

Drug-targetting of duplex and quadruplex DNA

by Evripidis Gavathiotis, BSc

Thesis submitted to the University of Nottingham for the
degree of Doctor of Philosophy, July 2002

Contents

Abstract.....	i
Acknowledgements.....	ii
Abbreviations.....	iii
Publications.....	v
1. Co-operative interactions in DNA recognition	1
1.1 Introduction.....	1
1.2 Co-operativity in protein–DNA interactions.....	3
1.3 Co-operativity in drug–DNA interactions.....	7
1.3.1 Co-operative binding in DNA intercalation.....	8
1.3.2 Co-operative binding in the DNA minor groove.....	12
1.4 Conclusions.....	18
1.5 References.....	20
2. Structural investigations on DNA recognition by Hoechst 33258	23
2.1 Introduction.....	23
2.2 Materials and methods.....	29
2.2.1 DNA synthesis.....	29
2.2.2 Sample preparation.....	29
2.2.3 NMR analysis.....	30
2.2.4 NOE restraints.....	31
2.2.5 Structure calculations of d(CTTTTGCAAAAG) ₂ and (GAAAAGCTTTTC) ₂	33
2.2.6 Structure calculations of 2:1 H33258-d(CTTTTGCAAAAG) ₂ complex.....	36
2.3 Results and Discussion.....	38
2.3.1 NMR assignments of the free d(CTTTTGCAAAAG) ₂	38
2.3.2 NMR titration of H33258 to d(CTTTTGCAAAAG) ₂	41
2.3.3 NMR assignment of the 2:1 complex.....	43
2.3.4 Structure of the 2:1 complex.....	49
2.3.5 Conformation and dynamics of bound and free d(CTTTTGCAAAAG) ₂	53
2.3.6 Evidence for co-operative A-tract stabilisation with dynamic GC junctions.....	60
2.3.7 Complex of d(GAAAAGCTTTTC) ₂ with Hoechst 33258.....	62

2.4 Conclusions.....	66
2.5 References.....	67
3. Co-operative interaction of Hoechst 33258 within the DNA minor groove	72
3.1 Introduction.....	72
3.2 Hoechst 33258 co-operative binding to DNA.....	73
3.3 Methods for detecting co-operativity in ligand-DNA recognition.....	75
3.3.1 NMR titration.....	75
3.3.2 Circular dichroism titration.....	76
3.3.3 Footprinting analysis.....	78
3.4 Materials and methods.....	80
3.4.1 Fluorescence spectroscopy.....	80
3.4.2 CD experiments.....	82
3.4.3 NMR experiments.....	83
3.4.4 Molecular dynamics.....	83
3.5 Results and discussion.....	84
3.5.1 Fluorescence studies.....	84
3.5.2 CD analysis.....	86
3.5.3 NMR binding studies of H33258.....	90
3.5.4 Footprinting analysis.....	99
3.5.5 Molecular dynamics analysis.....	103
3.6 Conclusions.....	112
3.7 References.....	115
4. DNA minor groove recognition by a bis-benzimidazole analogue of Hoechst 33258	117
4.1 Introduction.....	117
4.2 Materials and methods.....	120
4.2.1 Sample preparation and analysis.....	120
4.2.2 Structure calculations.....	121
4.3 Results and discussion.....	122
4.3.1 Structural analysis by NMR.....	122
4.3.2 Structure of the 1:1 H10–d(GCAAATTTGC) ₂ complex.....	129
4.4 Conclusions.....	138
4.5 References.....	139

5. Investigations into the structure of quadruplex DNA	141
5.1 Introduction.....	141
5.2 Materials and methods.....	148
5.2.1 DNA samples.....	148
5.2.2 NMR experiments.....	148
5.2.3 NMR restraints.....	149
5.2.4 Structure calculations.....	150
5.2.5 Analysis of thermodynamic parameters.....	153
5.3 Results and Discussion.....	154
5.3.1 NMR spectral features.....	154
5.3.2 Exchangeable protons assignments.....	158
5.3.3 Non-exchangeable protons assignments.....	164
5.3.4 NMR restraints.....	167
5.3.5 Restrained molecular dynamics calculations.....	168
5.3.6 Quadruplex structure.....	170
5.3.7 Distinctive features of d(TTAGGGT) ₄ structure.....	175
5.4 Conclusions.....	179
5.5 References.....	181
6. Recognition and stabilisation of quadruplex DNA by a potent new telomerase inhibitor	186
6.1 Introduction.....	186
6.2 Materials and Methods.....	191
6.2.1 Sample preparation.....	191
6.2.2 NMR experiments.....	192
6.2.3 NMR restraints.....	193
6.2.4 Structure calculations.....	194
6.3 Results and Discussion.....	196
6.3.1 1D NMR titration and spectral features.....	196
6.3.2 NMR analysis of the 2:1 complex.....	200
6.3.3 Molecular modelling of the 2:1 complex.....	209
6.4 Conclusions.....	216
6.5 References.....	218
Appendices	221
Appendix 1.....	221
Appendix 2.....	228
Appendix 3.....	247

Abstract

This thesis investigates structural and dynamic properties of drug recognition mechanisms to duplex and quadruplex DNA using primarily high field NMR techniques and molecular dynamics simulations.

The mechanism of co-operative binding of Hoechst 33258 to the DNA minor groove of duplexes that contain two binding sites such as d(CTTTTGCAAAG)₂, d(GAAAAGCTTTC)₂ and d(CTTTTGGCCAAAAG)₂ has been studied. NMR and other titration techniques have evidenced co-operative binding and no detection of an intermediate 1:1 complex. High-resolution NMR structure determination showed no evidence of direct contact between Hoechst 33258 molecules or DNA structure deformation that would facilitate co-operativity. Molecular dynamics simulations based on NMR data, allowed us to calculate thermodynamic quantities of the two binding events, and lead us to conclude that ligand binding can induce changes in DNA conformational flexibility in sites of the structure distant from the binding site and result in more favourable second ligand binding. The results highlight the general importance of flexibility in determining the properties of ligand–DNA interactions.

The relative importance of ligand isohelicity and phasing in DNA minor groove has been investigated by studying the structure and dynamics of the 1:1 complex of Hoechst 10–d(GCAAATTTGC)₂. The results suggest that DNA sequence-dependent structure and flexibility have significant role for the strong binding of Hoechst 10 to the duplex.

The formation, stability, structure and dynamics of the d(TTAGGGT)₄ quadruplex structure, which contains the human telomeric repeat TTAGGG, have been studied. Characteristic features of the quadruplex structure were determined and this information was used for understanding drug–quadruplex interactions. The complex of the fluorinated polycyclic methylacridinium cation RHPS4, lead compound for telomerase inhibition, with the d(TTAGGGT)₄ quadruplex structure has been investigated. RHPS4 forms a stable G-quadruplex complex by end-stacking externally to the G-tetrads of the ApG and GpT steps. This study presents detailed properties of the complex and provides further information for lead optimisation studies.

Acknowledgements

I would like to thank all of those who without their time and support throughout the course of this project the completion of this thesis would not have been possible.

First and most important, my supervisor Dr Mark Searle for his continual support and advice, his scientific guidance and many ideas were invaluable in both the work and the writing of the thesis.

All the members of the Searle group during 1998-2001, but mostly Dr Huw Williams and Dr Muriel Jordan for their time and help with technical aspects of the work, and all the people I have worked with in labs B56 and B52 for creating a supportive and pleasant working environment.

Dr Charlie Laughton and Dr Sarah Harris of the Cancer Research Laboratories for their valuable contribution and useful conversations on Hoechst 33258 project.

Dr Gary Sharman of AstraZeneca Pharmaceuticals for his critical comments on Hoechst 33258 project and also the constructive and enjoyable time during my visits to AstraZeneca.

Prof Malcolm Stevens and Dr Robert Heald for giving me the opportunity to work on the RHPS4 project.

Dr Henriëtte Willems for her help on proof reading this thesis.

Finally, I am grateful to my family for their love and encouragement without whom I would never have enjoyed so many opportunities in my life, and also all my friends that supported me and made my time during PhD study enjoyable.

I would like to acknowledge the financial support from AstraZeneca Pharmaceuticals, EPSRC and the University of Nottingham.

Abbreviations

DNA	deoxyribonucleic acid
RNA	ribonucleic acid
A-DNA	A conformation DNA
B-DNA	B conformation DNA
A	adenine
T	thymine
G	guanine
C	cytosine
I	inosine
pur	purine
pyr	pyrimidine
NOE	Nuclear Overhauser Effect
NMR	Nuclear Magnetic Resonance
NOESY	Nuclear Overhauser Effect Spectroscopy
COSY	Correlation Spectroscopy
DQF-COSY	Double Quantum Filtered Correlation Spectroscopy
TOCSY	Total Correlation Spectroscopy
HMQC	Heteronuclear Multiple Quantum Coherence
MD	molecular dynamics
rMD	restrained molecular dynamics
PME	particle mesh Ewald
CD	circular dichroism spectroscopy
UV	ultra violet spectroscopy

ITC	isothermal titration calorimetry
HPLC	High-Pressure Liquid Chromatography
H33258	Hoechst 33258
H10	Hoechst 10
ph	phenyl
bz	benzimidazole
pip	piperazine
RMSD	root mean squared deviation
T _m	melting temperature
TEAA	triethylamine acetate
EDTA	ethylenediaminetetraacetic acid
MIP	molecular interaction potential
GB/SA	generalised Born / surface area
dGTP	2'-deoxyguanosine 5'-triphosphate
P	pseudo rotational phase angle
x-disp	x-displacement
RESP	restrained electrostatic potential
PIE	polydenylation inhibition element
K	Kelvin degrees
sd	standard deviation

Publications

The following publications have appeared from this thesis:

Gavathiotis, E., Sharman, G. J. & Searle, M. S. 2000 Sequence-Dependent Variation in DNA Minor Groove Width Dictates Orientational Preference of Hoechst 33258 in A-tract Recognition: Solution NMR Structure of the 2:1 Complex with d(CTTTTGCAAAAG)₂. *Nucleic Acids Res.* **28**, 728-735.

Harris, S. A., Gavathiotis, E., Searle, M. S., Orozco, M. & Laughton, C. A. 2001 Cooperativity in Drug-DNA Recognition: A Molecular Dynamics Study. *J. Am. Chem. Soc.* **123**, 12658-12663.

Gavathiotis, E., Heald, R. A., Stevens, M. F. G. & Searle, M. S. 2001 Recognition and Stabilization of Quadruplex DNA by a Potent New Telomerase Inhibitor: NMR Studies of the 2:1 Complex of a Pentacyclic Methylacridinium Cation with (TTAGGGT)₄. *Angew. Chem. Int. Ed.* **40**, 4749-4751.

1. Co-operative interactions in DNA recognition

1.1 Introduction

Co-operative binding of ligands to nucleic acids is possible when multiple ligands can bind to the same nucleic acid molecule. Co-operativity is a thermodynamic effect that reflects the influence of one bound ligand on the binding affinity of the second ligand. Co-operative interactions arise when the changes in free energy from the binding of two or more ligands to the same nucleic acid are mutually dependent. Table 1.1 shows the ligand configurations and associated free energy states for co-operative ligand binding to two linked binding sites on a DNA molecule.

In the literature co-operativity is defined in two modes: positive or negative, reflecting the influence on the binding affinity of the second ligand which is enhanced or diminished according to the interaction of the first ligand. Negative co-operativity is usually related to statistical effects that result from the overlap of potential ligand binding sites, while positive co-operativity is induced by real changes to the intrinsic binding parameters. Co-operativity can arise from ligand–ligand interactions when ligands are bound to adjacent sites on DNA. It can also occur through short-range conformational changes that are induced in the DNA molecule on ligand binding. There are also examples for some DNA binding proteins that are bound to specific sites well separated along the length of the DNA molecule. The origins of co-operativity in such systems have generally become evident after structure elucidation.

Table 1.1 General configuration and associated free energy states for ligand binding to a two-sites DNA system. Binding sites are stated by L if liganded, co-operative interactions are stated with \leftrightarrow . The total free Gibbs energy of each configuration (ΔG_i) relative to the unliganded reference state is given from the sum of the contributions from six energy changes (column 3). ΔG_1 , ΔG_2 are the intrinsic free energy changes for binding to the individual sites, ΔG_{12} is the free energy of co-operative interaction between liganded sites, defined as the difference in free energy to fill the sites simultaneously and free energy to fill them individually. In the case that the two binding sites are identical $\Delta G_1=\Delta G_2$.

Two binding sites (general case)			
Species	Binding configuration		Total free energy
	Site 1	Site 2	
1	reference state		ΔG_{i1}
2	L		ΔG_{i2}
3		L	ΔG_{i3}
4	L \leftrightarrow L		ΔG_{i4}

It is particularly interesting to understand how co-operativity operates in ligand–DNA complexes and what are the factors that contribute to the overall favourable co-operative binding interactions that result in higher sequence specificity as well as greater sensitivity to changes in ligand concentration. This chapter reviews structural investigations of specific DNA complexes with proteins and drug molecules where co-operative binding is observed. The origins of co-operative interactions are discussed, together with sequence-dependent structural features, and the role of the structural flexibility of these DNA complexes is emphasised.

1.2 Co-operativity in protein–DNA interactions

The initiation of transcription is a primary control point for the regulation of gene expression in all organisms. The regulation of transcription initiation is mediated through specific interactions between complex regulatory assemblies. The latter can be proteins that interact with one another and with arrays of DNA sites that are often separated by considerable distances (1). A common feature of these regulatory assemblies is the co-operative binding of the regulatory proteins to DNA. Co-operative interactions of proteins with DNA have been recognised to play a key role in controlling the transcription regulation by enhancing binding selectivity and also increasing sensitivity for protein concentration (2). Thus understanding the nature of co-operative interactions between protein–DNA complexes is a crucial step in understanding the molecular mechanism of gene regulation.

Homeodomain proteins have become one of the most studied eukaryotic DNA-binding proteins because of their key role in development and gene regulation (3–5). The homeodomain is a compact 60 residue DNA-binding domain that consists of three α helices folded around a hydrophobic core and a flexible N-terminal arm that becomes ordered only upon binding to DNA. The most interesting characteristic feature of many homeodomains is that they do not function in the cells as monomers, but instead bind DNA co-operatively with other homeodomain proteins, such that the affinity of the second protein for DNA is higher in the presence of the binding of the first protein (6).

One of the most studied examples of co-operative DNA-binding homeodomain proteins is the "Paired" class. The co-operative DNA binding of the "Paired" class is based entirely on the conserved 60 amino acid homeodomain, unlike other cases of co-operative DNA binding of homeoprotein dimers that do need

protein–protein interactions extrinsic to the homeodomain (7). The high-resolution crystal structure of the "Paired" class, co-operative homeodomain dimer on the DNA sequence 5'-TAATCTGATTA-3', composed of two inverted TAAT motifs, has been determined at 2.0 Å resolution (figure 1.1) revealing the structural basis for the observed co-operative binding (8).

This study shows that the two-paired homeodomains do interact when bound to the DNA site through hydrophobic contacts, water mediated hydrogen bonds and electrostatic interactions. However, conformational changes in the DNA structure are essential for the homeodomains to come in position to interact. The deformation of the DNA structure that allows them to interact is mainly a bend (21°) at the centre of the palindrome site, primarily due to positive roll and tilt at the base steps from 5 to 7. Most importantly, the conformational change in the DNA that implies the co-operative interaction is a consequence of the single homeodomain binding to one half site and not a result of the dimerisation. This is strongly supported by the fact that the binding of a monomer or other homeodomain protein, brings about a similar conformational change in the DNA (9).

To summarise, there is strong evidence that co-operative binding is mediated by DNA conformational changes. However, it is not obvious how the binding of the first homeodomain promotes changes in DNA conformation, making the second binding site more favourable for the second homeodomain while maintaining the essential contacts between the DNA and the first homeodomain. In addition, how are these changes transferred from the one binding site to the other? The question that arises is what role do DNA structure and flexibility play in mediating co-operative binding of proteins at adjacent sites.

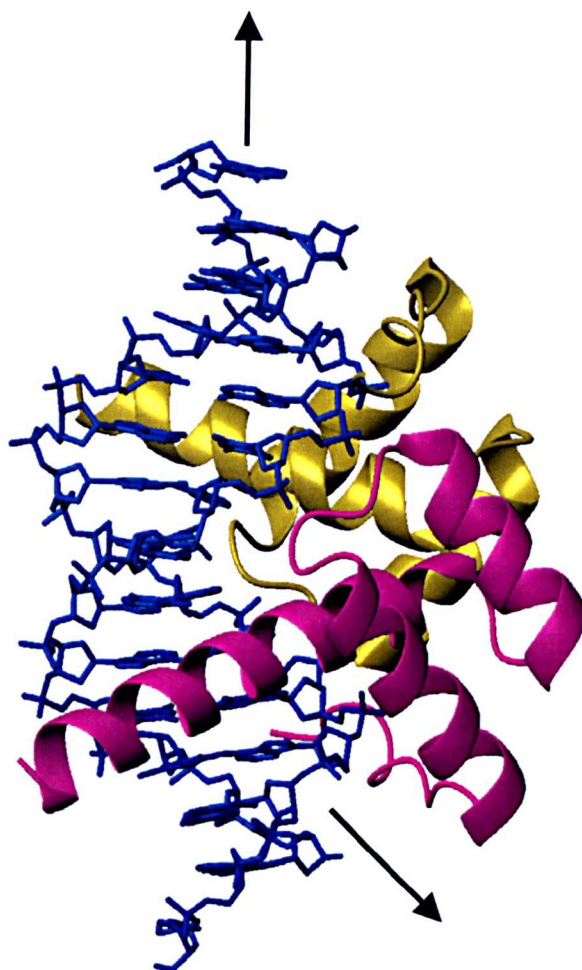


Figure 1.1 View rotated by about 90° along the DNA axis, showing the complex of a head-to-head arrangement of the paired homeodomains on the palindromic DNA site. The arrows indicate the 21° bend in the DNA (8).

It is well known that RNA–protein interactions play a central role in the regulation of gene expression. In eukaryotes there are proteins that bind to mRNA, to control their processing and export to the cytoplasm and also their translation into a protein. A well-studied example of such a protein is U1A protein, which regulates its own translation by binding to an RNA element called the polydenylation inhibition element, PIE. Two U1A molecules bind co-operatively to the PIE region. The resulting trimolecular complex interacts with the enzyme responsible for the formation of the poly (A) tail, poly(A) polymerase, and inhibits its activity (9). Co-

operative binding is necessary for the regulation of polydenylation as loss of the co-operative binding when the spacing between the two binding sites of U1A proteins is increased results in cancellation of the inhibition.

Figure 1.2 shows the structure of the trimolecular complex of two U1A proteins and the RNA element PIE (9). The dimer interface is formed by the interactions of each helix C between the two U1A proteins. The authors suggested in this case that co-operativity in protein binding is mediated by direct protein–protein interactions and not by protein-induced structural changes in the RNA element. This is in agreement with the RNA–protein interface being minimally affected by the protein–protein interactions in the trimolecular complex. The trimolecular U1A–PIE complex is the largest protein–RNA complex solved. The quality of the structure is clear with many of the intermolecular interactions in the protein–RNA interface and intramolecular interactions observed.

