatham III, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, K. M. Merz, B. Wang, D. A. Pearlman, M. Crowley, S. Brozell, V. Tsui, H. Gohlke, J. Mongan, V. Hornak, G. Cui, P. Beroza, C. Schafmeister, J. W. Caldwell, W. S. Ross & P. A. Kollman (2004). AMBER 8. University of California, San Francisco.

- [87] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman & D. A. Case (2004) Development and testing of a general amber force field. *J. Comput. Chem.* **25**, 1157–1174.
- [88] N. Spacková, I. Berger & J. Sponer (1999) Nanosecond molecular dynamics simulations of parallel and antiparallel guanine quadruplex DNA molecules. *J. Am. Chem. Soc.* **121**, 5519–5534.
- [89] R. Stefl, N. Spacková, I. Berger, J. Koca & J. Sponer (2001) Molecular dynamics of DNA quadruplex molecules containing inosine, 6-thioguanine and 6-thiopurine. *Biophys. J.* **80**, 455–468.
- [90] N. Spacková, I. Berger & J. Sponer (2001) Structural dynamics and cation interactions of DNA quadruplex molecules containing mixed guanine/cytosine quartets revealed by large-scale MD simulations. *J. Am. Chem. Soc.* **123**, 3295–3307.
- [91] S. Chowdhury & M. Bansal (2001) G-quadruplex structure can be stable with only some coordination sites being occupied by cations: A six-nanosecond molecular dynamics study. *J. Phys. Chem. B* **105**, 7572–7578.
- [92] S. Chowdhury & M. Bansal (2001) A nanosecond molecular dynamics study of antiparallel d(G)₇ quadruplex structures: Effect of the coordinated cations. *J. Biomol. Struct. Dyn.* **18**, 647–670.
- [93] R. Stefl, T. E. Cheatham III, N. Spacková, E. Fadrná, I. Berger, J. Koca & J. Sponer (2003) Formation pathways of a guanine-quadruplex DNA revealed by molecular dynamics and thermodynamic analysis of substates. *Biophys. J.* **85**, 1787.
- [94] E. Fadrná, N. Spacková, R. Stefl, J. Koca, T. E. Cheatham III & J. Sponer (2004) Molecular dynamics simulations of guanine quadruplex loops: Advances and force field limitations. *Biophys. J.* **87**, 227–242.
- [95] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, Jr., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell & P. A. Kollman (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **117**, 5179–5197.
- [96] P. Hobza, M. Kabeláč, J. Šponer, P. Mejzlík & J. Vondrášek (1997) Performance of empirical potentials (AMBER, CFF95, CVFF, CHARMM, OPLS, POLTEV), semiempirical quantum chemical methods (AM1, MNDO/M, PM3), and *ab initio*

- Hartree Fock method for interaction of DNA bases: Comparison with nonempirical beyond Hartree Fock results. *J. Comput. Chem.* **18**, 1136–1150.
- [97] W. D. Cornell, P. Cieplak, C. I. Bayly & P. A. Kollman (1993) Application of RESP charges to calculate conformational energies, hydrogen bond energies and free energies of solvation. *J. Am. Chem. Soc.* **115**, 9620–9631.
- [98] P. Cieplak, W. D. Cornell, C. Bayly & P. A. Kollman (1995) Application of the multimolecule and multiconformational RESP methodology to biopolymers: Charge derivation for DNA, RNA and proteins. *J. Comput. Chem.* **16**, 1357–1377.
- [99] J. R. Maple, U. Dinur & A. T. Hagler (1988) Derivation of force fields for molecular mechanics and dynamics from ab initio energy surfaces. *Proc. Natl. Acad. Sci. USA* **85**, 5350–5354.
- [100] *Molecular Simulations Inc, San Diego, CA, 1999.*
- [101] R. Elber & M. Karplus (1990) Enhanced sampling in molecular dynamics: use of the time-dependent Hartree approximation for a simulation of carbon monoxide diffusion through myoglobin. *J. Am. Chem. Soc.* **112**, 9161–9175.
- [102] C. Simmerling & R. Elber (1994) Hydrophobic "collapse" in a cyclic hexapeptide: Computer simulations of CHDLFC and CAAAAC in water. *J. Am. Chem. Soc.* **116**, 2534–2547.
- [103] X. Cheng, V. Hornak & C. Simmerling (2004) Improved conformational sampling through an efficient combination of mean-field simulation approaches. *J. Phys. Chem. B* **108**, 426–437.
- [104] R. H. Smith Jr., W. L. Jorgensen, J. Tirado-Rives, M. L. Lamb, P. A. Janssen, C. J. Michejda & M. B. Kroeger Smith (1998) Prediction of binding affinities for TIBO inhibitors of HIV-1 reverse transcriptase using Monte Carlo simulations in a linear response method. *J. Med. Chem.* **41**, 5272–5286.
- [105] J. Srinivasan, T. E. Cheatham III, P. Cieplak, P. A. Kollman & D. A. Case (1998) Continuum solvent studies of the stability of DNA, RNA and phosphoramidate-DNA helices. *J. Am. Chem. Soc.* **120**, 9401–9409.
- [106] B. Jayaram, D. Sprous, M. A. Young & D. L. Beveridge (1998) Free energy analysis of the conformational preferences of A and B forms of DNA in solution. *J. Am. Chem. Soc.* **120**, 10629–10633.