The structure of the bimolecular complex has been previously resolved and the differences with the trimolecular complex have been extracted. Characteristic NOE interactions between the hydrophobic residues of helix C and the remainder of the RNA-binding domain showed that the location of the helix C remained constrained by the requirement to maintain necessary interactions with the RNA, however helix C becomes longer and more rigid in the trimolecular structure. The structure explains that the observed co-operativity is a result of the optimal spacing of the two bulges. A change in the RNA structure would alter the spacing and thereby twist the two U1A monomers with respect to each other, pushing them either too close or too far apart. Formation of the trimolecular complex is favourable and offers a larger surface area of interaction with the poly(A) polymerase.

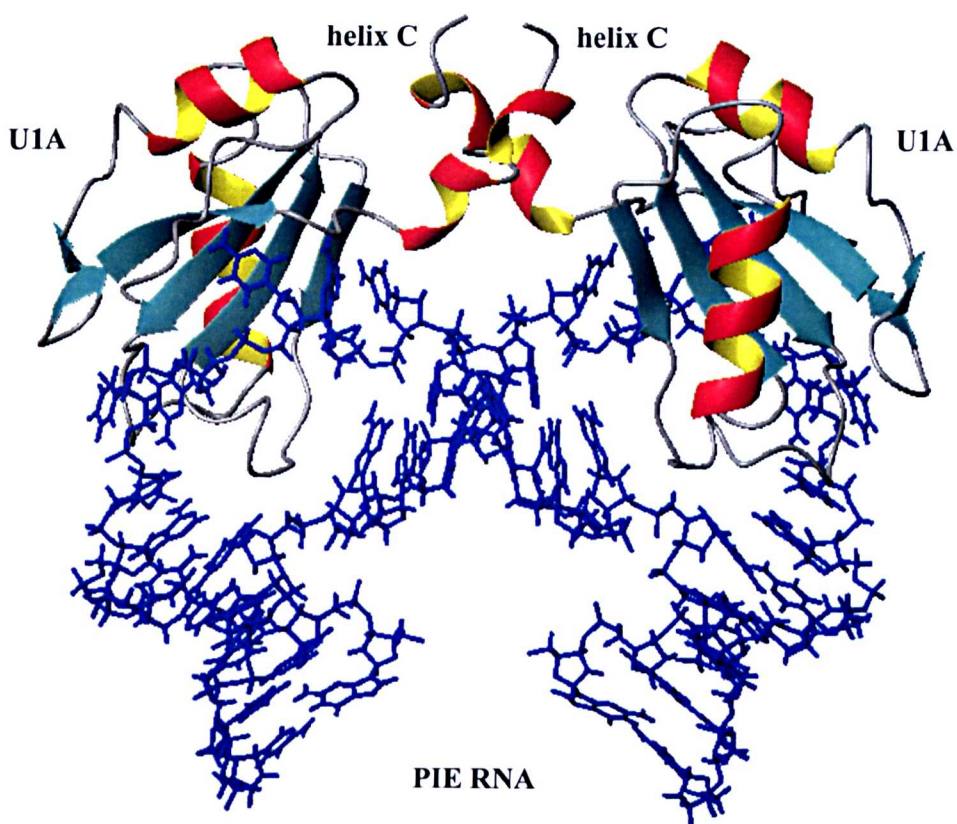


Figure 1.2 Structure of the trimolecular complex of two U1A proteins and PIE RNA. The dimer interface is formed by protein–protein interactions of the helix C from each U1A protein.

1.3 Co-operativity in drug–DNA interactions

Co-operative binding to DNA has been reported in a number of studies of natural or synthetic drug molecules that recognize specific DNA sequences. Although the factors that are responsible for co-operative binding have not been fully understood and in each case different factors may be responsible, it is likely that the phenomena associated with co-operative binding are important for the therapeutic activity of these substances. Moreover, co-operative binding appears between drug molecules that bind to DNA with different modes of interaction such as intercalation

(11), by inserting a planar aromatic chromophore between adjacent DNA base pairs, or groove binding by fitting into the DNA minor groove (12). In each case, DNA structure and flexibility should have a role in mediating co-operative binding of drug molecules, in other words DNA should allow the information from the first drug binding to be transferred through its sequence to the following site making the second binding more favourable.

1.3.1 Co-operative binding in DNA intercalation

The DNA intercalators (11) have a planar aromatic ring system that inserts between the DNA base pairs, figure 1.3. Intercalation involves separation of the DNA base pairs, distortion of the backbone conformation, and finally, unwinding of the duplex. The formation of the complexes with DNA is primarily stabilised by stacking interactions between the base pairs and the aromatic ring systems of the intercalator. Some intercalators have bulky groups at either end of the aromatic system that interact with the grooves of the DNA and offer extra binding stability. There are two types of intercalators: mono-intercalators, which have one aromatic system, and bis-intercalators, which have two aromatic systems attached with a linker.

One of the most demonstrated examples of co-operative binding in DNA intercalation has been reported for the antibiotic echinomycin, figure 1.3. Echinomycin is an antitumour antibiotic possessing a cyclic octadepsipeptide that binds tightly to DNA by the mechanism of bis-intercalation (13) and inhibits DNA-directed RNA synthesis (14). Various footprinting (15), NMR (16) and crystallographic (17) studies have demonstrated that echinomycin binds

preferentially to sites containing CpG steps in DNA. However, a weaker binding to CpA steps has also been established (18).

Figure 1.4 is a schematic representation of the echinomycin complex with the duplex d(ACGTTATACGT)₂, showing the intercalation of the quinoxaline rings at the ApC steps and the peptide binding in the minor groove (19). Its sequence selectivity is mediated by the formation of three hydrogen bonds between the alanine residues of the ligand and the N3 and 2-amino groups of the guanosine nucleotides in the recognition site. An unexpected feature of the echinomycin–oligonucleotide complexes from crystallographic studies (17) was the observation of Hoogsteen base pairing at the AT base pairs flanking the intercalation sites. Hoogsteen base pairing has been proposed to bring the sugar phosphate backbones 2 Å closer together than found for Watson–Crick pairs, allowing better Van der Waals contacts around the echinomycin molecule. However, Hoogsteen base pairing has not been observed in long DNA sequences and several NMR studies (16,19,20) confirmed that Hoogsteen base pairing depends both on the sequence and temperature.

Co-operative binding of echinomycin to specific sites of DNA has been studied by NMR spectroscopy and qualitative footprinting studies and the results from both studies agree to a high degree. Qualitative footprinting (21) showed a high degree of co-operativity in the binding of the echinomycin to certain closely spaced ACGT and TCGA sites. These results fully agree with the NMR studies of echinomycin binding to the sequences ACGTACGT (19) and ACGTATACGT (20) where echinomycin molecules bind co-operatively to each CpG step. Moreover, another NMR study showed that no co-operativity is observed between echinomycin molecules bound to TCGATCGA (19), where each molecule binds independently at each CpG step. Footprinting studies also detected co-operative interaction to the

sequence TCGAACGT (21). There is strong evidence that co-operative binding of echinomycin to DNA is sequence-dependent, and also that the distance between binding sites plays a role.

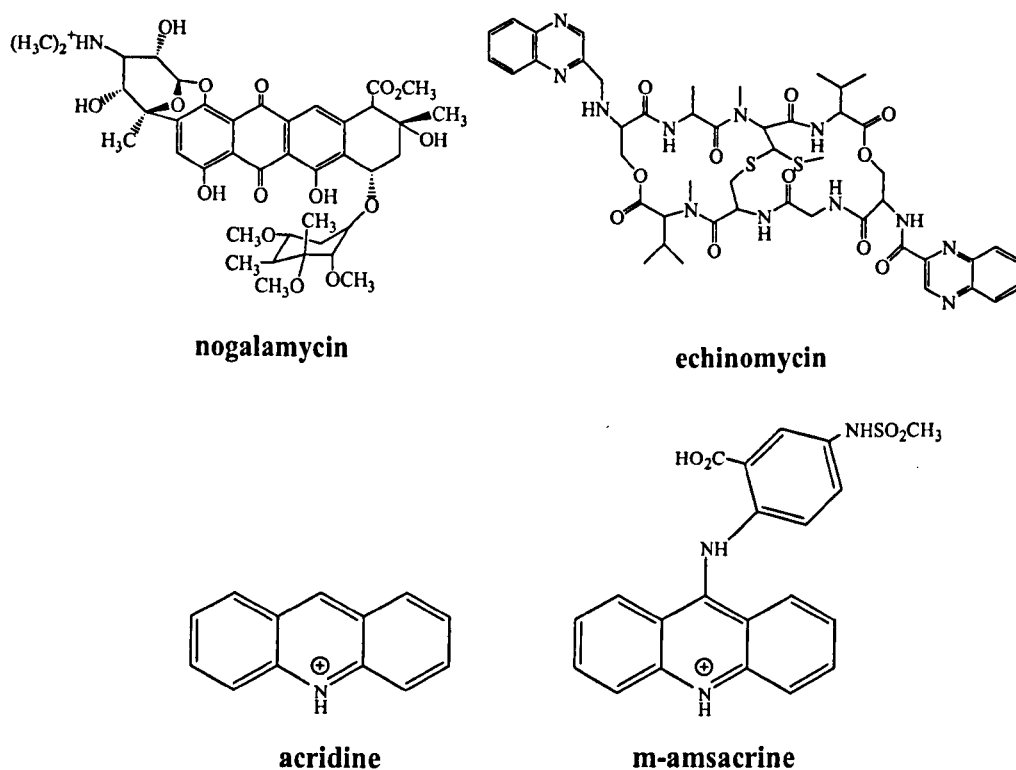


Figure 1.3 Characteristic DNA intercalators.



Figure 1.4 Schematic representation of the echinomycin complex with the duplex $d(ACGTTATACGT)_2$ (20).

Looking at both the NMR and footprinting data, it is unlikely that co-operativity originates from direct interaction between bound echinomycin molecules. The most likely origin of co-operativity, considering that it is sequence-dependent, is

drug-induced DNA conformational changes that are propagated from one binding site to the other. An interesting observation from an early NMR study of echinomycin with d(ACGTACGT)₂ (19) was that the central TA base steps are Hoogsteen paired. In contrast, the complex of echinomycin with d(ACGTATACGT)₂ (20) revealed that no Hoogsteen base pairs are formed in any central AT base pairs, showing that the extra ApT step prevents the formation of stable Hoogsteen base pairs immediately adjacent to the binding site. These NMR studies evidenced two major structural changes in the DNA duplex when two echinomycin molecules bind. The DNA duplex in the d(ACGTATACGT)₂-echinomycin complex is underwound at the central ApT step when compared with the free duplex. The presence of a cross peak between the A5H8 and T6H6 indicates an unwinding of the duplex by approximately 20 degrees. In addition, the internal AT base pairs are less stable in the bound DNA duplex than in the ligand-free duplex. The role of the destabilisation of base pairs in mediating co-operative interactions is not clear.

Footprinting results (22,23) have shown that the hyperactivity of AT-rich regions adjacent to echinomycin binding sites to DNA cleavage reagents, is propagated a significant distance from the binding sites. The unwinding of the duplex beyond the binding sites in AT-rich regions was proposed to increase the minor groove width, thus rendering the DNA more susceptible to DNA cleavage reagents.

On the basis of these observations, it is evident that bis-intercalation of echinomycin around a CpG step induces significant structural changes in the DNA helix. The unwinding of the duplex and accompanying changes in minor groove width may be slightly different between the two binding sites. Therefore these

structural changes upon the first binding of echinomycin may propagate differently to the second binding site and so may be differently perceived by the second echinomycin molecule, making the second interaction more favourable.

1.3.2 Co-operative binding in the DNA minor groove

Minor groove binders (11,12) interact with the DNA minor groove through non-covalent binding that causes small changes in DNA structure (figure 1.5). Many of them have planar and aromatic moieties that offer a complementary curvature to the bending surface of the narrow minor groove, enabling close Van der Waals contacts and hydrophobic interactions. Minor groove binders also have the potential for hydrogen bonding with the donors and acceptors on the floor of the groove. A characteristic of minor groove binders is a positively charged group that enables them to form electrostatic interactions with the negative phosphates and the negative potential of the AT tracts.

Naturally occurring oligopeptide distamycin A (figure 1.5) has been shown from numerous studies to bind to the minor groove of DNA in a co-operative mode. Distamycin A is potent antibacterial, antiviral, and antineoplastic agent whose pharmacological activity has been correlated to its ability to block DNA template function (25). Distamycin A forms non-covalent complexes with the DNA minor groove and exhibits considerable preference for AT-rich sequences. Crystallographic and NMR studies of 1:1 distamycin A–DNA complexes have shown that a minimal binding site consists of four AT base pairs, and that binding at asymmetrical sites is directional with the N-terminal formyl group pointing towards the 5' end of the AT-rich strand, figure 1.6 (11). The orientation of the first two

rings of the drug is parallel, while the third ring is inclined to conform to the rotation of the helix.

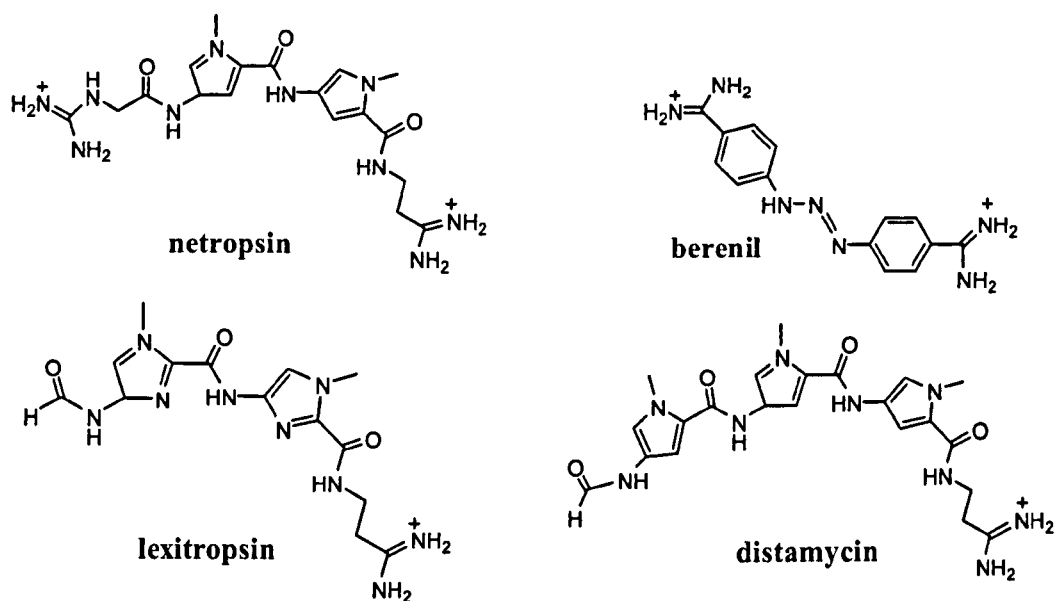


Figure 1.5 Characteristic minor groove binders.

The earlier studies (25–27) of complexes of distamycin A with binding sites consisting of four AT base pairs proposed that close Van der Waals contacts between adenine and drug H3 protons and potential three center hydrogen bonding between drug amide and adenine N3 and thymine O2 atoms on the floor of the minor groove, are common stabilizing factors. While hydrophobic interactions, hydrogen bonding and electrostatic interactions are proposed as general stabilizing factors, stacking interactions between DNA sugar O1'atoms and the three N-methylpyrrole rings are also thought to contribute significantly to complex stability (27).

NMR studies (12) have also established an interesting 2:1 binding motif of two distamycin molecules that bind simultaneously, side by side in an anti-parallel orientation, at binding sites of at least five base pairs, figure 1.6 (28). In these 2:1

complexes, distamycin molecules preserve their recognition elements of minor groove binders and are stacked side by side in the minor groove with the positively charged end groups pointing in opposite directions. The formyl group of each ligand lies at the 5' end of the adjacent strand. The binding of the second ligand in the 2:1 motif must widen the minor groove by approximately 3.5 Å relative to the 1:1 complex. The 2:1 binding motif can be highly co-operative and this depends strongly on the base sequence of the binding site, figure 1.7.

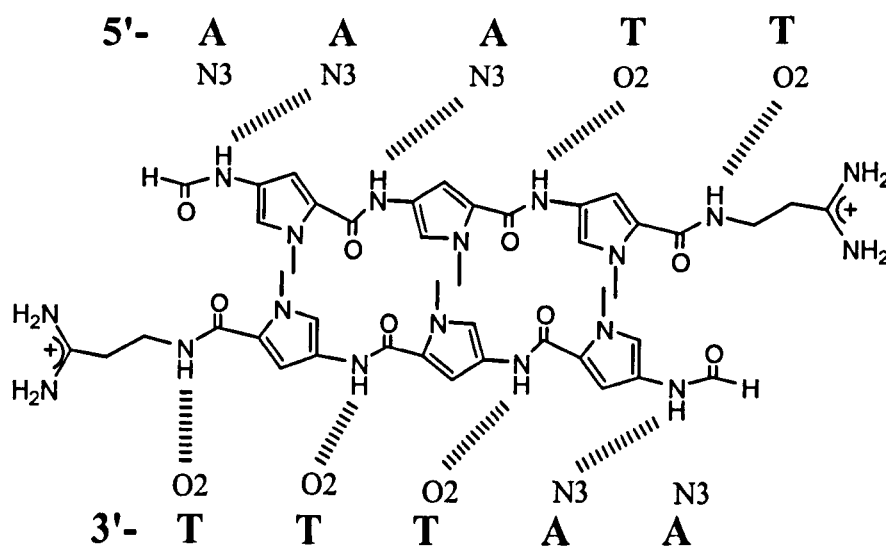


Figure 1.6 The antiparallel side-by-side motif composed of two distamycin molecules.

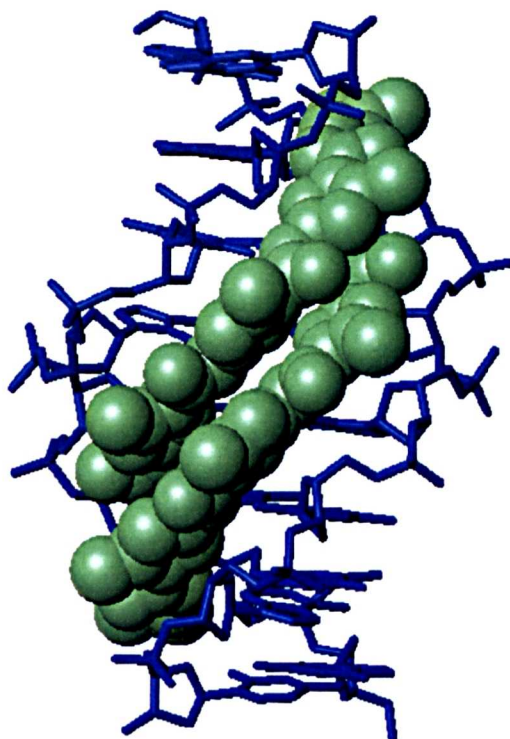


Figure 1.7 Structure of the 2:1 distamycin–d(CATATATG)₂ complex (29).

Molecular modelling studies based on NMR data of the 2:1 complex with the non-complementary duplex d(CGCAAATTGGC).d(GCCAATTTGCG) suggested that distamycin A molecules are staggered with respect to one another so that drug pyrrole rings of one antibiotic stack with the amide linkages of the another (30). Hydrogen bonds from the distamycin amide NHs to adenine N3 and thymine O2 are present again as stabilizing factors, however, three-centre hydrogen bonds with the DNA acceptors on different strands are not included in the 2:1 motif. The drug molecules are pushed to the sides of the groove rather than lying in the central region as observed in the 1:1 complex. In the 2:1 complex, the minor groove widens in order to accommodate two drug molecules side by side and this relies on the adequate flexibility of the phosphate backbone of the duplex, figure 1.7.