- [107] M. F. Sanner, A. J. Olson & J.-C. Spehner (1996) Reduced surface: An efficient way to compute molecular surfaces. *Biopolymers* **38**, 305–320.
- [108] D. Sitkoff, K. A. Sharp & B. Honig (1994) Accurate calculation of hydration free energies using macroscopic solvent models. *J. Phys. Chem.* **98**, 1978–1988.
- [109] G. D. Strahan, R. H. Shafer & M. A. Keniry (1994) Structural properties of the [d(G₃T₄G₃)]₂ quadruplex: evidence for sequential *syn* - *syn* deoxyguanosines. *Nucleic Acids Res.* **22**, 5447–5455.
- [110] J. Töhl & W. Eimer (1997) Interaction of a G-DNA quadruplex with mono- and divalent cations. A force field calculation. *Biophys. Chem.* **67**, 177–186.
- [111] V. M. Marathias & P. H. Bolton (1999) Determinants of DNA quadruplex structural type: Sequence and potassium binding. *Biochemistry* **38**, 4355–4364.
- [112] V. M. Marathias & P. H. Bolton (2000) Structures of the potassium-saturated, 2:1, and intermediate, 1:1, forms of a quadruplex DNA. *Nucleic Acids Res.* **28**, 1969–1977.
- [113] K. Padmanabhan & A. Tulinsky (1996) An ambiguous structure of a DNA 15-mer thrombin complex. *Acta Crystallogr.* **D52**, 272–282.
- [114] J. Srinivasan, J. Miller, P. A. Kollman & D. A. Case (1998) Continuum solvent studies of the stability of RNA hairpin loops and helices. *J. Biomol. Struct. Dyn.* **16**, 671–682.
- [115] M. R. Lee, Y. Duan & P. A. Kollman (2000) Use of MM-PB/SA in estimating the free energies of proteins: Application to native, intermediates, and unfolded villin headpiece. *Proteins* **39**, 309–316.
- [116] M. R. Lee, D. Baker & P. A. Kollman (2001) 2.1 and 1.8 Å average C_α RMSD structure predictions on two small proteins, HP-36 and S15. *J. Am. Chem. Soc.* **123**, 1040–1046.
- [117] M. R. Lee, J. Tsai, D. Baker & P. A. Kollman (2001) Molecular dynamics in the endgame of protein structure prediction. *J. Mol. Biol.* **313**, 417–430.
- [118] M. R. Lee & P. A. Kollman (2001) Free-energy calculations highlight differences in accuracy between X-ray and NMR structures and add value to protein structure prediction. *Structure* **9**, 905–916.

- [119] V. Hornak & C. Simmerling (2003) Generation of accurate protein loop conformations through low-barrier molecular dynamics. *Proteins* **51**, 577–590.
- [120] C. Simmerling, J. L. Miller & P. A. Kollman (1998) Combined locally enhanced sampling and particle mesh Ewald as a strategy to locate the experimental structure of a nonhelical nucleic acid. *J. Am. Chem. Soc.* **120**, 7149–7155.
- [121] Y. Duan & P. A. Kollman (1998) Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science* **282**, 740–744.
- [122] J. Schlitter (1993) Estimation of absolute and relative entropies of macromolecules using the covariance matrix. *Chem. Phys. Lett.* **215**, 617–621.
- [123] W. S. Ross & C. C. Hardin (1994) Ion-induced stabilization of the G-DNA quadruplex: Free energy perturbation studies. *J. Am. Chem. Soc.* **116**, 6070–6080.
- [124] J. Gu, J. Leszczynski & M. Bansal (1999) A new insight into the structure and stability of Hoogsteen hydrogen-bonded G-tetrad: an ab initio SCF study. *Chem. Phys. Lett.* **311**, 209–214.
- [125] M. Meyer, M. Brandl & J. Sühnel (2001) Are guanine tetrads stabilized by bifurcated hydrogen bonds? *J. Phys. Chem. A* **105**, 8223–8225.
- [126] T. van Mourik & A. J. Dingley (2005) Characterization of the monovalent ion position and hydrogen-bond network in guanine quartets by DFT calculations of NMR parameters. *Chem. Eur. J.* **11**, 6064–6079.
- [127] J. Gu & J. Leszczynski (2000) A remarkable alteration in the bonding pattern: An HF and DFT study of the interactions between the metal cations and the Hoogsteen hydrogen-bonded G-tetrad. *J. Phys. Chem. A* **104**, 6308–6313.
- [128] A. Wong & G. Wu (2003) Selective binding of monovalent cations to the stacking G-quartet structure formed by guanosine 5'-monophosphate: A solid state NMR study. *J. Am. Chem. Soc.* **125**, 13895–13905.
- [129] J. Gu & J. Leszczynski (2002) Origin of Na⁺/K⁺ selectivity of the guanine tetraplexes in water: The theoretical rationale. *J. Phys. Chem. A* **106**, 529–532.
- [130] M. Meyer, T. Steinke, M. Brandl & J. Sühnel (2001) Density functional study of guanine and uracil quartets and of guanine quartet/metal ion complexes. *J. Comput. Chem.* **22**, 109–124.

- [131] A. Kettani, R. A. Kumar & D. J. Patel (1995) Solution structure of a DNA quadruplex containing the fragile X syndrome triplet repeat. *J. Mol. Biol.* **254**, 638–656.
- [132] L. A. Kelley, S. P. Gardner & M. J. Sutcliffe (1996) An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally related subfamilies. *Protein Eng.* **9**, 1063–1065.
- [133] D. A. Case, D. A. Pearlman, J. W. Caldwell, T. E. I. Cheatham, J. Wang, W. S. Ross, C. L. Simmerling, T. A. Darden, K. M. Merz, R. V. Stanton, A. L. Cheng, J. J. Vincent, M. Crowley, V. Tsui, H. Gohlke, R. J. Radmer, Y. Duan, J. Pitera, I. Massova, G. L. Seibel, U. C. Singh, P. K. Weiner & P. A. Kollman (2002). AMBER 7. University of California, San Francisco.
- [134] *Delphi II, BIOSYM: San Diego, CA, 1998.*
- [135] A. Bondi (1964) Van der Waals volumes and radii. *J. Phys. Chem.* **68**, 441–451.
- [136] J. Åqvist (1990) Ion-water interaction potentials derived from free energy perturbation simulations. *J. Phys. Chem.* **94**, 8021–8024.
- [137] P. Emsley & K. Cowtan (2004) *Coot*: Model-building tools for molecular graphics. *Acta Crystallogr.* **D60**, 2126–2132.
- [138] E. A. Merritt & D. J. Bacon (1997) Raster3D: Photorealistic molecular graphics. *Methods in Enzymology* **277**, 505–524.
- [139] W. Humphrey, A. Dalke & K. Schulten (1996) VMD - Visual molecular dynamics. *J. Molec. Graphics* **14**, 33–38.
- [140] *Persistence of Vision Raytracer Pty. Ltd., 1994.*
- [141] <http://www.gnome.org/projects/dia/>.
- [142] T. R. Schneider & G. M. Sheldrick (2002) Substructure solution with *SHELXD*. *Acta Crystallogr.* **D58**, 1772–1779.
- [143] G. N. Murshudov, A. A. Vagin & E. J. Dodson (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr.* **D53**, 240–255.
- [144] C. R. Kissinger, D. K. Gehlhaar, B. A. Smith & D. Bouzida (2001) Molecular replacement by evolutionary search. *Acta Crystallogr.* **D57**, 1474–1479.

- [145] A. T. Brunger, P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J. S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson & G. L. Warren (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr.* **D54**, 905–921.
- [146] P. Hazel, J. Huppert, S. Balasubramanian & S. Neidle (2004) Loop-length-dependent folding of G-quadruplexes. *J. Am. Chem. Soc.* **126**, 16405–16415.
- [147] H. Gohlke & D. A. Case (2004) Converging free energy estimates: MM-PB(GB)SA studies on the protein-protein complex Ras-Raf. *J. Comput. Chem.* **25**, 238–250.
- [148] L. Ying, J. J. Green, H. Li, D. Klenerman & S. Balasubramanian (2003) Studies on the structure and dynamics of the human telomeric G-quadruplex by single-molecule fluorescence resonance energy transfer. *Proc. Natl. Acad. Sci. USA* **100**, 14629–14634.
- [149] N. Jing, C. Marchand, J. Liu, R. Mitra, M. E. Hogan & Y. Pommier (2000) Mechanism of inhibition of HIV-1 integrase by G-tetrad-forming oligonucleotides *in vitro*. *J. Biol. Chem.* **275**, 21460–21467.
- [150] S. M. Haider, G. N. Parkinson & S. Neidle (2003) Structure of a G-quadruplex-ligand complex. *J. Mol. Biol.* **326**, 117–125.
- [151] G. R. Clark, P. D. Pytel, C. J. Squire & S. Neidle (2003) Structure of the first parallel DNA quadruplex-drug complex. *J. Am. Chem. Soc.* **125**, 4066–4067.
- [152] O. Y. Fedoroff, M. Salazar, H. Han, V. V. Chemeris, S. M. Kerwin & L. H. Hurley (1998) NMR-based model of a telomerase-inhibiting compound bound to G-quadruplex DNA. *Biochemistry* **37**, 12367–12374.
- [153] M. J. Cocco, L. A. Hanakahi, M. D. Huber & N. Maizels (2003) Specific interactions of distamycin with G-quadruplex DNA. *Nucleic Acids Res.* **31**, 2944–2951.
- [154] A. K. Mehta, Y. Shayo, H. Vankayalapati, L. H. Hurley & J. Schaefer (2004) Structure of Quinobenzoxazine-G-quadruplex complex by REDOR NMR. *Biochemistry* **43**, 11953–11958.
- [155] H. Arthanari & P. H. Bolton (1999) Porphyrins can catalyze the interconversion of DNA quadruplex structural types. *Anticancer Drug Des.* **14**, 317–326.