NMR titration of distamycin A within the complementary duplex d(CGCAAATTTGCG)₂ revealed additional kinetic features of the 2:1 interaction (31). At low drug:DNA ratios (0.5 equiv), both the one-drug and the symmetric two-drug binding motif were observed, while at higher ratios (2 equiv) the two drug complex was the primary species in solution. The off rate for the drug from the 2:1 motif was found to be slow on the NMR time scale. NOE data from the drug pyrrole H3 to DNA deoxyribose H1' and adenine H2 protons together with observed line width changes of DNA protons as a function of temperature were consistent with a model, in which two drugs bind highly co-operatively in the minor groove, with each drug sliding between 5'-AATT-3' and 5'-ATTT-3' binding sites at a rate that is fast on the NMR time scale.

This dynamic equilibrium of the two drug molecules with the two possible binding sites on each strand is a consequence of one more A-T base pair added to the sequence, as the interaction of distamycin with the previously mentioned binding site 5'-AAATT-3' (30) resulted in one 'static' 2:1 complex. The proposed dynamic model, which is fully consistent with all the intermolecular NOE data, is illustrated schematically in figure 1.8.

NMR titration studies of distamycin A with the wider minor groove of an alternating site 5'-ATATAT-3' within the d(CGCATATATGCG)₂ duplex (32) gave also a symmetrical 2:1 complex, however the two-drug binding is more co-operative than the binding to the 5'-AAATTT-3' (31) site described above. This observation has been interpreted as weaker binding in the 1:1 mode as a consequence of poorer complementarity of the drug with the wider ATATAT minor groove.

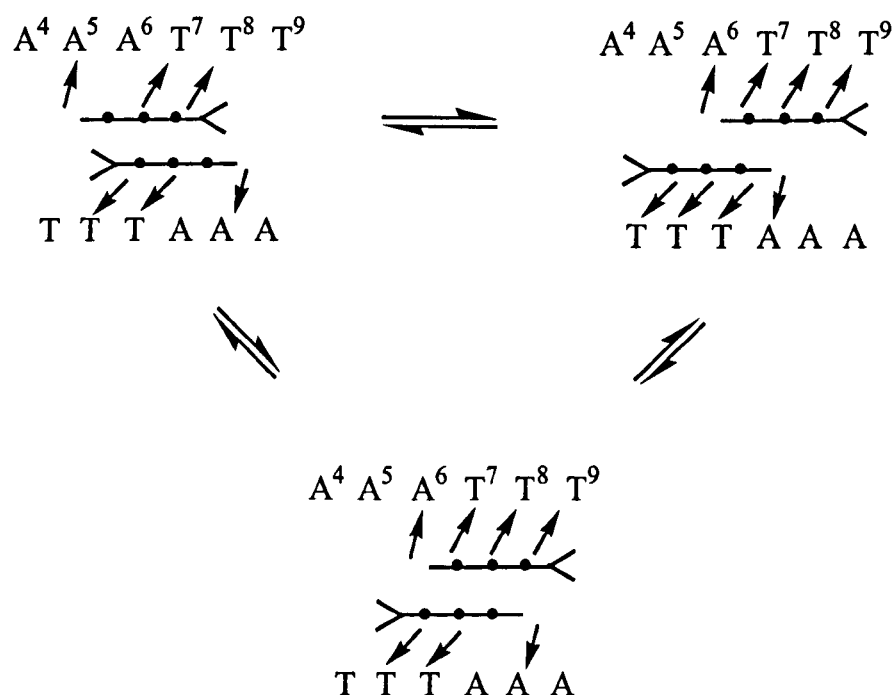


Figure 1.8 Schematic representation of the dynamic model for the 2:1 distamycin-d(CGAAATTTGCG)₂ complex (31).

Distamycin A binding was also examined with d.(CGCIIICCGGC).d(GCCIIICCGCG) (32) where I represents inosine(2-desamino-guanosine). An I-C base pair has minor groove functional groups equivalent to an A-T pair however, the groove width is similar to typical G-C rich regions (6–7 Å) rather than the narrower groove observed in AT-rich regions (3–4 Å). The electrostatic potential of I-C pairs is probably closer to that of G-C than A-T. Titration of distamycin to a 5'-IIICC site showed no evidence of an intermediate 1:1 complex, but only for the 2:1 even at very low drug concentrations, indicating a very high degree of co-operativity for the drug binding.

Comparing these titration results with the AAATTT minor groove results, these data strongly suggest that the IIICC minor groove is wider and can readily be filled with two ligands, each one making close contacts with each side of the groove.

However, one drug would make contact with only one side, lowering the binding affinity. Later NMR studies of distamycin analogues indicated only the presence of highly co-operative binding with sequences containing 5'-AAGTT-3' (33), 5'-TACGTA-3' (34) or 5'-TAACAA-3' (35) binding sites. These distamycin analogues contain imidazole nitrogens that can recognize a guanine amino group by hydrogen bonding.

The introduction of a GC base pair into an AT sequence of DNA widens the minor groove locally and results in distamycin binding preferentially in the 2:1 mode. These results emphasize that the binding behaviour of distamycin reflects the local variation in DNA groove, specifically the minor groove width geometry and flexibility. Distamycin forms tight complexes with DNA only if it can provide close Van der Waals contacts and hydrogen bonding, and also if the DNA minor groove can expand to accommodate two distamycin molecules. In addition, it seems energetically unfavourable to narrow a wide minor groove to increase the binding affinity of the 1:1 complex. Therefore, it was suggested that the loss of shape complementarity and contact surface between ligand and groove is more important than the depth of the groove (12). Clearly the co-operative 2:1 binding of distamycin depends on the sequence-dependent variations in minor groove structure and flexibility.

1.4 Conclusions

This chapter defines the term co-operativity and discusses mainly structural studies of ligands that show co-operative interactions when bound to DNA. We saw that co-operativity can operate when the ligand intercalates between base pairs or binds within the minor or major groove of the DNA molecule. Enhancement of the

binding interaction in co-operative systems is usually a consequence of favourable ligand–ligand interactions. We are interested to investigate co-operativity in a system where there is a communication between binding sites without the sites being in close proximity, preventing direct ligand–ligand interactions. Echinomycin is a very good example of co-operativity where binding sites are well separated. The intercalation reaction involves distortion and unwinding of the DNA helix, which can allow for preorganisation of the second binding site when the first echinomycin molecule binds. However, when the ligand interacts with the minor groove it is unlikely that this causes similar distortion in the DNA structure as that induced by intercalators, so the origins of the co-operativity are not obvious. In the next two chapters the efforts of understanding co-operativity in the DNA minor groove, when binding sites are well separated, are described. First, our efforts are concentrated on establishing co-operativity and understanding its origins in our system. Then sequence dependent effects of the DNA as well as increase of the distance between sites are considered in order to obtain insights into the factors important in information transfer between binding sites that leads to co-operativity.

1.5 References

1. Mitchell, P. J. & Tjian, R. 1989 Transcriptional Regulation in Mammalian-Cells by Sequence-Specific DNA-Binding Proteins. *Science* **245**, 371-378.
2. Sorger, P. K. & Pelham, H. R. B. 1988 Yeast Heat-Shock Factor is an Essential DNA-Binding Protein That Exhibits Temperature-Dependent Phosphorylation. *Cell* **54**, 855-864.
3. Gehring, W. J., Affolter, M. & Burglin, T. 1994 Homeodomain Proteins. *Annu. Rev. Biochem.* **63**, 487-526.
4. Laughon, A. 1991 DNA-Binding Specificity of Homeodomains. *Biochemistry* **30**, 11357-11367.
5. Duboule, D. 1994 Guidebook to the Homeodomain Genes. Oxford University Press, Oxford.
6. Wilson, D. S. & Desplan, C. 1995 Homeodomain Proteins – Cooperating to Be Different. *Curr. Biol.* **5**, 32-34.
7. Wilson, D., Sheng, G. J., Lecuit, T., Dostatni, N. & Desplan, C. 1993 Cooperative Dimerization of Paired Class Homeo Domains on DNA. *Genes Dev.* **7**, 2120-2134.
8. Wilson, D. S., Guenther, B., Desplan, C. & Kuriyan, J. 1995 High-Resolution Crystal-Structure of a Paired (Pax) Class Cooperative Homeodomain Dimer on DNA. *Cell* **82**, 709-719.
9. Varani, L., Gunderson, S. I., Mattaj, I. W., Kay, L. E., Neuhaus, D. & Varani, G. 2000 The NMR Structure of the 38 kDa UIA Protein-PIE RNA Complex Reveals the Basis of Co-operativity in Regulation of Polyadenylation by Human UIA protein. *Nat. Struct. Biol.* **7**, 329-335.
10. Kissinger, C. R., Liu, B. S., Martinblanco, E., Kornberg, T. B. & Pabo, C. O. 1990 Crystal-Structure of an Engrailed Homeodomain-DNA Complex at 2.8 Å Resolution – A Framework for Understanding Homeodomain-DNA Interactions. *Cell* **63**, 579-590.
11. Searle, M. S. 1993 NMR-Studies of Drug-DNA Interactions. *Prog. Nucl. Magn. Reson. Spectrosc.* **25**, 403-480.
12. Geierstanger, B. H. & Wemmer, D. E. 1995 Complexes of the Minor-Groove of DNA. *Ann. Rev. Biophys. Biomolec. Struct.* **24**, 463-493.
13. Waring, M. J. & Wakelin, L. P. G. 1974 Echinomycin: A Bifunctional Intercalating Antibiotic. *Nature* **252**, 653-657.
14. Neidle, S. & Waring, M. 1994 Molecular Aspects of Anti-cancer Drug-DNA interactions. Vol 1, The Macmillan Press LTD, UK.

15. Vandyke, M. M. & Dervan, P. B. 1984 Echinomycin Binding-Sites on DNA. *Science* **225**, 1122-1127.
16. Gao, X. L. & Patel, D. J. 1988 NMR-Studies of Echinomycin Bisintercalation Complexes with d(A1-C2-G3-T4) and d(T1-C2-G3-A4) Duplexes in Aqueous-Solution – Sequence-Dependent Formation of Hoogsteen A1.T4 and Watson-Crick T1.A4 Base-Pairs Flanking the Bisintercalation Site. *Biochemistry* **27**, 1744-1751.
17. Wang, A. H. J., Ughetto, G., Quigley, G. J., Hakoshima, T., Vandermarel, G. A., Vanboom, J. H. & Rich, A. 1984 The Molecular-Structure of a DNA Triostin-A Complex. *Science* **225**, 1115-1121.
18. Phillips, D. R., White, R. J., Dean, D. & Crothers, D. M. 1990 Monte-Carlo Simulation of Multisite Echinomycin DNA Interactions Detected by In Vitro Transcription Analysis. *Biochemistry* **29**, 4812-4819.
19. Gilbert, D. E. & Feigon, J. 1991 The DNA-Sequence at Echinomycin Binding Sites Determines the Structural-Changes Induced by Drug-Binding – NMR Studies of Echinomycin Binding to [d(ACGTACGT)]₂ and [d(TCGATCGA)]₂. *Biochemistry* **30**, 2483-2494.
20. Gilbert, D. E. & Feigon, J. 1992 Proton NMR Study of the [d(ACGTATACGT)]₂-2 Echinomycin Complex – Conformational Changes Between Echinomycin Binding Sites. *Nucleic Acids Res.* **20**, 2411-2420.
21. Bailly, C., Hamy, F. & Waring, M. J. 1996 Cooperativity in the Binding of Echinomycin to DNA Fragments Containing Closely Spaced CpG Sites. *Biochemistry* **35**, 1150-1161.
22. Mendel, D. & Dervan, P. B. 1987 Hoogsteen Base-Pairs Proximal and Distal to Echinomycin Binding-Sites on DNA. *Proc. Nat. Acad. of Sci. U.S.A.* **84**, 910-914.
23. Fox, K. R. & Kentebe, E. 1990 Echinomycin Binding to the Sequence CG(AT)_n CG Alters the Structure of the Central at Region. *Nucleic Acids Res.* **18**, 1957-1963.
24. Krey, A. K., Olenick, J. G. & Hahn, F. E. 1976 Interactions of the Antibiotic Distamycin A With Homopolymeric Single-Stranded Polydeoxyribonucleotides and With Phix174 Deoxyribonucleic Acid. *Mol. Pharmacol.* **12**, 185-190.
25. Kopka, M. L., Yoon, C., Goodsell, D., Pjura, P. & Dickerson, R. E. 1985 The Molecular-Origin of DNA Drug Specificity in Netropsin and Distamycin. *Proc. Nat. Acad. of Sci. U.S.A.* **82**, 1376-1380.
26. Klevit, R. E., Wemmer, D. E. & Reid, B. R. 1986 H¹-NMR Studies on the Interaction Between Distamycin-A and a Symmetrical DNA Dodecamer. *Biochemistry* **25**, 3296-3303.

27. Pelton, J. G. & Wemmer, D. E. 1988 Structural Modeling of the Distamycin-A-d(CGCGAATTCGC)₂ Complex Using 2D NMR and Molecular Mechanics. *Biochemistry* **27**, 8088-8096.
28. Chen, Y. H., Yang, Y. W. & Lown, J. W. 1996 Design of Distamycin Analogues to Probe the Physical Origin of the Antiparallel Side-By-Side Oligopeptide Binding Motif in DNA Minor Groove Recognition. *Biochem. Biophys. Res. Commun.* **220**, 213-218.
29. Mitra, S. N., Wahl, M. C. & Sundaralingam, M. 1999 Structure of the Side-By-Side Binding of Distamycin to d(GTATATAC)₂. *Acta Crystallogr. Sect. D-Biol. Crystallogr.* **55**, 602-609.
30. Pelton, J. G. & Wemmer, D. E. 1989 Structural Characterization of a 2-1 Distamycin A. d(CGCAAATTGGC) Complex by Two-Dimensional NMR. *Proc. Nat. Acad. of Sci. U.S.A* **86**, 5723-5727.
31. Pelton, J. G. & Wemmer, D. E. 1990 Binding Modes of Distamycin-A with d(CGCAAATTTGCG)₂ Determined by 2-Dimensional NMR. *J. Am. Chem. Soc.* **112**, 1393-1399.
32. Fagan, P. & Wemmer, D. E. 1992 Co-operative Binding of Distamycin-A to DNA in the 2-1 Mode. *J. Am. Chem. Soc.* **114**, 1080-1081.
33. Geierstanger, B. H., Dwyer, T. J., Bathini, Y., Lown, J. W. & Wemmer, D. E. 1993 NMR Characterization of a Heterocomplex Formed by Distamycin and Its Analog 2-Imd with d(CGCAAGTTGGC)-d(GCCAAGTTGCG) – Preference for the 1:1:1 2-Imd-Dst-DNA Complex over the 2:1 2-Imd-DNA and the 2:1 Dst-DNA Complexes. *J. Am. Chem. Soc.* **115**, 4474-4482.
34. Capobianco, M. L., Colonna, F. P., Forni, A., Garbesi, A., Iotti, S., Moretti, I., Samori, B. & Tondelli, L. 1991 Interactions of Nucleic Acids with Distamycins – Binding of Dst-3 to d(CGTTTAAACG)₂ and d(CGTACGTACG)₂. *Nucleic Acids Res.* **19**, 1695-1698.
35. Geierstanger, B. H., Jacobsen, J. P., Mrksich, M., Dervan, P. B. & Wemmer, D. E. 1994 Structural and Dynamic Characterization of the Heterodimeric and Homodimeric Complexes of Distamycin and 1-Methylimidazole- 2-Carboxamide-Netropsin Bound to the Minor-Groove of DNA. *Biochemistry* **33**, 3055-3062.

2. Structural investigations on DNA recognition by Hoechst 33258

2.1 Introduction

Studies of the interactions of small drug molecules with DNA have been carried out for several decades in the hope of learning the recognition code that governs the targeting of specific DNA sequences in order to control gene expression (1–3). There are a number of natural antibiotics and synthetic molecules that bind to DNA in different modes and act as inhibitors of DNA transcription and replication processes (4). Their mechanism of action usually involves the inhibition of topoisomerases I and II. This prevents DNA religation and stabilisation of the protein–DNA complex during the processes of replication, resulting in protein-induced DNA strand breaks. A number of these drug molecules have been proven clinically useful as anti-tumour, anti-viral or anti-microbial agents (5–9) although their exact mode of action has not been defined completely. Understanding the molecular basis of action of such molecules will permit the rational design of new improved therapeutic agents with enhanced or more selective activity.

Ligands with low molecular weight, e.g. drug molecules that interact non-covalently with the DNA have been used as model compounds for DNA binding proteins, in order to make the analysis of these systems relatively simple. Drug molecules represent good model systems both in their way of interacting with the DNA, as the same forces are responsible for protein–DNA interaction, and by exhibiting similar binding site size. Thus drug–DNA complexes may serve as good

models to investigate the molecular basis of DNA binding and increase our knowledge about the mechanism of recognition involved in gene expression.

Hoechst 33258, figure 2.1, abbreviated to H33258, is a synthetic compound containing two linked benzimidazole rings with phenol and N-methylpiperazine rings attached at either end of the structure, giving the molecule a crescent shape. H33258 is the best known and studied of the bis-benzimidazole family of minor groove binders and it has been used widely as a fluorescent cytological DNA stain (10,11). It possesses anthelmintic properties (12) and shows some activity against L210 and P388 leukemias (13). Recently, phase II clinical trials were carried out for the treatment of pancreatic cancer, but efforts were abandoned due to high toxicity (14). A number of H33258 analogues have been designed and synthesised usually by substituting the phenol and piperazine rings with other moieties. Some analogues have shown increased DNA specificity and significant cytotoxicity against several human cancer cell lines (15–16).

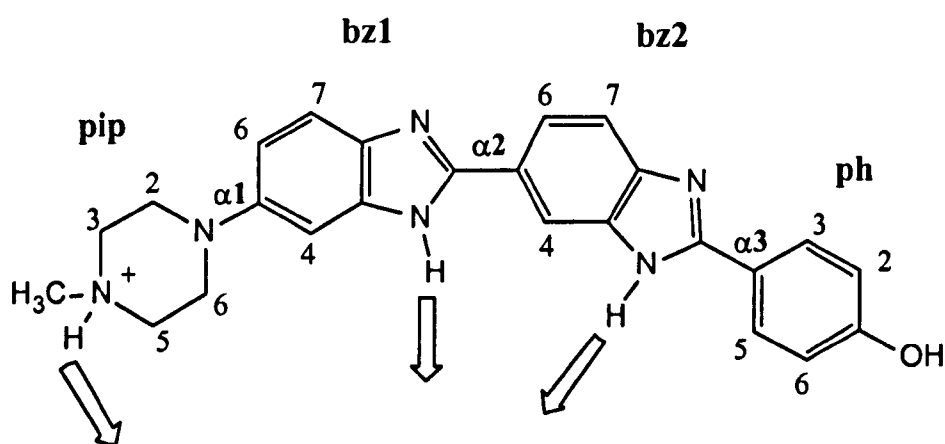


Figure 2.1 Hoechst 33258 scheme with atoms, torsion angles and residues labelled. Arrows indicate hydrogen-bonding potential.