- [156] E. M. Rezler, J. Seenisamy, S. Bashyam, M.-Y. Kim, E. White, W. D. Wilson & L. H. Hurley (2005) Telomestatin and diseleno sapphyrin bind selectively to two different forms of the human telomeric G-quadruplex structure. *J. Am. Chem. Soc.* **127**, 9439–9447.
- [157] M. A. Read, A. A. Wood, J. R. Harrison, S. M. Gowan, L. R. Kelland, H. S. Dosanjh & S. Neidle (1999) Molecular modeling studies on G-quadruplex complexes of telomerase inhibitors: Structure-activity relationships. *J. Med. Chem.* **42**, 4538–4546.
- [158] M. A. Read & S. Neidle (2000) Structural characterization of a guanine-quadruplex ligand complex. *Biochemistry* **39**, 13422–13432.
- [159] M. Read, R. J. Harrison, B. Romagnoli, F. A. Tanious, S. H. Gowan, A. P. Reszka, W. D. Wilson, L. R. Kelland & S. Neidle (2001) Structure-based design of selective and potent G-quadruplex-mediated telomerase inhibitors. *Proc. Natl. Acad. Sci. USA* **98**, 4844–4849.
- [160] I. Haq, J. O. Trent, B. Z. Chowdhry & T. C. Jenkins (1999) Intercalative G-tetraplex stabilization of telomeric DNA by a cationic porphyrin. *J. Am. Chem. Soc.* **121**, 1768–1779.
- [161] P. Cieplak (2002) Application of the free energy calculations to study drug-enzyme and drug-DNA complexes. *Mol. Simul.* **28**, 173–186.
- [162] N. Spacková, T. E. Cheatham III, F. Ryjacek, F. Lankas, L. van Meervelt, P. Hobza & J. Sponer (2003) Molecular dynamics simulations and thermodynamics analysis of DNA-drug complexes. Minor groove binding between 4',6-diamidino-2-phenylindole and DNA duplexes in solution. *J. Am. Chem. Soc.* **125**, 1759–1769.
- [163] S. Yan, M. Wu, D. J. Patel, N. E. Geacintov & S. Broyde (2003) Simulating structural and thermodynamic properties of carcinogen-damaged DNA. *Biophys. J.* **84**, 2137–2148.
- [164] O. A. T. Donini & P. A. Kollman (2000) Calculation and prediction of binding free energies for the matrix metalloproteinases. *J. Med. Chem.* **43**, 4180–4188.
- [165] S. Huo, J. Wang, P. Cieplak, P. A. Kollman & I. D. Kuntz (2002) Molecular dynamics and free energy analyses of cathepsin D-inhibitor interactions: Insight into structure-based ligand design. *J. Med. Chem.* **45**, 1412–1419.

- [166] L. T. Chong, Y. Duan, L. Wang, I. Massova & P. A. Kollman (1999) Molecular dynamics and free-energy calculations applied to affinity maturation in antibody 48G7. *Proc. Natl. Acad. Sci. USA* **96**, 14330–14335.
- [167] B. Kuhn, P. Gerber, T. Schulz-Gasch & M. Stahl (2005) Validation and use of the MM-PBSA approach for drug discovery. *J. Med. Chem.* **48**, 4040–4048.
- [168] B. Guyen, C. M. Schultes, P. Hazel, J. Mann & S. Neidle (2004) Synthesis and evaluation of analogues of 10*H*-indolo[3,2-*b*]-quinoline as G-quadruplex stabilising ligands and potential inhibitors of the enzyme telomerase. *Org. Biomol. Chem.* **2**, 981–988.
- [169] J.-M. Zhou, X.-F. Zhu, Y.-J. Lu, R. Deng, Z.-S. Huang, Y.-P. Mei, Y. Wang, W.-L. Huang, Z.-C. Liu, L.-Q. Gu & Y.-X. Zeng (2005) Senescence and telomere shortening induced by novel potent G-quadruplex interactive agents, quindoline derivatives, in human cancer cell lines. *Oncogene* Advance online publication.
- [170] *Insight II manual*.
- [171] M. W. Schmidt, K. K. Baldrige, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis & J. A. Montgomery, Jr (1993) General atomic and molecular electronic structure system GAMESS. *J. Comput. Chem.* **14**, 1347–1363.
- [172] C. C. Hardin, A. G. Perry & K. White (2001) Thermodynamic and kinetic characterization of the dissociation and assembly of quadruplex nucleic acids. *Biopolymers* **56**, 147–194.
- [173] N. Gresh, J. E. Sponer, N. Spacková, J. Leszczynski & J. Sponer (2003) Theoretical study of binding of hydrated Zn(II) and Mg(II) cations to 5'-guanosine monophosphate. Toward polarizable molecular mechanics for DNA and RNA. *J. Phys. Chem. B* **107**, 8669–8681.
- [174] M. Meyer & J. Sühnel (2003) Interaction of cyclic cytosine-, guanine-, thymine-, uracil- and mixed guanine-cytosine base tetrads with K⁺, Na⁺ and Li⁺ ions - a density functional study. *J. Biomol. Struct. Dyn.* **20**, 507–518.