DNA footprinting and spectroscopic studies have shown that H33258 exhibits a distinct selectivity for AT-rich DNA sequences with a binding site of 4–5 bases (17,18). Although H33258 requires at least 4 consecutive AT bases for strong binding, additional studies with repeating synthetic DNA polymers and natural DNA showed that H33258 has a weaker binding mode which is reduced in the presence of high salt and has no apparent base specificity (19). Linear dichroism studies show minor groove binding with poly(dA–dT)·poly(dA–dT), poly(dA)·poly(dT) and DNA from calf thymous and partially intercalating binding with poly(dG–dC)·poly(dG–dC), poly(dA–dG)·poly(dT–dC) and poly(dG)(dC) (20). Association constants K_a for binding of H33258 with AT-rich self-complementary duplexes determined using fluorescence titrations lie in the range of $2\text{--}5 \times 10^8$ (duplex M)⁻¹ (21). A detailed calorimetric and fluorescence study with the oligonucleotide d(CGCAAATTTGCG)₂ has concluded that H33258 DNA recognition is an entropy-driven process and that Van der Waals contacts along with hydrogen-bonding interactions play only a fine tuning role in the binding site (22).

A number of NMR (23–27) and X-ray crystallographic (28–33) studies of H33258 complexes with AT-containing oligonucleotides have been reported and, in part, shed light on the molecular basis for DNA minor groove recognition. These studies position the drug in the minor groove with the benzimidazole groups following the groove curvature. Each nitrogen atom forms hydrogen bonds in a bifurcated manner to a pair of hydrogen-bond donors of adjacent AT base pairs found on the floor of the groove. While both NMR and X-ray studies agree that the two almost planar aromatic benzimidazoles and phenol rings are accommodated firmly within the narrow AT-tract, there is some controversy about the degree of GC tolerance of the bulky positively charged piperazine ring. Specifically, there are

some crystal structures (29,30) where the piperazine ring is bound at the wider GC minor groove close to the end of the AT-tract, while most of the structures suggest that binding preferentially occurs within the central region of the AT groove. This structural heterogeneity may arise from the way crystals are grown as well as the temperature at which data were collected, or may be due to X-ray crystal structures being a series of snap shots of a dynamic complex.

Hoechst 33258 possesses the common recognition properties found for minor groove binders such as complementary shape with the floor of the minor groove, hydrogen bonding with the acceptors on the floor of the groove and cationic charge for electrostatic interactions with negative phosphates and negative potential of AT-tracts. The NMR study (27) of H33258 with an extended AT-tract in the sequence d(GCAAAATTTTCG)₂ identified one binding site and no evidence of movement (sliding) along the length of the AT-tract. H33258 was bound at the central AATT site, where the groove is narrower and restricts the mobility of the drug. This study suggests that sequence specificity may be dominated by the ability to select binding sites on the basis of groove width, allowing optimum Van der Waals complementarity and hydrophobic surface burial, rather than specific hydrogen-bonding interactions with the groove floor. Similar conclusions have been drawn from the crystallographic study of a meta-hydroxy Hoechst analogue (34).

The vast majority of the structural data on Hoechst 33258 and its analogues bound to DNA have reported complexes containing a single binding site of 4–6 A–T base pairs within self-complementary AATT, ATAT and AAATTT binding sites (23–33). The limitations imposed by these sequences are two fold: (i) the limited size of these binding sites has complicated the analysis of sequence-dependent A-tract recognition by restricting the number of high affinity binding sites available in

one sequence; (ii) the orientational preference of the bound ligand has been obscured by choosing A-tracts containing self-complementary (A_nT_n) binding sites where the drug can bind in one of two symmetry related orientations. It is known that T_n or A_n tracts ($n \geq 4$) are not uniform; recent studies of the localisation of ammonium ions in the minor groove of various A-tracts show sequence-dependent effects which correlate with groove narrowing towards the 3'-end of A_n tracts ($n > 4$) (35), making symmetrical A_nT_n tracts narrowest at the centre of the sequence. The sensitivity of ligand binding to the direction of groove narrowing in isolated A_n and T_n tracts, together with the effects of A-tract junctions on binding orientation, has not been addressed.

To examine the effects of sequence-dependent structural features on groove recognition and ligand orientation we describe a high resolution NMR study of H33258 binding to the dodecamer duplex $d(CTTTTGCAAAG)_2$ containing two TTTT tracts. Previously, titration studies and subsequent low-resolution structural analysis revealed that two bound drug molecules are accommodated in symmetry-related orientations such that the overall dyad symmetry of the duplex is retained in solution (23). Binding at the two sites was found to be highly co-operative with only the 2:1 complex detected in equilibrium with the free duplex. Capillary electrophoresis studies have demonstrated a similar co-operative behaviour with other AT-rich sequences (36). Furthermore, the two TTTT sites are separated by an intervening GC pair, which isolates the two ligands, figure 2.2. Thus, a direct-contact mechanism cannot explain the observed co-operativity. This contrasts with the anti-parallel side-by-side dimer motif of distamycin (figure 1.7) where there is a highly complementary fit within a widened minor groove with a favourable π -stacking interaction between the two drug molecules (37, 38).

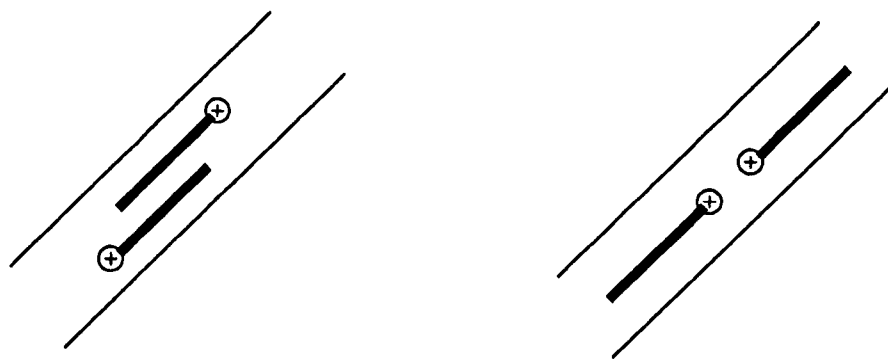


Figure 2.2 Left, side by side, head to tail co-operative binding of distamycin molecules in the 2:1 complex, right, co-operative head to head binding of H33258 molecules in the minor groove of the 2:1 complex with d(CTTTTGCAAAAG)₂.

We investigate the molecular basis of co-operative processes in ligand–DNA interactions. In this work, we have attempted to determine whether the drug binds co-operatively to the two sites by further analysis of the H33258 complex. In other words, is there an enhancement of the binding affinity to the second site as soon as the first is occupied? Following this, we are interested to investigate the origins of the observed binding co-operativity and also examine the role of DNA structure and flexibility in such a phenomenon, since the DNA is a highly flexible molecule and permits structural changes on ligand binding, as demonstrated by the work with distamycin.

In this study, we examined extensively the complex of H33258 with d(CTTTTGCAAAAG)₂ by NMR spectroscopy and restrained molecular dynamics using the AMBER 95 force field with an explicit solvent model. We will examine sequence-dependent effects on H33258 groove preference and orientation and also shed light on the origins of the co-operative process. The structure and dynamics of the high-resolution complex are analysed and compared with those of the unbound DNA to examine how the DNA structure is affected upon ligand binding. The

dynamics and the structure of the AT-tracts in the free duplex are also examined by NMR based on imino proton exchange times τ_{ex} . The complex of H33258 with a related duplex d(GAAAAGCTTTTC)₂, where the orientation of the AT-tracts has been inverted, is also studied.

2.2 Materials and methods

2.2.1 DNA synthesis

The oligonucleotides d(CTTTTGCAAAAG)₂ and d(GAAAAGCTTTTC)₂ were synthesized on an Applied Biosystems DNA synthesiser using standard solid-phase phosphoramidite chemistry, and purified trityl-on by reverse phase HPLC using TEEA buffer, pH=7.0 and acetonitrile gradient. The trityl group was cleaved by treatment with 50% aqueous acid for 1 h at 35° C. Acetic acid was extracted into ether (3 x 200 ml) and the oligonucleotides finally dialysed to remove residual TEEA and acetic acid then dialysed against 1 M NaCl and 100 mM Na₂PO₄ (2000 ml) to introduce Na⁺ as the counter ion. The Na⁺ ions bind the negative phosphate backbone of the DNA and stabilise the duplex. Finally, the sample was dialysed twice with water (2000 ml) to reduce the salt concentration. The oligonucleotides were shown to be > 95% pure in duplex form by ¹H NMR spectroscopy.

2.2.2 Sample preparation

The final salt concentration of the purified oligonucleotides was adjusted to 100 mM NaCl and 10 mM NaH₂PO₄ with the addition of stock solutions (125 μ l NaCl 0.5 M, 10 μ l NaH₂PO₄ 0.5 M). The oligonucleotide samples were quantified spectrophotometrically using extinction coefficients determined by the nearest-neighbour method (39). The oligonucleotides were lyophilised and dissolved in 0.6

ml D₂O, with the addition of 2,2,3,3-[²H₄]-trimethylsilylpropanoic acid TSP (3 µl of a 1 mM solution) as an internal NMR reference (0.0 ppm). The oligonucleotides were also prevented from decomposing by additions of sodium azide NaN₃ (10 µl of a 6 mM solution), an antibacterial agent, and the disodium salt of ethylenediaminetetraacetic acid EDTA (10 µl of a 6 mM solution), a metal chelating agent. Finally the pH of the sample was adjusted to 7.0 with appropriate additions of NaOD solution in D₂O.

The sample of Hoechst 33258 was purchased from Sigma and was used without further purification after checking its purity by ¹H NMR. Its concentration was determined spectrophotometrically at 25 °C using the following extinction coefficient $\epsilon_{338} = 42000 \text{ M}^{-1}\text{cm}^{-1}$. The 2:1 complex of DNA and H33258 was generated by titrating small aliquots of the drug solution in D₂O (~10 mM) into a solution of the duplex. The stoichiometry was monitored by ¹H NMR and stopped when the resonances of the free DNA were fully replaced with those of the 2:1 complex. The NMR sample was lyophilised again and dissolved in 90% H₂O and 10% D₂O solution (0.6 ml), for the determination of the exchangeable proton resonances.

2.2.3 NMR analysis

All ¹H NMR spectra were recorded at 500 MHz on a Bruker Avance DRX-500 spectrometer and NMR data collected and processed on R4600PC and R5000SC Silicon Graphics Indy Workstations using Bruker X-WinNMR software. Standard phase-sensitive 2D NMR pulse sequences were used throughout, including NOESY, DQF-COSY, TOCSY, HMQC and WATERGATE-NOESY for solvent suppression in 90% H₂O solutions (26). Quadrature detection for f1 was achieved using time

proportional phase incrementation (TPPI). In D₂O solutions, the carrier frequency was placed at the centre of the spectrum at the frequency of the HOD solvent resonance to allow presaturation by low power continuous wave irradiation during the relaxation delay. In H₂O solutions, the transmitter was placed at the frequencies of the H₂O resonance and solvent suppression was achieved by using the WATERGATE pulse sequence.

NOESY spectra were acquired at mixing times between 50 and 300 ms and, typically, 2048 complex data points in t_2 were collected for each of 512 t_1 increments with 64 transients for each. WATERGATE-NOESY spectra at 250 ms, 100 ms and 75 ms mixing times were collected using a sample in 90% H₂O and 10% D₂O solution over 2048 data points in t_2 and 512 points in t_1 with 64 transients for each, and employed a 10 ppm spectral width. Spectra were zero filled to 2048 x 1024 prior to Fourier transformation, apodised typically with a sine-squared window function in both dimensions and also treated with automatic baseline correction. Measurements of imino proton line widths as a function of [NH₃] used the jump-return pulse sequence for solvent suppression.

2.2.4 NOE restraints

Intensities of the assigned peaks for each mixing time were converted to distances by using the inverse sixth power relationship between NOE intensity and interproton distance:

$$D_{ij} = \sqrt[6]{\frac{D_{ref}^6}{I_{ij}/I_{ref}}}$$

Where D_{ij} is the interproton distance calculated for the cross-peak between i and j protons, D_{ref} is the interproton distance of the reference cross-peak, I_{ij} is the intensity

of the NOE cross-peak between i and j and I_{ref} is the intensity of the reference NOE cross-peak of known distance. In order to be more accurate in getting the distance restraints, different internal distances were used (D_{ref}) for the three different sets of protons: sugar protons, base protons and methyl protons. The NOEs that involve sugar protons were referenced with sugar H2'-H2'' distance (1.85 Å), the base protons were referenced with cytosine H5-H6 distance (2.45 Å) and the thymine CH₃-CH₆ reference distance (3.0 Å) was applied to the NOEs involving methyl protons.

Interproton distances in both the 2:1 complex and the free DNA were derived by integration of NOEs assigned from 75, 100 and 125 ms NOESY spectra in D₂O, and 200 and 100 ms NOESY data in H₂O solutions. Final distances were calculated using the extrapolation method as described by Baleja *et al.* (40). The extrapolation method is the most accurate method of estimating the correct distance by minimising the spin diffusion effects on NOE data. Distances from each mixing time were extrapolated back to zero mixing time. The inherent error associated with distance calculations from the NOE data led to a 20% upper bound error being applied to all distances involving the non-exchangeable protons. All distances involving the exchangeable protons were given a 40% upper bound error due to the degree of exchange with the solvent. Finally, distance restraints were compared visually with the distances of a structure derived from unrestrained molecular dynamics simulation, by using MolMol software for molecular graphics and display (41). A total of 326 and 420 distance restraints were used for structural refinements of the free duplexes d(CTTTTGCAAAAG)₂ and d(GAAAAGCTTTTC)₂ and 464 for the 2:1 complex, including 56 drug-DNA NOEs. Tables with the list of the NOE restraints are found in appendix 2.

2.2.5 Structure calculations of d(CTTTTGCAAAAG)₂ and d(GAAAAGCTTTTC)₂

Molecular modelling was performed on an Origin 200 Silicon Graphics server using the AMBER 94 force field (43) with explicit solvent models and Particle Mesh Ewald method (PME) (44) for the treatment of electrostatics. Energy minimisations, restrained and unrestrained molecular dynamics were carried out using the SANDER module of AMBER 4.1. Calculations with SANDER were performed with a 2 fs time step; the SHAKE algorithm (tolerance 0.00005 Å) was applied to all bonds to remove bond stretching and a 9 Å cut off was applied to the Lennard Jones interactions. The restrained molecular dynamics were performed at constant pressure 1.0 atm with isotropic position scaling and at 300 K, utilising the Berendsen algorithm for temperature coupling. Translational and rotational motions were removed every 100 fs. All calculations were carried out with the PME method using a 9 Å cut-off for direct space non-bonded calculations and a 0.00001 Ewald convergence tolerance for the inclusion of long-range electrostatics in our calculations.

The initial model of the dodecamer d(CTTTTGCAAAAG)₂ as a canonical B-DNA was created using the AMBER 4.1 modules *NUKIT*, *NUCGEN* and *LEAP*. Previous work by Bostock-Smith *et al.* (42) showed that when starting from A- or B-form DNA, the final conformation is independent of starting structure, with equilibration faster for B-form DNA. The duplex system was neutralized, requiring 22 sodium ions placed at the most negative locations but no closer than 5 Å to any other atom in the system using Coloumbic potential terms with the LEAP module. The DNA molecule was solvated in a periodic TIP3 water box of approximate dimensions 40 Å x 40 Å x 60 Å that extended to a distance of 5 Å from any solute

atom and contained 1748 water molecules. The final information about the structure is included in the topology and coordinate files, and these were used for molecular dynamics simulations.

The starting structure was allowed to equilibrate fully before the distance restrained MD calculations were applied using the following specific protocol. Minimisation with 5000 step conjugate gradient was applied first to water only, then to water and sodium ions and finally to all components of the system. Next, 10ps of dynamics were run at 100K on the water only with the DNA and sodium ions constrained. A further 10 ps of dynamics with PME allowed the water and sodium ions to move, but not the DNA. PME was applied to all subsequent simulations. In the following 5 ps of dynamics the temperature of the system was increased from 100 K to 300 K. In the next runs, each of them of 10 ps dynamics, the DNA force constant is gradually reduced from 100 to 50, 25, 10, 5 and 2.5 kcal mol⁻¹Å⁻². The equilibration step ends with 100 ps of dynamics on the whole, fully unrestrained system. This equilibration protocol of free DNA is described step by step in table 2.1.

Distance restraints with a force constant of 30 kcal mol⁻¹ Å⁻² were introduced gradually to the system over a period of 20 ps, with the temperature stable at 300 K and PME on. The NOE-restrained B-DNA dynamics simulation was extended to 1 ns and snapshots of each ps were extracted from the whole simulation. The energy penalty from 36 distance restraint violations was 32.15 kcal mol⁻¹, most of them found at the terminal residues. The maximum distance violation was 0.46 Å, 21 restraints were violated by < 0.1 Å, a further 9 restraints by < 0.2 Å, and 5 restraints by < 0.4 Å. A 1 ns NOE restrained simulation of d(CAAAAGCTTTTG)₂ was performed and analysed following the same methods as for the

d(CTTTTGCAAAAG)₂. Distance restraint violations from 62 restraints contributed to an energy penalty of 62.15 kcal mol⁻¹. The maximum distance violation was 0.29 Å, 15 restraints were violated by < 0.29 Å, 27 restraints by < 0.2 Å, and 20 restraints by < 0.1 Å. The average structure and the RMSDs over the course of the trajectory were calculated using the CARNAL module of AMBER 4.1 and helicoidal structural properties were analysed using CURVES 5.1 (45). Average structures were calculated over the final 500 ps of the simulations and energy-minimised.

Table 2.1 Molecular dynamics equilibration protocol of the free d(CTTTTGCAAAAG)₂ and d(GAAAAGCTTTC)₂.

Steps	Time	Notes
1	5000 steps	Minimisation of water
2	5000 steps	Minimisation of water and Na ⁺
3	5000 steps	Minimisation of whole system
4	0-10 ps	MD at 100K on water, DNA and Na ⁺ constrained
5	10-20 ps	MD at 100K on water and Na ⁺ , DNA constrained
6	20-25 ps	Temperature rise to 300K
7	35-45 ps	MD on system, DNA is held with 100 kcal mol ⁻¹ Å ⁻²
8	45-55 ps	MD on system, DNA is held with 50 kcal mol ⁻¹ Å ⁻²
9	55-65 ps	MD on system, DNA is held with 25 kcal mol ⁻¹ Å ⁻²
10	65-75 ps	MD on system, DNA is held with 10 kcal mol ⁻¹ Å ⁻²
11	75-85 ps	MD on system, DNA is held with 5 kcal mol ⁻¹ Å ⁻²
12	85-95 ps	MD on system, DNA is held with 2.5 kcal mol ⁻¹ Å ⁻²
13	95-195 ps	MD on system, DNA fully free, no restraints

2.2.6 Structure calculations of 2:1 H33258–d(CTTTTGCAAAG)₂ complex

Force field parameters of Hoechst 33258 atoms were determined by using AMBER 94 force field parameters and missing parameters were estimated from similar atom types within the AMBER parameter set. The electrostatic potential around the Hoechst 33258 was first calculated as a grid of points by an *ab initio* method at the G-13C* basis set level using Gaussian 98 software (46). The charges at the atom centres (appendix 3.1) were fitted to reproduce the electrostatic potential using the RESP program in the AMBER suite and maintain consistency with the AMBER force field charges. Molecular dynamics *in vacuo* of the drug alone showed that the torsion angles of the aromatic rings are inherently non-planar and the piperazine ring has some increased flexibility.