Appendix A

Clustering methodology

Clustering was carried out according to the method used by the NMRCLUST program¹³², which is based on a hierarchical clustering algorithm. A Python script was written to enable the clustering of an unlimited number of structures. Unlike the NMRCLUST program, this script does not calculate the rms deviations, and they must be input from a file. Pairwise rms deviations were obtained from Insight II in this work.

The rms distance $\text{dist}(m, n)$ between each pair of clusters m and n containing X_m and X_n structures i and j , is calculated according to:

$$\text{dist}(m, n) = \frac{[\sum_{i=1}^{X_m} \sum_{j=1}^{X_n} \text{dist}(i, j)]}{X_m X_n}$$

At each step, the two closest clusters are merged into a single cluster. This is repeated until all the clusters have been merged. The “spread” of each cluster is also calculated at each step. In this case, i and k are structures of the same cluster m , which contains X_m members.

$$\text{spread}_m = \frac{[\sum_{k=1}^{X_m} \sum_{i=1, i>k}^{X_m} \text{dist}(i, k)]}{X_m(X_m - 1)/2}$$

The average spread of all clusters at stage i , avgSpread_i , is normalised for use in the penalty function calculation. cnum_i is the number of clusters at stage i , $\text{max}(\text{avgSpread})$ and $\text{min}(\text{avgSpread})$ are the maximum and minimum values of avgSpread_i , and N is the total number of structures which are being clustered.

$$\text{avgSpread}_i = \frac{\sum_{m=1}^{\text{cnum}_i} \text{spread}_m}{\text{cnum}_i}$$

$$\text{avgSpread}(\text{norm})_i = \frac{(N - 2)(\text{avgSpread}_i - \min(\text{avgSpread}))}{\max(\text{avgSpread}) - \min(\text{avgSpread})} + 1$$

The penalty function is calculated at each step, according to:

$$\text{penalty}_i = \text{avgSpread}(\text{norm})_i + \text{cnum}_i$$

The optimum number of clusters is determined by the lowest penalty value, corresponding to the most populated clusters possible with the smallest spread within each cluster. The Python script written also includes the option to define a threshold, so that clusters will not contain conformations which have greater rms deviations than the set threshold.

Generally, when clustering loop conformations, the penalty calculation was found to provide good cluster distributions, with rms deviations around 1.0 to 1.5 Å between clusters. However, when few similar structures were obtained in simulated annealing runs, the clusters often contained structures which were judged too dissimilar (by visual inspection of the clusters). In the latter case, an rms deviation threshold of between 1.0 and 1.5 Å was set explicitly.

The clustering script can be downloaded from <http://www.pascalehazel.org/cluster/>.