The starting model structure canonical B-DNA of the dodecamer d(CTTTTGCAAAG)₂ was generated using the NUCGEN module of AMBER 4.1. Counterions (20 Na⁺) were added around the phosphates to neutralise charges and the system was solvated by a periodic box of 216 TIP3P molecule waters to a minimum distance of 5 Å around the solute. This resulted in a solvent box size of 40 Å x 40 Å x 60 Å, solvating the complex with around 1700 water molecules. The two drug molecules were docked in three different positions to avoid biasing the binding sites: position a) into the minor groove and positioned by observing several NOE distances, positions b) and c) place the drug by one base pair along the groove and out of the groove by approximately 3 Å.

Minimisation was performed with a 50 step steepest descent and a 5000 step conjugate gradient to water and counterions first with the DNA and drug coordinates frozen, followed by a further 5000 steps on all the components of the system. Next molecular dynamics were run at 100 K on the water alone with the DNA, the drug

and counterions constrained to their initial coordinates with a force constant of 100 kcal mol⁻¹ Å⁻². After this initial equilibration subsequent dynamics were carried out using the particle mesh Ewald method within AMBER 4.1. Successive 10 ps of dynamics, with the PME method turned on, allowed the counterions first and then the drug to move, but the DNA remained constrained. In further 5 ps of dynamics the systems were heated to 300 K and held there for another 5 ps, giving a total of 40 ps equilibration.

In the following 10 ps of dynamics a set of 56 intermolecular restraints were introduced gradually with the drug molecules fully free but the DNA still constrained with a force constant of 100 kcal mol⁻¹ Å⁻². After a further 40 ps of dynamics with the same conditions the system was cooled down from 300 K to 1 K over 5 ps with the restraints on and held there at 1 K for 5 ps. The systems were minimised with no restraints applied and RMSD values were calculated between the starting and final minimised structures using MolMol software.

One structure from the three complexes that satisfied the intermolecular distance restraints best was selected for the continuation of dynamics and allowed to run for 1 ns. The temperature of the system was ramped to 300 K from 1 K over a period of 10 ps with the DNA constrained with a force constant of 100 kcal mol⁻¹ Å⁻². At the same time, the whole set of drug–DNA and DNA–DNA distance restraints were introduced with a force constant of 30 kcal mol⁻¹ Å⁻². In successive 10 ps runs the dynamics of the complex continued with the force constant on the DNA reducing to 50, 25, 10, 5 and finally 2 kcal mol⁻¹ Å⁻². Restrained dynamics followed for 1 ns with the drug and DNA fully free. Snapshots of the simulations were collected every ps and the structures were determined to be equilibrated by RMSD analysis. The energy penalty from distance restraint violations was 21.36 kcal mol⁻¹ with only 28

distance violations. The maximum distance violation was 0.24 Å, 8 restraints were violated by < 0.1 Å, 14 restraints by < 0.2 Å, and 6 restraints by < 0.24 Å. Helicoidal parameters were analysed with CURVES 5.1 and average structures were calculated over the equilibrated trajectories with the CARNAL module of AMBER 4.1.

2.3 Results and Discussion

2.3.1 NMR assignments of the free d(CTTTTGCAAAAG)₂

Proton resonances of both exchangeable and non-exchangeable protons were assigned using well-established procedures (47). Nucleotides are numbered sequentially from the 5'-end of the dodecamer d-(C1-T2-T3-T4-T5-G6-C7-A8-A9-A10-A11-G12). Firstly, the scalar spin-spin couplings observed in TOCSY and DQF-COSY permitted almost complete identification of the deoxyribose spin-systems, although there is significant overlap of the H3'-H4' and H4'-H5'/H5'' peaks. Scalar couplings also permitted the identification of cytosine H5-H6 and thymine CH₃-H6 resonances. Secondly, through-space interactions observed in NOESY spectra allowed sequential assignment of the base and the deoxyribose spin systems. Chemical shift assignments of the DNA protons are presented in appendix 1.1.

The highlighted regions A and B of the 200 ms NOESY spectrum at 298 K in figure 2.3 include inter- and intranucleotide connectivities of H6/H8(i)-H1'(i)-H6/H8(i+1) and H6/H8(i)-H2'/2''(i)-H6/H8(i+1), respectively. Sequential assignment in figure 2.3A starts with the cross peak between internucleotide protons C1H6-C1H1', proceeds to the cross peak between intranucleotide protons C1H1'-T2H6 and continues to the next neighbouring nucleotide along the sequence.

In figure 2.3B sequential connectivities between H6/H8 protons and H2'/H2'' protons are assigned independently using the same procedure. The distinction between H2' and H2'' protons was achieved by examination of the intensity of the NOE cross peaks between these protons and H1' protons, as was suggested previously (47). The distance between H1'–H2' protons is always longer than the distance between H1'–H2'' protons, in respect to predominantly S-type sugar conformation indicated by DQF-COSY and TOCSY data. Sequential NOEs AH2(i)–AH2(i+1) are observed between neighbouring adenines along the AT-tract in the duplex and are characteristic of base stacking interactions in AT-tracts. There are also interstrand NOEs between adenine H2 and deoxyribose H1' of the complementary nucleotide to the 5' neighbouring residue such as A9H2–T5H1', A10H2–T4H1' and A11H2–T3H1'. These indicate that the AT-tract is highly propeller twisted, figure 2.3.

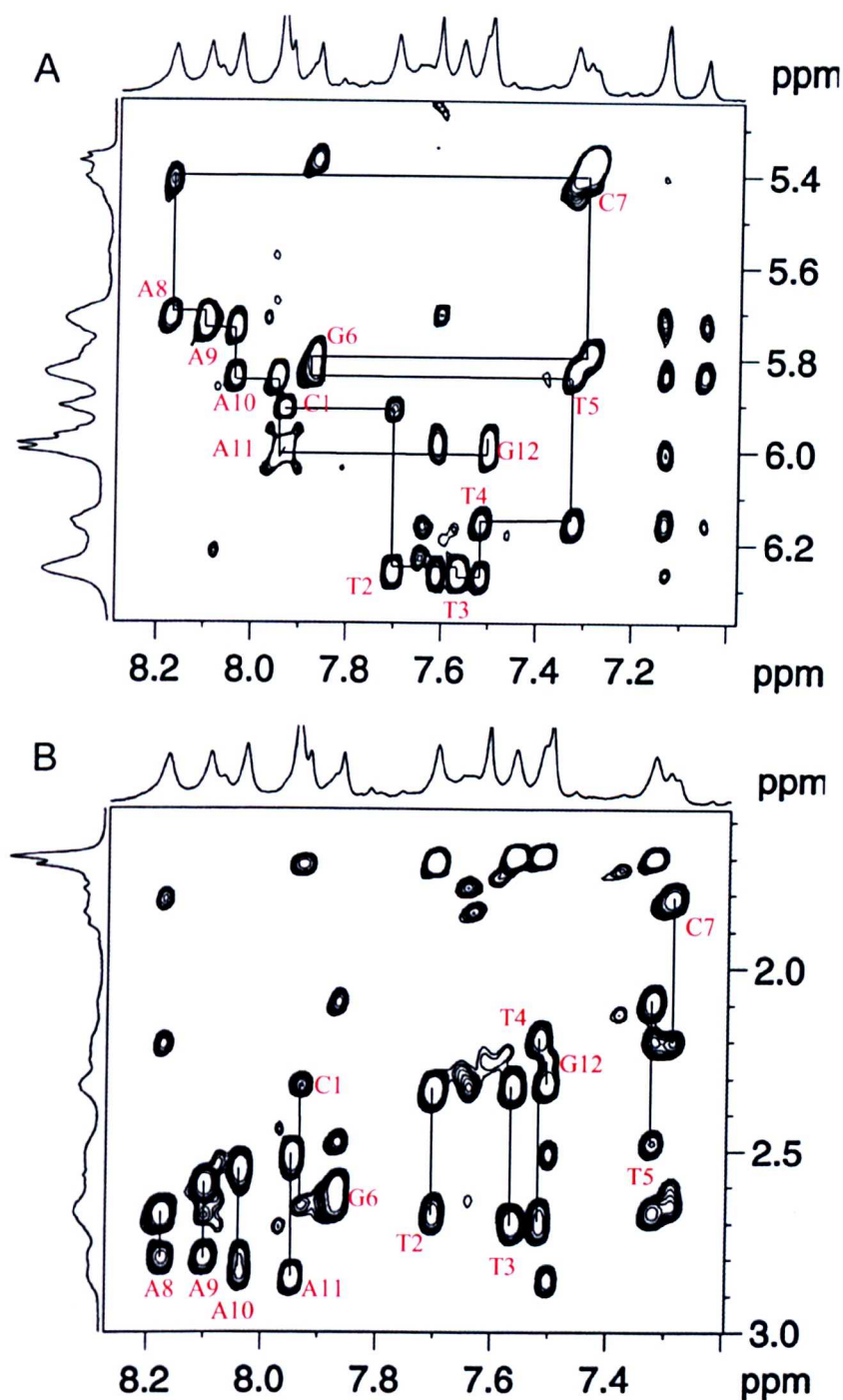
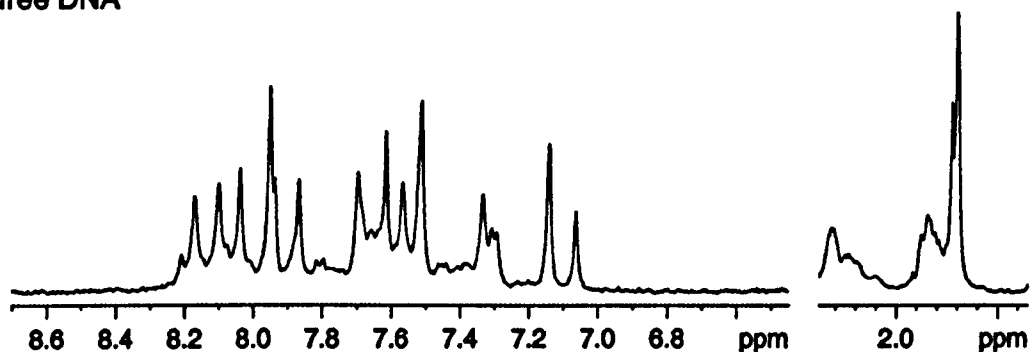


Figure 2.3 Expanded regions of the 200 ms NOESY spectrum of $d(CTTTTGCAAAAG)_2$ illustrating sequential assignment pathways between A) base H6/H8 and deoxyribose H1' protons and B) base H6/H8 and deoxyribose H2'/H2'' protons.

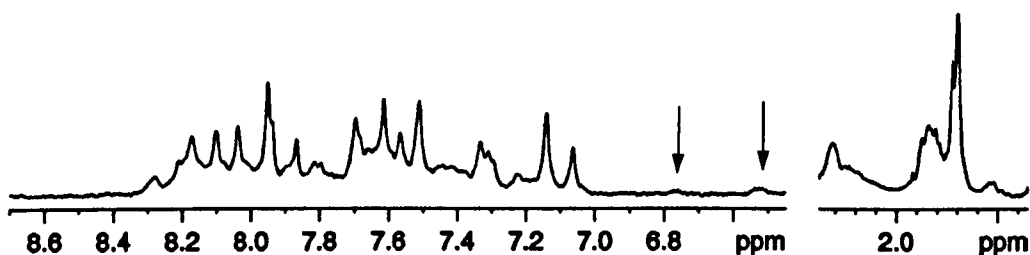
2.3.2 NMR titration of H33258 to d(CTTTTGCAAAAG)₂

Figure 2.4 shows portions of the 1D ¹H-NMR spectra of the free d(CTTTTGCAAAAG)₂ and the different ratios of Hoechst 33258–d(CTTTTGCAAAAG)₂ complex. Upon drug titration, resonances corresponding to the unbound DNA decreased and finally disappeared and new resonances corresponding to the bound DNA appeared. At the point of one equivalent of Hoechst 33258 per duplex there are resonances corresponding to both the free and bound DNA. Observing easily distinguished regions of the spectrum, such as the methyl region (1.3–1.8 ppm) and the aromatic region, enables new resonances to be followed. Resonances of the complex formed can also be observed in previously unoccupied spectral regions (6.5–6.0 ppm). Addition of further quantities of the drug results in a decrease of the intensity of the signals of the free DNA until, at a ratio of two H33258 molecules per duplex, a single species in solution is retained. This is best illustrated from the thymine methyl resonances between 1.0–1.5 ppm, where in the 2:1 complex four methyl resonances are clearly resolved while at 1:1 ratio only resonances from the 2:1 complex and the free DNA are detected. The observation of four thymine methyl peaks in the NMR spectrum of the 2:1 complex confirms that two asymmetric H33258 molecules bind the duplex in symmetrical orientation and so the duplex keeps the dyad symmetry of its palindromic sequence, figure 2.5. Binding in parallel orientations would give rise to a doubling of oligonucleotide resonances as this would lift the dyad symmetry of the duplex. The absence of any recordable resonance of the 1:1 complex and the presence of the 2:1 complex from low drug:duplex ratio such as 0.25:1 during the titration suggests that the two H33258 molecules bind the duplex d(CTTTTGCAAAAG)₂ in a highly co-operative manner.

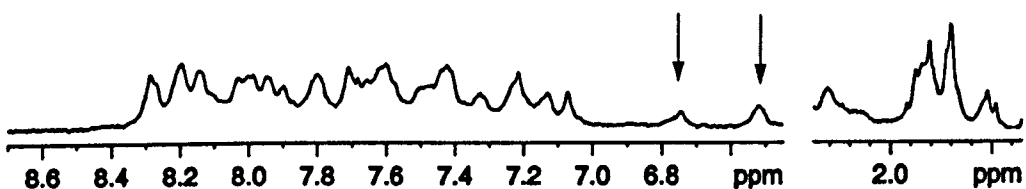
free DNA



0.5:1 ratio



1:1 ratio



2:1 ratio

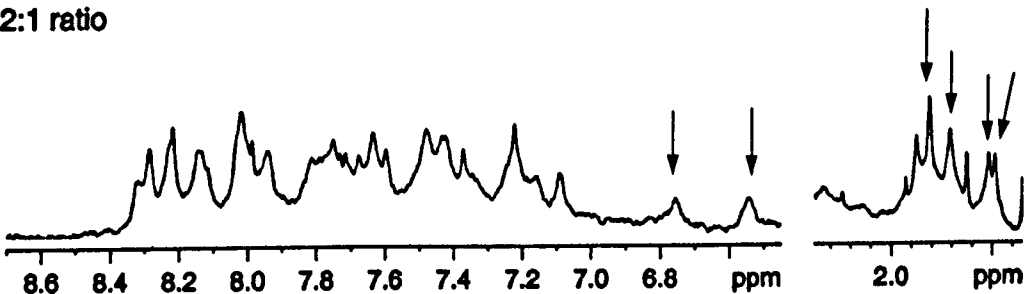


Figure 2.4 Titration of H33258 with d(CTTTTGCAAAAG)₂. ¹H-NMR spectra recorded at 298 K illustrate the thymine methyl region and the aromatic region for different drug:duplex ratios.



Figure 2.5 Scheme of the two types of symmetrical binding of H33258 molecules along the duplex. The head of the arrow represents the bulky N-methylpiperazine ring of H33258.

2.3.3 NMR assignment of the 2:1 complex

Using scalar spin spin couplings observed from the TOCSY spectrum along with through space interactions via NOE data it was possible to identify the deoxyribose spin systems of the duplex. The NOESY spectrum, figure 2.6, of the complex in D₂O solution recorded at 300 ms mixing time was fully assigned by following the same sequential pathways described for the free duplex. In figure 2.7, two expanded regions A and B of the 300 ms NOESY spectrum are illustrated showing the inter- and intranucleotide sequential connectivities of H6/H8(i)–H1'(i)–H6/H8(i+1) and H6/H8(i)–H2'/2''(i)–H6/H8(i+1). Figure 2.7A shows the correlation between H6(i)–CH₃(i+1)–H6(i+1) along the C1–T5 segment of the duplex and figure 2.7B indicates a number of intermolecular interactions. In NOESY spectra at 298 K a number of chemical exchange cross peaks are identified between corresponding protons of the two forms of the duplex, suggesting that drug is dissociating slowly from the complex.

Sequential NOEs AH2(i)–AH2(i+1) and interstrand NOEs between adenine H2 and deoxyribose H1' of the complementary nucleotide to the 5' neighbouring residue such as A9H2–T5H1', A10H2–T4H1' and A11H2–T3H1', are also observed as in free DNA. Such interactions are indicative of highly propeller-

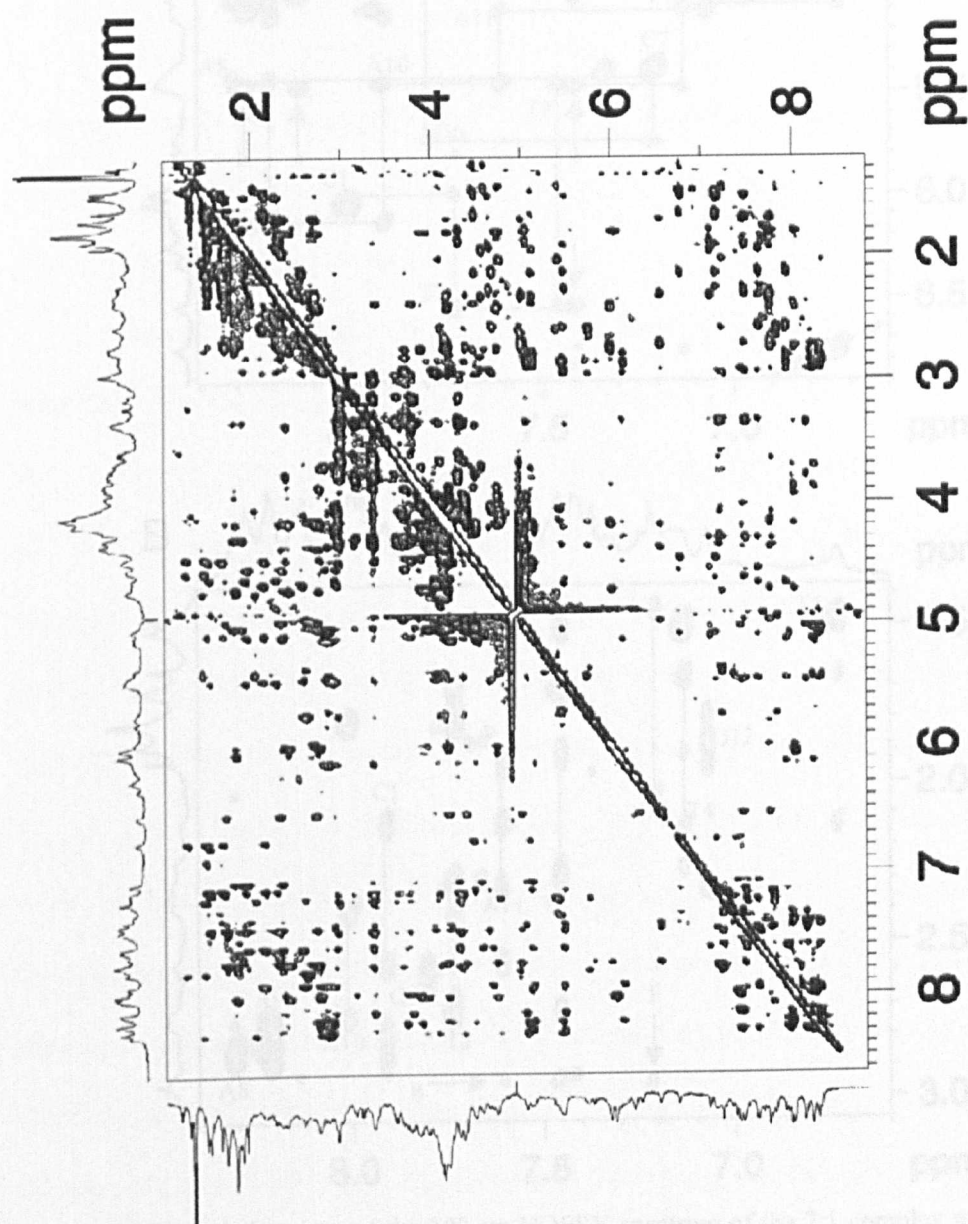


Figure 2.6 NOESY spectrum of the 2:1 H33258-DNA complex in D₂O solution, recorded at 298 K with mixing time 300 ms.