Appendix B

Ligand charges

B.1 JH ligand charges

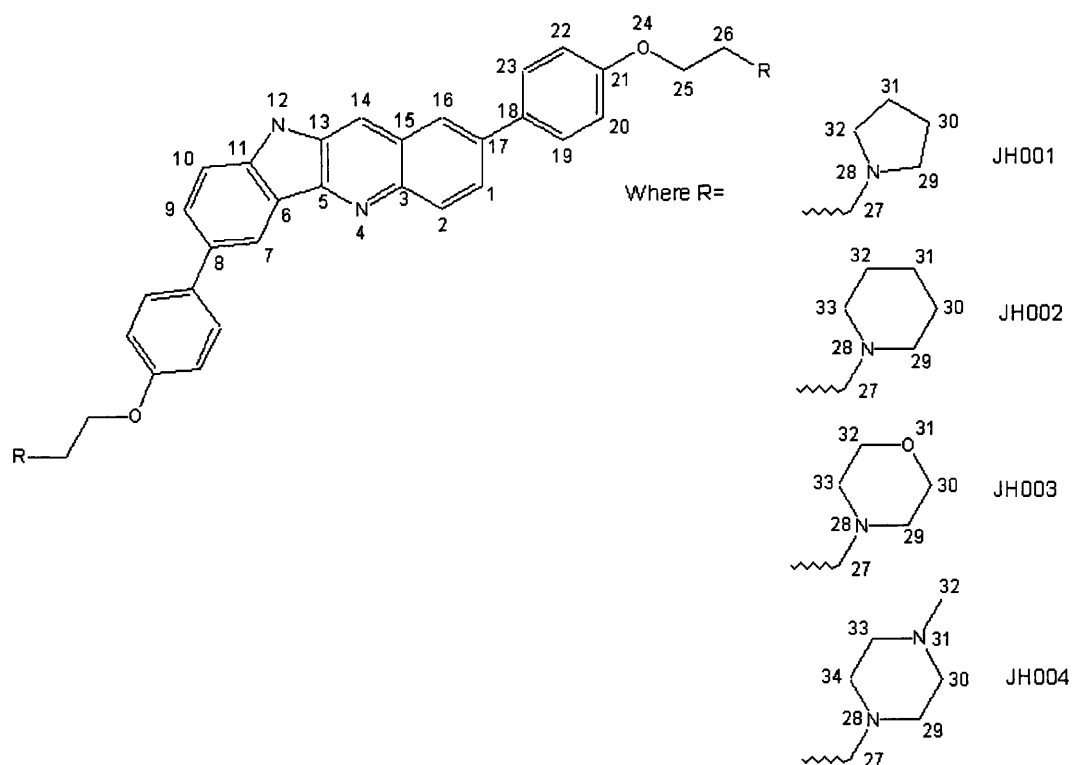


Figure B.1: JH ligand numbering. The numbering scheme is arbitrary. The two side chains were constrained to have equivalent charges, and only one is numbered.

B.1.1 JH001

Table B.1: JH001 atomic charges calculated using RESP after a 6-31G* minimisation and single point calculation. Atom numbers are shown in Figure B.1.

atom	gaff atom type	charge	atom	gaff atom type	charge
C1	ca	-0.160	C19	ca	-0.140
H1	ha	0.150	H19	ha	0.151
C2	ca	-0.270	C20	ca	-0.196
H2	ha	0.170	H20	ha	0.136
C3	ca	0.452	C21	ca	0.192
N4	nb	-0.614	C22	ca	-0.196
C5	cp	0.417	H22	ha	-0.136
C6	cp	-0.041	C23	ca	-0.140
C7	ca	-0.139	H23	ha	0.151
H7	ha	0.151	O24	os	-0.270
C8	cp	0.002	C25	c3	0.026
C9	ca	-0.150	1,2H25	h1	0.067
H9	ha	0.133	C26	c3	-0.010
C10	ca	-0.203	1,2H26	hc	0.083
H10	ha	0.164	C27	c3	-0.381
C11	ca	0.093	1,2H27	hx	0.203
N12	na	-0.394	N28	n4	-0.006
H12	hn	0.257	H28	hn	0.291
C13	ca	0.027	C29	c3	-0.124
C14	ca	-0.225	1,2H29	hx	0.130
H14	ha	0.184	C30	c3	0.003
C15	ca	-0.042	1,2H30	hc	0.055
C16	ca	-0.191	C31	c3	0.003
H16	ha	0.132	1,2H31	hc	0.055
C17	cp	0.022	C32	c3	-0.124
C18	cp	0.036	1,2H32	hx	0.130

B.1.2 JH002

Table B.2: JH002 atomic charges calculated using RESP after a 6-31G* minimisation and single point calculation. Atom numbers are shown in Figure B.1.

atom	gaff atom type	charge	atom	gaff atom type	charge
C1	ca	-0.158	H19	ha	0.149
H1	ha	0.146	C20	ca	-0.178
C2	ca	-0.251	H20	ha	0.144
H2	ha	0.165	C21	ca	0.160
C3	ca	0.437	C22	ca	-0.178
N4	nb	-0.608	H22	ha	0.144
C5	cp	0.405	C23	ca	-0.147
C6	cp	-0.048	H23	ha	0.149
C7	ca	-0.126	O24	os	-0.287
H7	ha	0.148	C25	c3	0.044
C8	cp	0.005	1,2H25	h1	0.063
C9	ca	-0.148	C26	c3	-0.013
H9	ha	0.134	1,2H26	hc	0.085
C10	ca	-0.205	C27	c3	-0.378
H10	ha	0.163	1,2H27	hx	0.197
C11	ca	0.100	N28	n4	-0.056
N12	na	-0.389	H28	hn	0.320
H12	hn	0.355	C29	c3	-0.155
C13	ca	0.030	1,2H29	hx	0.146
C14	ca	-0.225	C30	c3	-0.040
H14	ha	0.181	1,2H30	hc	0.057
C15	ca	-0.030	C31	c3	-0.014
C16	ca	-0.193	1,2H31	hc	0.058
H16	ha	0.140	C32	c3	-0.040
C17	cp	0.003	1,2H32	hc	0.057
C18	cp	0.060	C33	c3	-0.155
C19	ca	-0.147	1,2H33	hx	0.146

B.1.3 JH003

Table B.3: JH003 atomic charges calculated using RESP after a 6-31G* minimisation and single point calculation. Atom numbers are shown in Figure B.1.

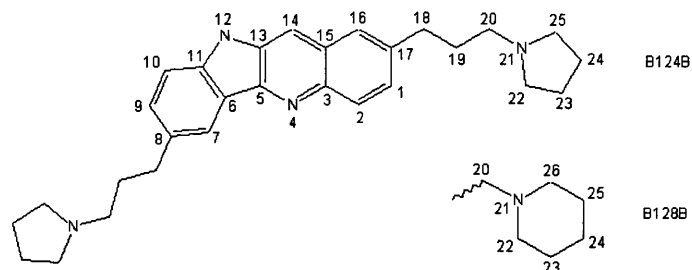
atom	gaff atom type	charge	atom	gaff atom type	charge
C1	ca	-0.133	C19	ca	-0.165
H1	ha	0.138	H19	ha	0.138
C2	ca	-0.255	C20	ca	-0.140
H2	ha	0.161	H20	ha	0.139
C3	ca	0.386	C21	ca	0.138
N4	nb	-0.588	C22	ca	-0.140
C5	cp	0.420	H22	ha	0.139
C6	cp	-0.049	C23	ca	-0.165
C7	ca	-0.134	H23	ha	0.138
H7	ha	0.139	O24	os	-0.313
C8	cp	0.008	C25	c3	0.056
C9	ca	-0.144	1,2H25	h1	0.051
H9	ha	0.144	C26	c3	0.049
C10	ca	-0.211	1,2H26	hc	0.010
H10	ha	0.152	C27	c3	-0.025
C11	ca	0.098	1,2H27	hx	0.048
N12	na	-0.419	N28	n4	-0.334
H12	hn	0.363	C29	c3	-0.039
C13	ca	0.033	1,2H29	h1	0.076
C14	ca	-0.224	C30	c3	0.076
H14	ha	0.175	1,2H30	h1	0.060
C15	ca	-0.026	O31	os	-0.392
C16	ca	-0.202	C32	c3	0.076
H16	ha	0.140	1,2H32	h1	0.060
C17	cp	0.022	C33	c3	-0.039
C18	cp	0.034	1,2H33	h1	0.076

B.1.4 JH004

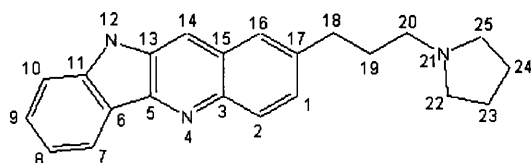
Table B.4: JH004 atomic charges calculated using RESP after a 6-31G* minimisation and single point calculation. Atom numbers are shown in Figure B.1.

atom	gaff atom type	charge	atom	gaff atom type	charge
C1	ca	-0.150	C20	ca	-0.164
H1	ha	0.145	H20	ha	0.143
C2	ca	-0.266	C21	ca	0.091
H2	ha	0.169	C22	ca	-0.164
C3	ca	0.453	H22	ha	0.143
N4	nb	-0.642	C23	ca	-0.138
C5	cp	0.466	H23	ha	0.146
C6	cp	-0.054	O24	os	-0.256
C7	ca	-0.152	C25	c3	-0.020
H7	ha	0.152	1,2H25	h1	0.102
C8	cp	0.020	C26	c3	-0.033
C9	ca	-0.138	1,2H26	hc	0.087
H9	ha	0.143	C27	c3	-0.360
C10	ca	-0.219	1,2H27	hx	0.187
H10	ha	0.171	N28	n4	-0.026
C11	ca	0.085	H28	hn	0.322
N12	na	-0.413	C29	c3	-0.129
H12	hn	0.372	1,2H29	hx	0.146
C13	ca	0.043	C30	c3	-0.116
C14	ca	-0.258	1,2H30	h1	0.133
H14	ha	0.192	N31	n3	-0.363
C15	ca	-0.012	C32	c3	-0.005
C16	ca	-0.197	1,2,3H32	h1	0.077
H16	ha	0.138	C33	c3	-0.116
C17	cp	-0.007	1,2H33	h1	0.133
C18	cp	0.067	C34	c3	-0.129
C19	ca	-0.138	1,2H34	hx	0.146
H19	ha	0.146			

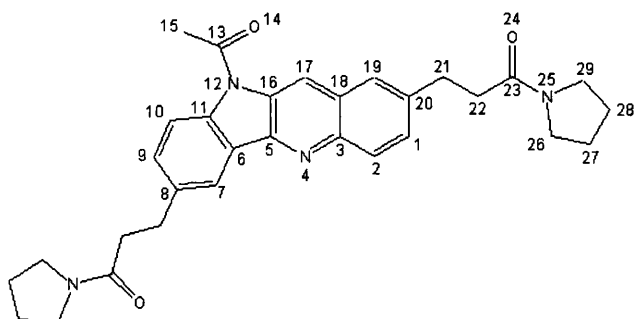
B.2 B ligand charges



(a) B124B and B128B



(b) B120B



(c) B123B

Figure B.2: B ligand numbering. The numbering scheme is arbitrary. The two side chains were constrained to have equivalent charges, and only one is numbered.