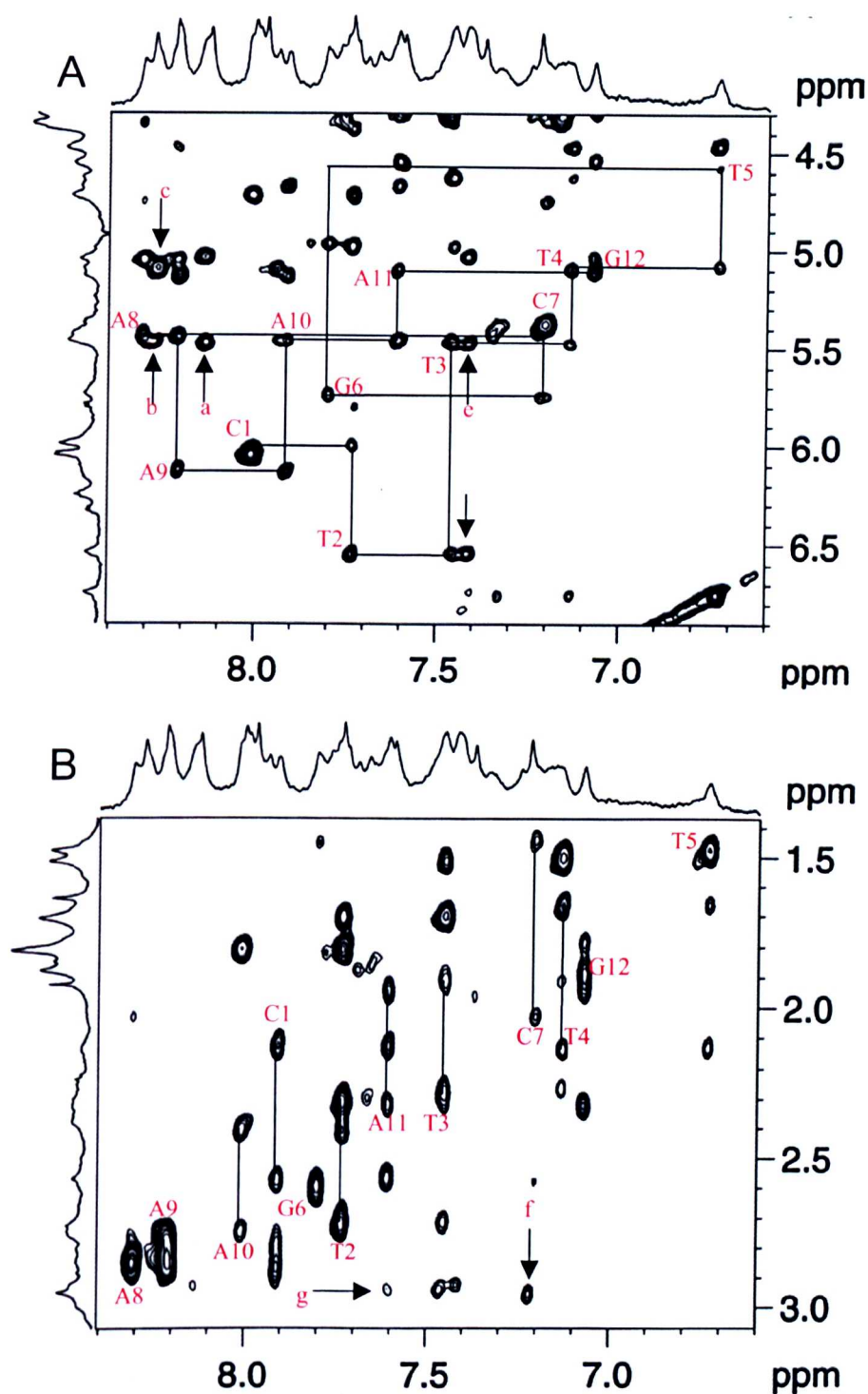


Figure 2.7 Expanded regions of the 300 ms NOESY spectrum of the 2:1 complex recorded at 288 K in D_2O , illustrating sequential assignment pathways A) base H6/H8 and deoxyribose H1' protons and B) base H6/H8 and deoxyribose H2'/H2'' protons. NOEs labelled (a)-(g) are assigned as: (a) ph H3/H5-T3H1', (b) bz1 H4-T3H1', (c) bz1 H4-T4H1', (d) ph H2/H6-T2H1', (e) ph H2/H6-T3H1', (f) pip NMe-A8H2, (g) pip NMe-A9H2.

twisted AT-tract; this is characteristic of H33258–DNA complexes and seems to play a key role in drug specificity and stability (28–30). A high degree of propeller twisting is consistent with the groove width of the AT-tract being significantly narrow. H33258 proton resonances were primarily assigned from the aromatic region of the TOCSY spectrum and from specific intermolecular NOEs with the DNA protons in D₂O and 90% H₂O solutions, following a similar assignment procedure (23). Chemical shift assignments of the DNA and H33258 protons are presented in appendices 1.3 and 1.8.

Previous studies have shown that the differences of the ¹H chemical shift values between the free and the bound form of the DNA indicate the drug-binding site (25,26). In figure 2.8 chemical shift differences $\Delta\delta$ of base resonances H6/H8, H2, H5/CH₃ and deoxyribose H1' are highlighted. Negative shift differences correspond to resonances that move upfield while positive shifts indicate a shift downfield on H33258 binding. The largest chemical shift perturbations are experienced along the AT-tracts. Specifically, the deoxyribose H1' resonances of T2, T3, T5, A11 and G12 have experienced a large upfield shift (0.26–1.25 ppm) while the base H2 resonances of A9, A10 and A11 incur large downfield shifts (0.54–0.80 ppm). In contrast, the resonances of the base protons H6/H8 and H5/CH₃ experience only small shift perturbations. The largest drug-induced perturbations are found on the floor (AH2) and on the walls (H1') of the minor groove and this locates the two drug molecules with some precision in the minor groove along the AT-tracts.

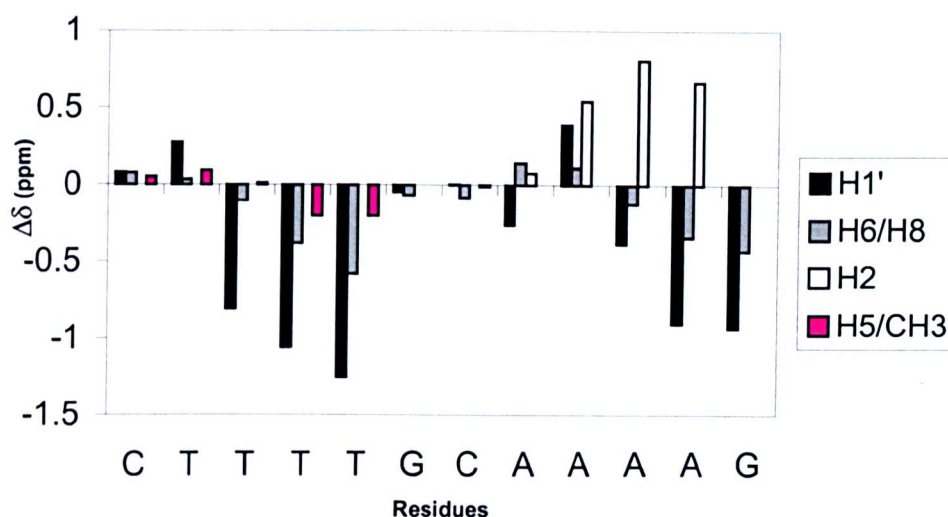


Figure 2.8 H33258 induced chemical shift perturbations of base H2, H6/H8, H2 protons and deoxyribose H1' protons.

The 2:1 complex of H33258-d(CTTTTGCAAAAG)₂ is characterised by 38 intermolecular NOEs between drug and DNA protons. Many of the relevant NOE cross peaks are shown in figure 2.7. The intermolecular NOEs are summarised in table 2.2 and presented schematically in figure 2.9. Specifically, intermolecular NOEs are observed between protons on the concave edge of the drug and protons found on the floor and the walls of the minor groove, such as AH2 and deoxyribose H1' along the full length of the AT-tract. Looking at the schematic representation of the intermolecular NOEs, we conclude unambiguously that drug molecules are positioned within each AT-tract and that the symmetrical binding of the two drugs keeps the N-methylpiperazine groups of the two molecules at the center of the duplex and not at its edges. For example, NOEs from the N-methylpiperazine ring to A8H2, A9H2 and A9H1' at one end of the binding site, and from the phenyl ring protons ph H2,H6 to T14H1', T15H1' and A11H2 at the other end of the A-tract establish that both bound H33258 molecules are located at the centre of the 5'-TTTT and 5'-AAAA sequences each close to the T5–G6 steps. The N-methylpiperazine rings

point towards the centre of the duplex and are accommodated within AT-tracts, while the phenol rings are located close to the 3'-end of the duplex. There is no NOE interaction between the G6 or C7 nucleotides of the duplex and the N-methylpiperazine rings that might indicate a degree of GC recognition by H33258 molecules.

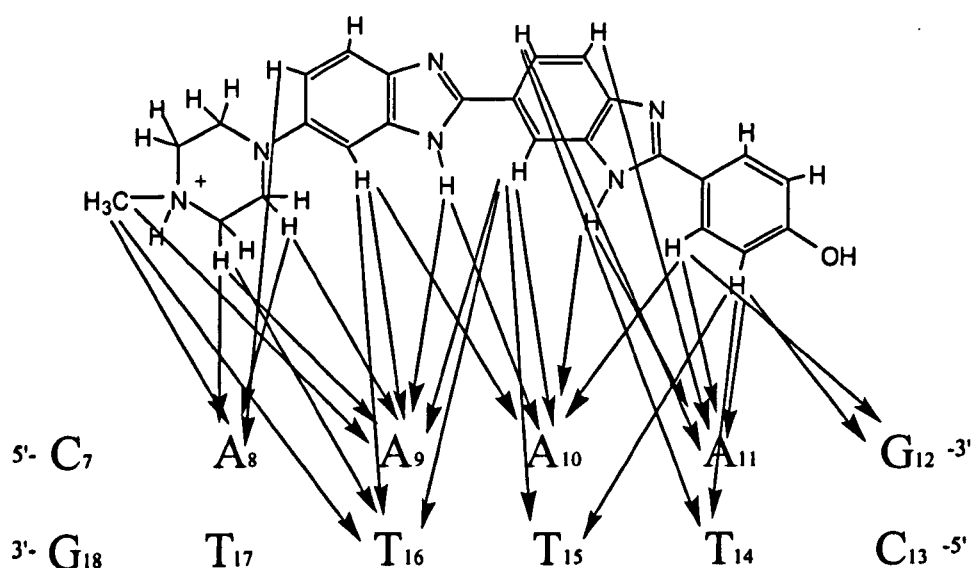


Figure 2.9 Schematic representation of the intermolecular NOEs of the complex.

Table 2.2 Intermolecular NOEs of the 2:1 complex from NOESY recorded in D₂O and H₂O at 288 K.

Drug protons	DNA protons
ph H2/H6	T2H1', T3H1', A11H2, G12H1', G12H5'/5'', T2H3
ph H3/H5	A11H2, A10H2, G12H1', T3H1'
pip H2/H6	A9H2, A8H2, A9H1', A8H5'/H5'', A8H4'
pip H3/H5	A9H2, A9H2, T4H1'
bz1 H4	A10H2, T3H1', T4H1', A9H2, T3H3
bz1 H6	A11H4'
bz1 H7	A11H4'
bz1 H3	A10H2, A11H2
bz2 H4	T2H6, A9H2, A9H1', A10H2, T3H3
bz2 H6	A8H5'/5''
bz2 H3	A10H2, A9H2
pip N-Me	A8H2, A9H2, A9H1'

2.3.4 Structure of the 2:1 complex

The NMR restrained molecular dynamics simulation was judged to equilibrate and converge well within the 1ns of the run as judged by RMSD analysis. The last 500 ps of the simulation were selected for structural analysis. The pairwise RMS deviation from the mean structure over all heavy atoms over the final 500 ps of the simulation was 1.13 (± 0.17) Å. The average structure from the last 500 ps of the simulation of the 2:1 complex is shown in figure 2.10. Each drug is bound across the TTTT tract such that the complex retains a pseudo two-fold symmetry. The aromatic rings of the drug are accommodated well across the TTTT tract with the piperazine rings fitted close to the TG (CA) steps of the duplex. The view into the minor groove illustrates the proximity of the two charged piperazine rings. These are separated by ~ 15 Å (N–N distance) with the intervening cavity filled with water molecules. The pattern of hydrogen-bonding interactions with the floor of the groove is shown in figure 2.11, with the drug forming cross-strand bifurcated interactions between adenine N3 and thymine O2 hydrogen bond acceptors (table 2.3). Such interactions have been observed with H33258 complexes in several X-ray and NMR structures (26,29,31,42). The pattern of interactions in the two binding sites is essentially the same, hydrogen bonds of NH donors of the drug with the N3 of A10 and A11 are generally longer than those with the O2 of T3 and T4 on the complementary strand with standard deviations in these distances over the course of the equilibrated portions of the simulation of ~ 0.3 Å. The angles of the hydrogen bonds N–H...A (A meaning the acceptor atom) shown on table 2.3 agree well with the pattern of the three-centre hydrogen bonds. Small changes in the hydrogen-bond distances and angles during the simulation are apparent due to the dynamic nature of the H33258 binding and do not preclude the tight fit of the aromatic rings of the

drugs in the minor groove. Although the charged piperazinium ring does not form direct hydrogen bonds to the floor of the groove, analysis of the solvation shell during the MD simulation shows persistent water-mediated interactions with the N3 of G6 in both ligand binding sites. Thus, electrostatic interactions mediated by water molecules with the central part of the sequence may be important for the orientation of the drug molecules within the particular complex.

The torsion angles between the phenyl–benzimidazole (ph–bz1), the benzimidazole–benzimidazole (bz1–bz2) and the benzimidazole–piperazine (bz2–pip) rings define the conformation of the bound ligands, figure 2.1. The average values along with standard deviations over the last 500 ps are shown in table 2.3. To maximise groove complementarity and optimise the hydrogen-bond interactions the drug molecules have a slightly twisted conformation. The ph–bz1 and bz1–bz2 torsion angles show only small deviations from planarity ($<20^\circ$ in the mean structure) while standard deviations from the mean values are also relatively small during the course of the dynamics simulations. The planar aromatic rings of both bound drug molecules appear to be conformationally restricted by Van der Waals interactions with the groove walls and hydrogen bonds to the groove floor. In contrast, the bz2–pip torsion angle shows greater dynamic freedom, sampling a range of conformations within the wider region of the groove. The flexibility of the H33258 conformation within the binding site is illustrated in figure 2.12 with three random chosen snapshots of the trajectory. The NMR data are consistent with a degree of flexibility for the N-methylpiperazine and phenol rings as the presence of resonance averaging for ph H2, H6, ph H3, H5, and pip H2, H6, pip H3, H5 is a result of rapid ring-flipping motion within the lifetime of drug binding.

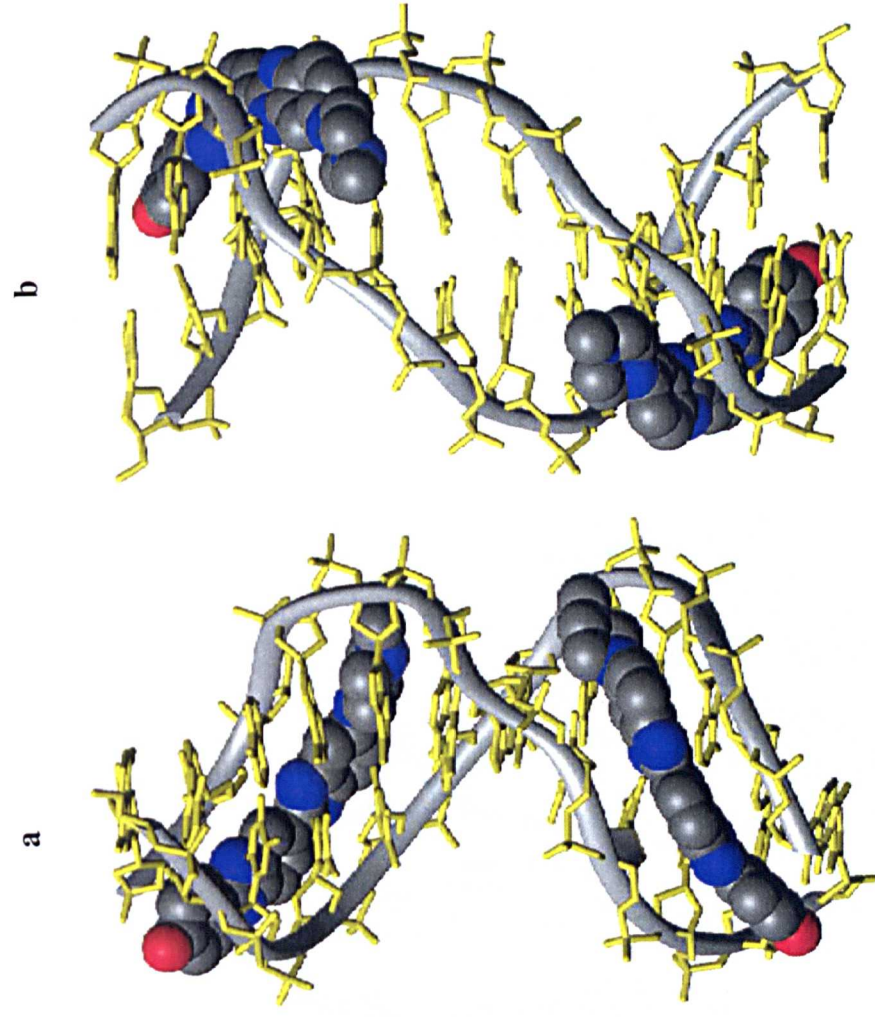


Figure 2.10 Structure of the 2:1 complex of H33258 bound to d(CTTTTGCAAAAG)₂. (a) Illustrating the change in groove width along the sequence with widening at the junctions with the central GC steps (b) view into the minor groove indicating the relative separation of the two N-methylpiperazine rings.