B.2.1 B124B

Table B.5: B124B atomic charges calculated using RESP after a 6-31G* minimisation and single point calculation. Atom numbers are shown in Figure B.2.

atom	gaff atom type	charge	atom	gaff atom type	charge
C1	ca	-0.226	C15	ca	-0.050
H1	ha	0.166	C16	ca	-0.216
C2	ca	-0.198	H16	ha	0.160
H2	ha	0.160	C17	ca	-0.001
C3	ca	0.424	C18	c3	-0.083
N4	nb	-0.615	1,2H18	hc	0.083
C5	cp	0.407	C19	c3	-0.002
C6	cp	-0.056	1,2H19	hc	0.033
C7	ca	-0.177	C20	c3	-0.181
H7	ha	0.159	1,2H20	hx	0.118
C8	cp	-0.003	N21	n4	0.004
C9	ca	-0.152	H21	hn	0.297
H9	ha	0.149	C22	c3	-0.122
C10	ca	-0.257	1,2H22	hx	0.131
H10	ha	0.178	C23	c3	-0.016
C11	ca	0.235	1,2H23	hc	0.060
N12	na	-0.591	C24	c3	-0.016
H12	hn	0.422	1,2H24	hc	0.060
C13	ca	0.118	C25	c3	-0.122
C14	ca	-0.204	1,2H25	hx	0.131
H14	ha	0.176			

B.2.2 B128B

Table B.6: B128B atomic charges calculated using RESP after a 6-31G* minimisation and single point calculation. Atom numbers are shown in Figure B.2.

atom	gaff atom type	charge	atom	gaff atom type	charge
C1	ca	-0.185	C16	ca	-0.202
H1	ha	0.140	H16	ha	0.160
C2	ca	-0.206	C17	ca	0.005
H2	ha	0.156	C18	c3	-0.070
C3	ca	0.416	1,2H18	hc	0.073
N4	nb	-0.609	C19	c3	-0.025
C5	cp	0.430	1,2H19	hc	0.040
C6	cp	-0.055	C20	c3	-0.111
C7	ca	-0.178	1,2H20	hx	0.106
H7	ha	0.157	N21	n4	-0.019
C8	cp	-0.013	H21	hn	0.278
C9	ca	-0.171	C22	c3	-0.086
H9	ha	0.168	1,2H22	hx	0.105
C10	ca	-0.232	C23	c3	-0.018
H10	ha	0.171	1,2H23	hc	0.053
C11	ca	0.199	C24	c3	-0.026
N12	na	-0.525	1,2H24	hc	0.050
H12	hn	0.411	C25	c3	-0.018
C13	ca	0.059	1,2H25	hc	0.053
C14	ca	-0.184	C26	c3	-0.086
H14	ha	0.171	1,2H26	hx	0.105
C15	ca	-0.065			

B.2.3 B120B

Table B.7: B120B atomic charges calculated using RESP after a 6-31G* minimisation and single point calculation. Atom numbers are shown in Figure B.2.

atom	gaff atom type	charge	atom	gaff atom type	charge
C1	ca	-0.213	H14	ha	0.175
H1	ha	0.156	C15	ca	-0.040
C2	ca	-0.224	C16	ca	-0.223
H2	ha	0.165	H16	ha	0.152
C3	ca	0.438	C17	ca	-0.010
N4	nb	-0.624	C18	c3	-0.031
C5	cp	0.430	1,2H18	hc	0.056
C6	cp	-0.015	C19	c3	0.023
C7	ca	-0.164	1,2H19	hc	0.020
H7	ha	0.158	C20	c3	-0.144
C8	ca	-0.145	1,2H20	hx	0.105
H8	ha	0.142	N21	n4	0.024
C9	ca	-0.134	H21	hn	0.285
H9	ha	0.156	C22	c3	-0.125
C10	ca	-0.205	1,2H22	hx	0.130
H10	ha	0.155	C23	c3	-0.015
C11	ca	0.131	1,2H23	hc	0.058
N12	na	-0.469	C24	c3	-0.015
H12	hn	0.382	1,2H24	hc	0.058
C13	ca	0.062	C25	c3	-0.125
C14	ca	-0.229	1,2H25	hx	0.130

B.2.4 B123B

Table B.8: B123B atomic charges calculated using RESP after a 6-31G* minimisation and single point calculation. Atom numbers are shown in Figure B.2.

atom	gaff atom type	charge	atom	gaff atom type	charge
C1	ca	-0.184	C17	ca	-0.152
H1	ha	0.162	H17	ha	0.191
C2	ca	-0.233	C18	ca	-0.025
H2	ha	0.151	C19	ca	-0.243
C3	ca	0.405	H19	ha	0.164
N4	nb	-0.565	C20	ca	-0.012
C5	cp	0.357	C21	c3	-0.012
C6	cp	-0.027	1,2H21	hc	0.035
C7	ca	-0.103	C22	c3	0.135
H7	ha	0.148	1,2H22	hc	-0.041
C8	cp	-0.017	C23	c	0.441
C9	ca	-0.166	O24	o	-0.574
H9	ha	0.147	N25	n	-0.137
C10	ca	-0.179	C26	c3	-0.022
H10	ha	0.133	1,2H26	h1	0.057
C11	ca	-0.007	C27	c3	-0.048
N12	na	-0.068	1,2H27	hc	0.035
C13	c	0.558	C28	c3	-0.048
O14	o	-0.545	1,2H28	hc	0.035
C15	c3	-0.120	C29	c3	-0.022
1,2,3H15	hc	0.049	1,2H29	h1	0.057
C16	ca	-0.056			