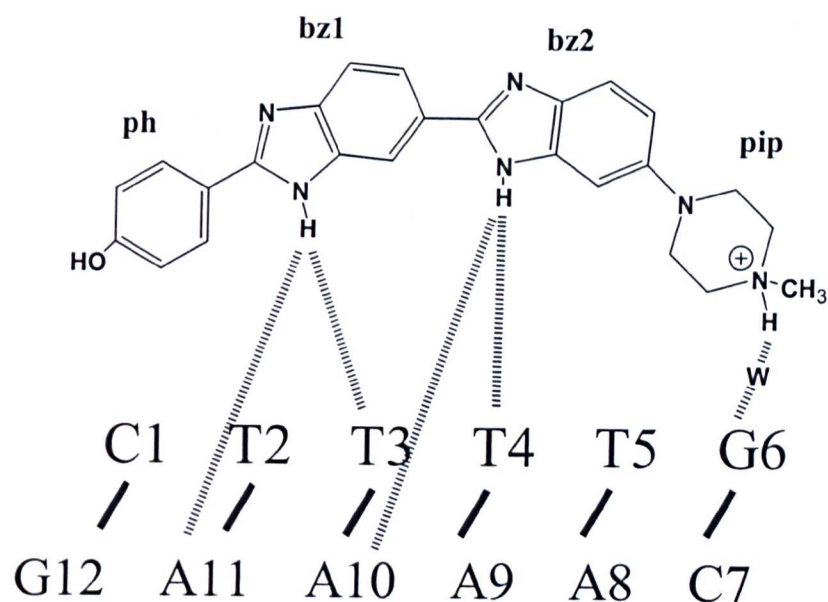


Figure 2.11 Direct and solvent-mediated hydrogen-bonding interactions of H33258 with the floor of the AT minor groove (adenine N3, thymine O2 and w- water).

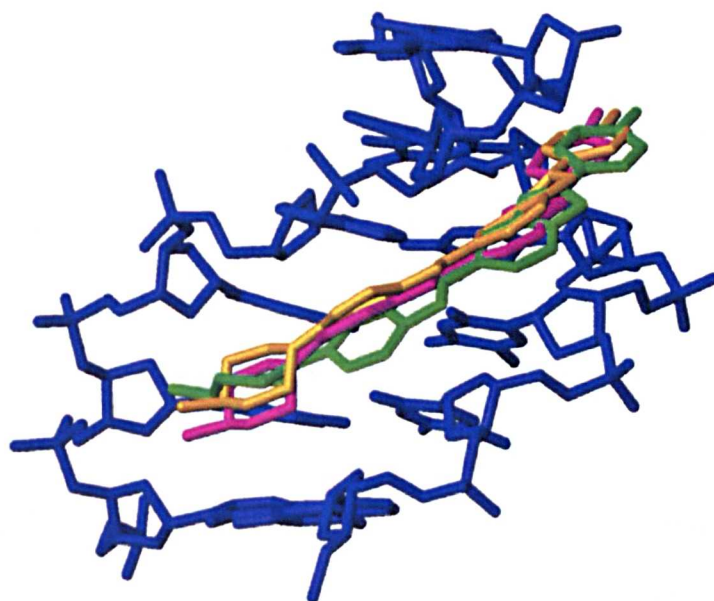


Figure 2.12 Close view of one binding site of the 2:1 complex of H33258 with $d(\text{CTTTTGCAAAAG})_2$ illustrating a) close Van der Waals contacts between the bound ligand and the walls of the minor groove and b) the flexibility of the H33258 conformation within the binding site from three random chosen snapshots of the trajectory.

Table 2.3 Mean intermolecular hydrogen-bonding distances and angles and ligand torsion angles, in the 2:1 complex with d(CTTTTGCAAAAG)₂ taken over the final 500 ps of restrained molecular dynamics simulations.

Hydrogen-bond distance (H—X, Å) ^a and angles [N—H—A(N3)/T(O2), degrees] ^b		
H33258 (1)		
Bz1 NH	A11 N3 2.91 (0.29) [117.8 (13.9)]	T3 O2 2.11 (0.24) [148.5 (17.3)]
Bz2 NH	A10 N3 2.54 (0.33) [132.4 (19.2)]	T4 O2 2.53 (0.31) [129.7 (22.1)]
H33258 (2)		
Bz1 NH	A11 N3 2.99 (0.28) [112.4 (12.0)]	T3 O2 2.09 (0.19) [154.8 (12.3)]
Bz2 NH	A10 N3 2.76 (0.22) [117.6 (12.6)]	T4 O2 2.22 (0.25) [146.3 (13.7)]

Torsion angles (degrees) ^a			
	ph-bz1	bz1-bz2	bz2-pip
H33258 (1)	11.7 (12.1)	9.0 (13.9)	17.4 (39.4)
H33258 (2)	12.3 (11.4)	17.6 (10.3)	-24.3 (15.7)

^a Mean values given with standard deviations in parentheses.

^b Hydrogen-bond angles for bifurcated interactions typically lie between 120–160°.

2.3.5 Conformation and dynamics of bound and free d(CTTTTGCAAAAG)₂

The DNA helical parameters have been analysed over the final 500 ps of the dynamics simulation and average values along with standard deviations calculated, table 2.4. Sequence-dependent variations in structure are illustrated in figure 2.13 for propeller twist, groove width and roll angle. Most striking is the apparent widening of the groove in the 5'→3' direction along both TTTT tracts, culminating in distinct junctions with the central G–C base pairs. The mean propeller twist at the 5'-end of

each TTTT sequence is highly negative ($>20^\circ$). However, it diminishes towards the centre of the sequence. Such a pronounced negative propeller twist across AT-tracts is observed along with particularly narrow minor groove width of duplexes bound tightly with H33258 in many X-ray structures (28–34). Minor groove width, measured from the P_i to P_{i+3} distance (less 5.8 Å), also increases significantly from the 5'-end and is widest across the central G–C step, showing a strong correlation with propeller twist. Widening of the groove is evident from the structures shown in figure 2.10 with the N-methylpiperazine rings sitting at the end of each TTTT tract close to the TpG step in the wider part of the groove. The average increase from the terminal residue to the central GC residues is ~ 3.5 Å for both TTTT tracts. While the roll angles within the T-tracts are relatively small, close to 0° , large positive roll ($>10^\circ$) at the TpG and CpA steps where the piperazine rings are located are clearly evident, typical of pyrimidine-purine steps (48). The above structural parameters among others emphasise strongly the variation in structure between the TTTT tracts and the central GC part of the duplex.

The structural analysis identifies distinct sequence-dependent features in general agreement with earlier conclusions that the A-tract is not uniform in width but tends to widen from the 5'- towards the 3'-end of T_n sequences (49, 50). The N-methylpiperazine ring is accommodated at the end of each of the T-tracts precisely in the region where the groove widens towards the junction with the central GC base pairs. As a consequence we see a distinct orientational preference for H33258, which takes advantage of this sequence-dependent groove widening. Similar bifurcated hydrogen-bonding interactions are possible in either orientation of the H33258 molecules and this should not play an important role in determining the orientational preference. The planar phenyl and benzimidazole portions of the

molecule are accommodated in the region where the minor groove is at its narrowest at 5'-end of the TTTT tract, where intermolecular Van der Waals interactions are maximised. The tight fit between the walls of the groove and the bound ligand are illustrated for one site in figure 2.12.

Table 2.4 Average values and standard deviations for pseudorotation angle, twist, roll, rise, tilt, shift, slide, propeller twist, buckle angle, inclination, x-displacement and minor groove distances. Values calculated from the final 500 ps of the 1 ns restrained simulations of the the 2:1 complex H33258-d(CTTTTGCAAAG)₂ (A), the free DNA d(CTTTTGCAAAG)₂ (B) and the free DNA d(GAAAAGCTTTC)₂ (C).

	A		B		C	
	average	sd	average	sd	average	sd
Pucker (°)	116.0	31.9	121.1	41.8	134.3	34.1
Twist (°)	33.1	4.8	31.6	8.2	30.6	7.0
Roll (°)	3.7	6.6	6.2	11.6	5.3	6.5
Rise (°)	3.7	6.6	3.5	0.4	3.4	0.3
Tilt (°)	3.5	4.0	3.6	4.2	3.0	6.7
Slide (°)	-0.9	0.4	-0.6	0.7	-0.9	0.6
Shift (°)	-0.9	0.4	0.5	0.7	0.2	0.6
Propeller (°)	-16.6	7.9	-23.5	12.9	-10.7	6.2
Buckle (°)	-1.4	10.5	-0.6	17.4	-0.2	8.4
Inclination (°)	2.2	1.5	7.4	2.3	4.1	1.4
X-disp (Å)	-1.7	0.1	-1.7	0.2	-1.9	0.2
P24-P5	4.6	0.5	4.0	0.6	7.9	0.8
P23-P6	4.9	0.5	4.3	0.6	10.3	0.7
P22-P7	7.2	0.8	6.6	1.1	10.3	0.8
P21-P8	8.3	1.0	6.6	0.8	5.3	0.8
P20-P9	8.0	1.0	6.7	0.7	4.9	0.7
P19-P10	8.1	0.6	6.7	1.0	9.4	0.8
P18-P11	4.9	0.5	4.2	1.0	9.7	1.08
P17-P12	4.7	0.4	4.1	1.0	7.2	1.48

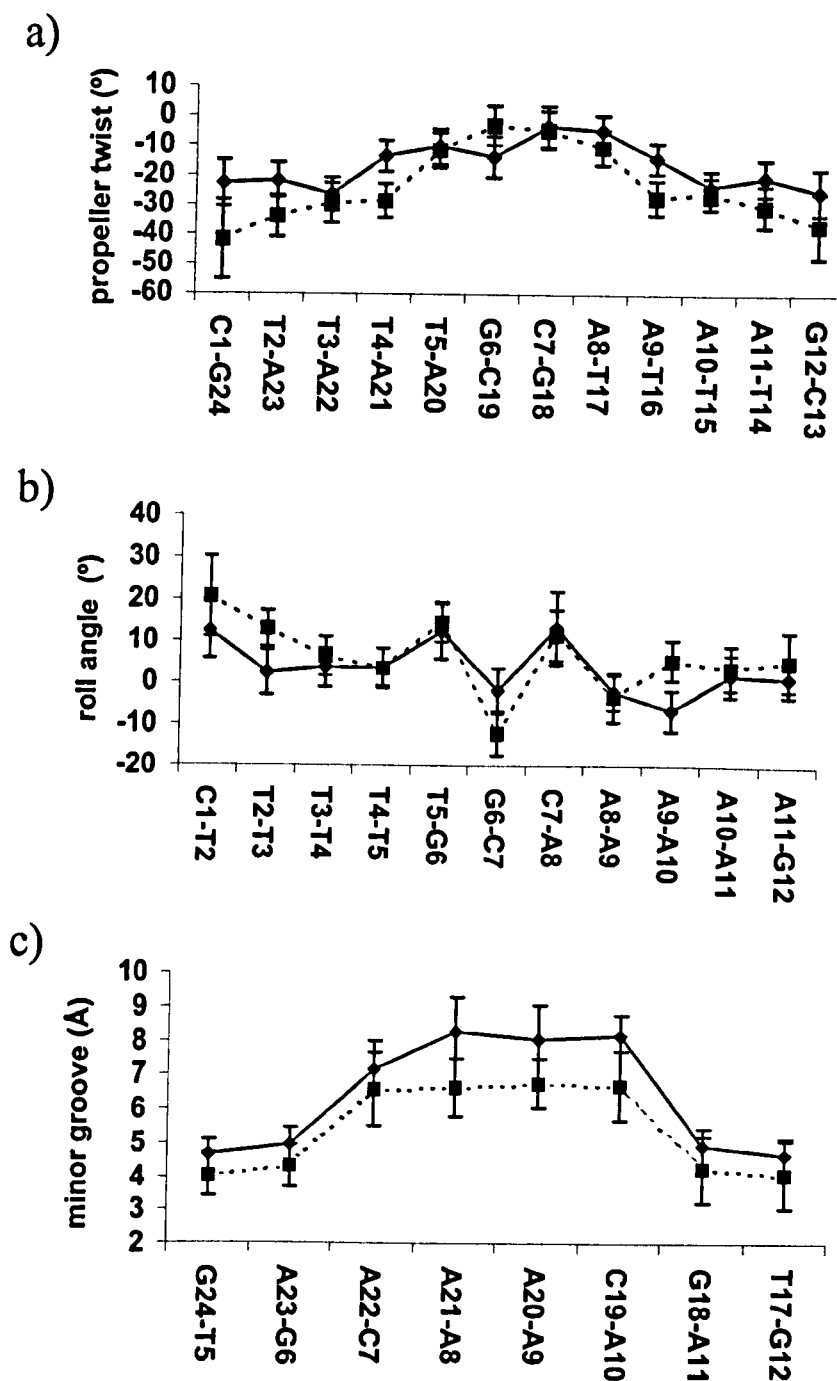


Figure 2.13 Sequence-dependent analysis of propeller twist (a), roll angle (b) minor groove width (P_i-P_{j+3} distance minus 5.8 Å) (c), in the 2:1 complex of H33258 with d(CTTTTGCAAAAG)₂ (solid lines), and for the free DNA (dotted lines); errors bars represent one standard deviation from the mean over the final 500 ps of the 1 ns dynamics simulation.

The structural features identified in the 2:1 complex appear to account reasonably clearly for the orientational preference observed for H33258. However, to what extent are these structural features a consequence of induced fit by the bound ligand, reflecting sequence-dependent DNA flexibility, rather than intrinsic sequence-dependent DNA structure? Moreover, what are the changes in DNA structure from the free to the fully bound state that would allow the co-operative binding of H33258 molecules? To this end, we have calculated structures of the free d(CTTTTGCAAAG)₂ duplex using a very similar NOE-restrained MD protocol to examine these questions in more detail.

Simulation of the free d(CTTTTGCAAAG)₂ duplex was performed on a canonical B-form starting structure for 1 ns utilising a set of 396 NOE distance restraints. A stereo view of 5 structures is illustrated in figure 2.14, showing overlaid snapshots from the rMD simulation taken at 100 ps intervals between 500 and 1000 ps. The RMSD from the mean structure is 0.82 (± 0.13) Å. DNA structural parameters have been calculated over the final 500 ps of the dynamics simulation and average values along with standard deviations calculated and presented in table 2.4. Plots of propeller twist, groove width and roll for d(CTTTTGCAAAG)₂ are shown in figure 2.13 with the data for the 2:1 complex. Many of the structural features identified for the 2:1 complex are also apparent for the structure of the free DNA. A superposition of energy-minimised mean DNA structures for the free DNA and 2:1 complex (figure 2.15) shows a relatively small RMSD over all heavy atoms of 1.55 Å.

Propeller twist is found to be greatest for A–T base pairs ($>30^\circ$) at the 5'-end of the TTTT tract, but this again decreases significantly towards the 3'-end of the tract and across the central GC step. In parallel with this we again see a narrow

groove in the A-tracts, which widens considerably towards the 3'-end of the TTTT sequences suggesting that the two parameters are correlated (51). The minor groove width in free DNA is wider at the central part of the sequence and widens further with the binding of the H33258 molecules allowing optimal fit of the N-methylpiperazine moieties. Analysis of sugar puckers shows that with the exception of the more dynamic terminal residues, the deoxyribose rings adopt C2'-endo/C1'-exo conformations ($P=144-188^\circ$), the typical S configuration of B-DNA. Particularly, T5 and C7 at the TpG (CpA) junctions show a larger standard deviation which indicates a greater degree of dynamic flexibility (see below).

Helical twist also shows some degree of sequence-dependence. While the two A-tracts have high positive twist, the TpG and CpA steps have lower values $< 30^\circ$, in contrast to previous studies where the twist is large ($35^\circ-38^\circ$) at these steps (52, 53), but in good agreement with other H33258 duplexes where binding across alternating pur-pyr sequences (ATAT) is also observed (30). Junctions are clearly visible again at the TpG and CpA steps at the centre of the sequence where the roll angles become large and positive. Roll angle shows a similar but opposite trend to helical twist. Roll increases to a maximum at the TpG step, decreases at the neighbour GpC step and again increase to a maximum at the CpA step. These observations are consistent with positive roll at TpG (CpA) steps, reducing cross-strand steric clashes in the minor groove (48). Helical twist and roll angle appear to be anti-correlated in these simulations, as has been observed for MD simulations of a 30-mer duplex $d(C_6T_4A_4CGT_4A_4G_6)_2$ (54).

The structural features identified in the complex are clearly evident in the structure of the free DNA, suggesting that these are intrinsic sequence-dependent features that appear to account for ligand binding in the observed position and

orientation. In addition, a comparison of the DNA structure in free and fully bound form provides no explanation for the observed co-operativity. Both mean structures are very similar as judged by RMSD analysis and comparison of the fluctuation of their structural features during the equilibrated portions of the rMD simulations. The origins of the co-operativity may be understood when the structure of the 1:1 complex can be determined. This will allow us to characterise any structural changes induced in the second binding site when the first H33258 molecule binds to the DNA. However, to isolate the 1:1 complex by NMR spectroscopy is not possible as the specific system is highly co-operative and 2:1 complex is formed at very low drug:DNA ratios.

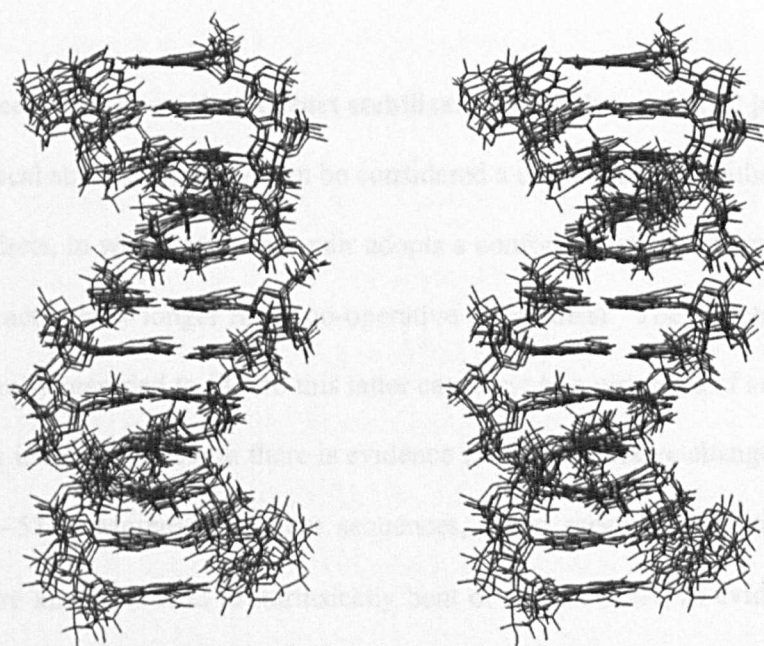


Figure 2.14 Stereoviews of five overlaid structures of d(CTTTTGCAAAAG)₂ taken at 600, 700, 800, 900, 1000 ps during the MD simulation, viewed into the minor groove. The RMSD between structures over all heavy atoms is 0.82 (± 0.13) Å.

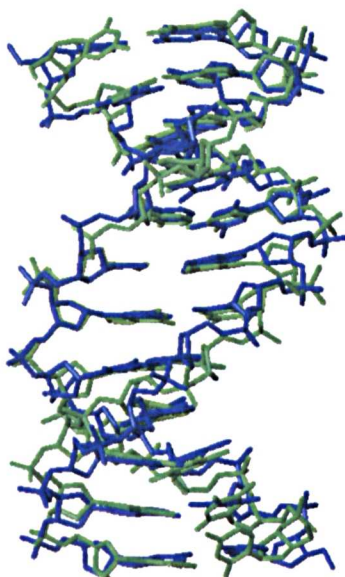


Figure 2.15 Energy-minimised mean structure of the DNA from simulations of the free DNA (green) and the 2:1 complex (blue) (drug removed for clarity), the RMSD over all heavy atoms is 1.55 Å.

2.3.6 Evidence for co-operative A-tract stabilisation with dynamic GC junctions

The local structure of DNA can be considered a consequence of either nearest neighbour effects, in which each base pair adopts a conformation dependent only on adjacent interactions, or longer range co-operative phenomena. The structure of A-tracts is generally regarded to fit into this latter category; as a sequence of successive adenine bases increases in length there is evidence for a co-operative change in helix structure (55–57). Moreover, A-tract sequences, when repeated in phase every helical turn are known to lead to intrinsically bent or curved DNA, as evident from gel retardation and cyclisation kinetic measurements (55). A nucleation length of at least 4 consecutive adenine bases appears to be necessary to induce the "A-tract" structure that is proposed to be stabilised by a minor groove hydration network (58). This solvation network is disrupted by GC rich sequences and also TpA, but not ApT steps. When A_N and T_N tracts are interrupted by intervening GC base pairs they act

as independent structural elements (55, 56), as might be predicted for the sequence d(CTTTTGCAAAAG)₂. We have investigated, from a dynamic perspective, to what extent the formation of co-operative A-tract structure might be significant in the present context of minor groove recognition, and to what extent the junctions with the central GC base pairs identified in the structural analysis disrupt this structure.

Earlier studies have established from measurements of imino proton exchange rates that AT base pairs within A_n tracts (n ≥ 4) show anomalously long lifetimes, reflecting A-tract stabilisation (59). The imino proton spectrum of d(CTTTTGCAAAAG)₂ was examined at various concentrations of ammonia, which is added as an exchange catalyst (56, 59, 60). The imino proton exchange times τ_{ex} measured from the excess line width in the presence and absence of ammonia provide a direct measure of the lifetime of the base-paired (closed) state. When τ_{ex} is extrapolated to infinite catalyst concentration ($1/[\text{NH}_3] = 0$), base-pair opening becomes the rate-limiting step for proton exchange (60). Exchange times (τ_{ex}) versus $1/[\text{NH}_3]$ are plotted in figure 2.16 from data at 15°C. Imino protons of C1–G12 and T2–A11 exchange rapidly as a consequence of end fraying effects, resulting in base pair lifetimes too short to measure (<1 ms), even at 5°C. The lifetime for T3–A10 is ~4 ms, typical of AT base pairs in mixed sequence DNA and, in the present context, may also be influenced by end effects. In contrast, the lifetime of T4–A9 is anomalously long (~23 ms) and similar to that observed for the central AT base pair of a number of A-tract sequences (56), indicative of enhanced stability. The lifetime of the T5–A8 base pair, by comparison, is essentially normal (~6 ms), indicating a breakdown of A-tract stabilisation at the GC junction. The central GC base pairs are much more stable (τ_{ex} ~43 ms), consistent with previously

observed lifetimes (59). The data suggest that $d(\text{CTTTTGCAAAAG})_2$ does have some of the features of co-operative A-tract structures previously identified (56, 59), such as the distinct junction at the TpG (CpA) step, identified in the structural analysis. This is supported by our kinetic measurements of base-pair lifetimes. Breakdown of A-tract structure, and greater intrinsic DNA flexibility, may be important factors in site-specific recognition, and, in the present context orientational specificity, by minor groove binding agents.

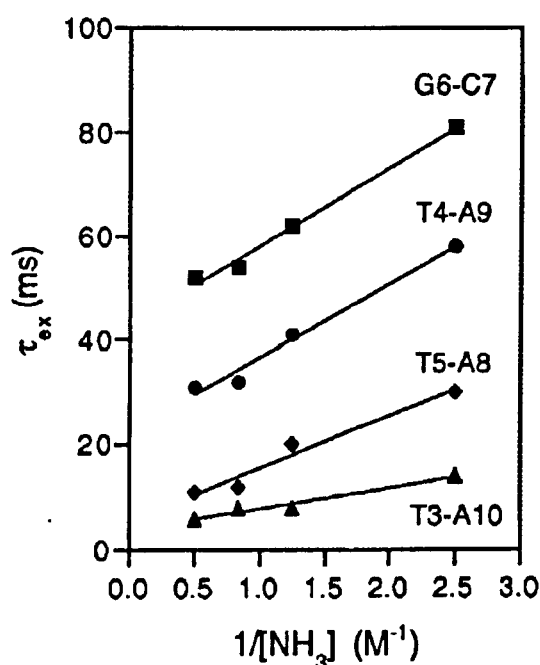


Figure 2.16 Plot of imino proton exchange time τ_{ex} (ms) versus $1/[\text{NH}_3]$ (M^{-1}) for the base pairs of the $d(\text{CTTTTGCAAAAG})_2$ at 15 °C. Extrapolation to $1/[\text{NH}_3] = 0$ provides a measure of the base pair lifetime where imino proton exchange becomes limited only by base pair opening.

2.3.7 Complex of $d(\text{GAAAAGCTTTTC})_2$ with Hoechst 33258

We have carried out complementary studies of the dodecamer duplex $d(\text{GAAAAGCTTTTC})_2$ in which the orientation of the two A-tracts has been inverted while simultaneously eliminating the TpG (CpA) junctions at the ends of these tracts. The duplex $d(\text{GAAAAGCTTTTG})_2$ should be able to accommodate

two bound H33258 molecules within the minor groove of both the A-tracts in a similar way as the duplex d(GTTTTGCAAAC)₂. However, we want to test if the elimination of the TpG (CpA) junctions from the sequence changes the orientational preference of H33258, as we suggested earlier. In addition, can we still observe co-operative binding of H33258 or are the junctions in the sequence important for transferring the information from one site to the other after the initial binding process? Thus we carried out analogous NMR and rMD studies.

¹H-NMR titration studies (see chapter 3) show that H33258 does not bind with the same orientational specificity seen with d(CTTTTGCAAAG)₂. At a ratio of drug to duplex of 2:1, we see evidence for multiple bound species in solution consistent with the drug binding in a number of different orientations. The complexity of the spectrum does not permit us to resolve the mixture of bound species present. It was suggested that lowering the temperature would isolate a symmetrical complex, as maybe the high temperature gives the energy needed for switching between the different orientations or increase the presence of the asymmetric complex. However, low temperatures again resulted in a 1D ¹H-NMR spectrum of broad resonances. 2D NOESY spectra at several temperatures also showed an excess of cross peaks at the aromatic region agreeing with a mixture of bound species. These observations seem to add weight to our conclusion that the TpG junctions strongly influence the orientational preference, while an ApG junction does not result in distinct structural features capable of influencing H33258 minor groove recognition and that the duplex d(GAAAAGCTTTTC)₂ accommodates the molecules easily with either orientation.

We have also carried out full NMR and rMD analysis of the free duplex d(GAAAAGCTTTTC)₂ in order to examine the sequence-dependent features of the

structure and possibly explain the reasons of the observed binding preference of H33258. A table with the assignments of the proton resonances and the NOE distance restraints used for the molecular dynamics simulation can be found in appendices 1.2 and 2.2. The DNA structural parameters have been analysed over the equilibrated portion of the dynamics simulation and results are presented in table 2.4. Ten overlaid snapshots from the rMD simulation are illustrated in figure 2.17 showing a sample of the conformational space occupied by the structure and the variation in minor groove width from two views. The RMSD of the last 500 snapshots from the mean structure over all heavy atoms is 1.15 (± 0.18) Å. Although most of the DNA helical parameters of the d(GAAAAGCTTTTC)₂ duplex have similar mean values to d(CTTTTGCAAAG)₂ there are significant differences in sequence variation of propeller twist and minor groove width.

Propeller twist is found moderately negative between 20°–10° towards the end of the AT-tracts and the central GpC step, while at the ApG and CpT steps values are small ~0°. This duplex with inverted AT-tracts has a wider AT minor groove and narrows significantly towards the 3'-end of the AT-tract sequences, close to the centre of the sequence. The minor groove width (figure 2.18) in this duplex behaves exactly the opposite way compared to the duplex d(CTTTTGCAAAG)₂, as it widens towards AG steps and generally has a wider average groove width. The difference in minor groove width is possibly responsible for the orientational preference of H33258 since there are no distinct structural features to strongly influence the orientation in the complex with d(GAAAAGCTTTTC)₂. It seems that the bulky N-methylpiperazine moiety can fit and allow for optimal interactions with the groove in either orientation within the A-tract.

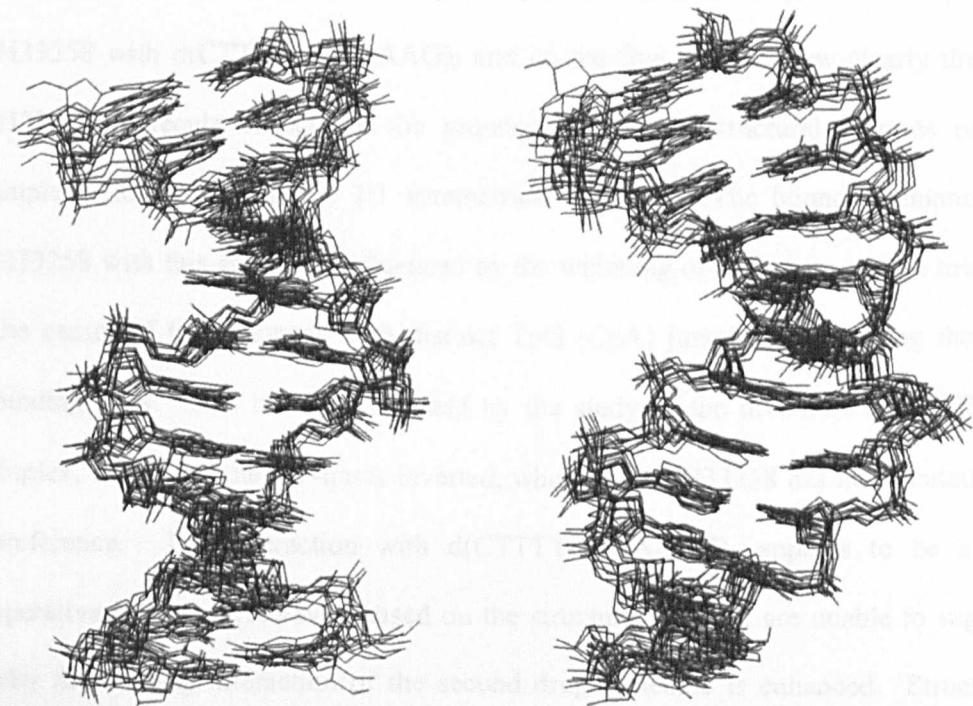


Figure 2.17 Two views of the ten overlaid snapshots from the last 100 ps of the rMD simulation showing the variation in minor groove width along the duplex of d(GAAAAGCTTTTC)₂.

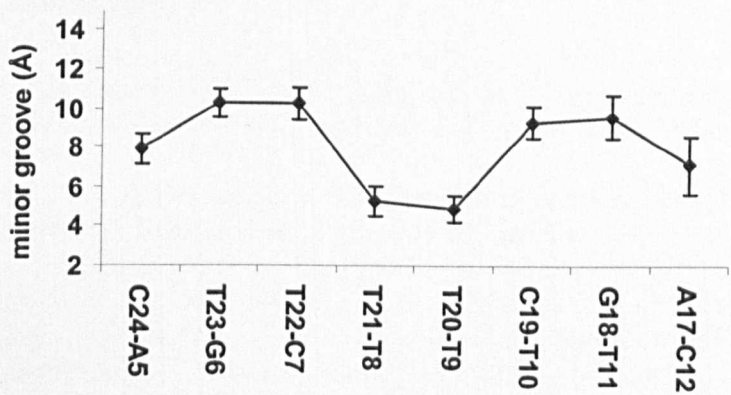


Figure 2.18 Sequence-dependent analysis of minor groove width (P_i-P_{j+3} distance minus 5.8 Å) in the free duplex d(GAAAAGCTTTTC)₂.

2.4 Conclusions

NMR and restrained molecular dynamics studies on the 2:1 complex of H33258 with d(CTTTTGCAAAAG)₂ and on the free duplex show clearly that the H33258 molecule recognises the sequence-dependent structural features of the duplex and forms a single 2:1 symmetrical complex. The bound orientation of H33258 with this duplex is influenced by the widening of the minor groove towards the center of the sequence with distinct TpG (CpA) junctions connecting the two binding sites. This is also confirmed by the study of the d(GAAAAGCTTTTC)₂ duplex, which has the AT-tracts inverted, where bound H33258 has no orientational preference. The interaction with d(CTTTTGCAAAAG)₂ appears to be a co-operative process. However, based on the structural data we are unable to suggest why the binding interaction of the second drug molecule is enhanced. Structural analysis of the duplex shows that binding sites are largely predisposed to accommodate H33258 and this can play a significant role when the first molecule binds the duplex. In the following chapter H33258 co-operativity is investigated further in this system and other duplexes.

2.5 References

1. Krugh, T. R. 1994 Drug–DNA Interactions. *Curr. Opin. Struct. Biol.* **4**, 351-364.
2. Wemmer, D. E. & Dervan, P. B. 1997 Targeting the Minor Groove of DNA. *Curr. Opin. Struct. Biol.* **7**, 355-361.
3. Chaires, J. B. 1998 Drug–DNA Interactions. *Curr. Opin. Struct. Biol.* **8**, 314-320.
4. S. Neidle, M. Waring, Molecular Aspects of Anti-Cancer Drug–DNA Interactions (1994) vol 2, The Macmillan Press LTD.
5. Bell, C. A., Dykstra, C. C., Naiman, N. A., Cory, M., Fairley, T. A. & Tidwell, R. R. 1993 Structure-Activity Studies of Dicationically Substituted Bis-Benzimidazoles against *Giardia-Lambli*a – Correlation of Antigiardial Activity with DNA-Binding Affinity and Giardial Topoisomerase II Inhibition. *Antimicrob. Agents Chemother.* **37**, 2668-2673.
6. Tidwell, R. R., Jones, S. K., Naiman, N. A., Berger, L. C., Brake, W. B., Dykstra, C. C. & Hall, J. E. 1993 Activity of Cationically Substituted Bis-Benzimidazoles against Experimental *Pneumocystis-Carinii* Pneumonia. *Antimicrob. Agents Chemother.* **37**, 1713-1716.
7. Fairley, T. A., Tidwell, R. R., Donkor, I., Naiman, N. A., Ohemeng, K. A., Lombardy, R. J., Bentley, J. A. & Cory, M. 1993 Structure, DNA Minor-Groove Binding, and Base-Pair Specificity of Alkyl-Linked and Aryl-Linked Bis(Amidinobenzimidazoles) and Bis(Amidinoindoles). *J. Med. Chem.* **36**, 1746-1753.
8. Chen, A. Y., Yu, C., Gatto, B. & Liu, L. F. 1993 DNA Minor Groove-Binding Ligands – A Different Class of Mammalian DNA Topoisomerase-I Inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 8131-8135.
9. Patel, S. R., Kvols, L. K., Rubin, J., O'Connell, M. J., Edmonson, J. H., Ames, M. M. & Kovach, J. S. 1991 Phase-I-II Study of Pibenzimol Hydrochloride (NSC-322921) in Advanced Pancreatic-Carcinoma. *Invest. New Drugs* **9**, 53-57.
10. Ellwart, J. W. & Dormer, P. 1990 Vitality Measurement Using Spectrum Shift in Hoechst-33342 Stained Cells. *Cytometry* **11**, 239-243.
11. Loontjens, F. G., McLaughlin, L. W., Diekmann, S. & Clegg, R. M. 1991 Binding of Hoechst 33258 and 4',6-Diamidino-2-Phenylindole to Self-Complementary Decadeoxynucleotides with Modified Exocyclic Base Substituents. *Biochemistry* **30**, 182-189.
12. Denham, D. A., Suswillo, R. R., Rogers, R., McGreevy, P. B., Andrews, B. J. 1976 Studies on *Brugia Pahangi*. 13. The Anthelmintic Effect of Compounds F151 (Friedheim), HOE 33258 (Hoechst) and their Reaction Product. *J. Helminthol.* **50**, 243-250.

13. Screening Data on NSC 322921 (Pibenzimol), National Cancer Institute, Bethesda, USA, 1983.
14. Kraut, E. H., Fleming, T., Segal, M., Neidhart, J. A., Behrens, B. C. & Macdonald, J. 1991 Phase-II Study of Pibenzimol in Pancreatic-Cancer – A Southwest Oncology Group-Study. *Invest. New Drugs* **9**, 95-96.
15. Reddy, B. S. P., Sondhi, S. M. & Lown, J. W. 1999 Synthetic DNA Minor Groove-Binding Drugs. *Pharmacol. Ther.* **84**, 1-111.
16. Singh, A. K. & Lown, J. W. 2000 Design, Synthesis and Antitumor Cytotoxicity of Novel Bis-Benzimidazoles. *Anti-Cancer Drug Des.* **15**, 265-275.
17. Harshman, K. D. & Dervan, P. B. 1985 Molecular Recognition of B-DNA by Hoechst 33258. *Nucleic Acids Res.* **13**, 4825-4835.
18. Portugal, J. & Waring, M. J. 1988 Assignment of DNA-Binding Sites for 4',6-Diamidine-2-Phenylindole and Bisbenzimidazole (Hoechst-33258) – A Comparative Footprinting Study. *Biochimica Et Biophysica Acta* **949**, 158-168.
19. Jorgenson, K. F., Varshney, U. & Vandesande, J. H. 1988 Interaction of Hoechst-33258 with Repeating Synthetic DNA Polymers and Natural DNA. *J. Biomol. Struct. Dyn.* **5**, 1005-1023.
20. Colson, P., Houssier, C. & Bailly, C. 1995 Use of Electric Linear Dichroism and Competition Experiments with Intercalating Drugs to Investigate the Mode of Binding of Hoechst-33258, Berenil and Dapi to GC Sequences. *J. Biomol. Struct. Dyn.* **13**, 351-366.
21. Loontjens, F. G., McLaughlin, L. W., Diekmann, S. & Clegg, R. M. 1991 Binding of Hoechst 33258 and 4',6-Diamidino-2-Phenylindole to Self-Complementary Decadeoxynucleotides with Modified Exocyclic Base Substituents. *Biochemistry* **30**, 182-189.
22. Haq, I., Ladbury, J. E., Chowdhry, B. Z., Jenkins, T. C. & Chaires, J. B. 1997 Specific Binding of H33258 to the d(CGCAAATTTGCG)₂ Duplex. Calorimetric and Spectroscopic Studies. *J. Mol. Biol.* **271**, 244-257.
23. Searle, M. S. & Embrey, K. J. 1990 Sequence-Specific Interaction of Hoechst-33258 with the Minor Groove of an Adenine-Tract DNA Duplex Studied in Solution by H¹ NMR Spectroscopy. *Nucleic Acids Res.* **18**, 3753-3762.
24. Parkinson, J. A., Barber, J., Douglas, K. T., Rosamond, J. & Sharples, D. 1990 Minor-Groove Recognition of the Self-Complementary Duplex d(CGCGAATTCGCG)₂ by Hoechst-33258 – A High-Field NMR-Study. *Biochemistry* **29**, 10181-10190.
25. Fede, A., Labhardt, A., Bannwarth, W. & Leupin, W. 1991 Dynamics and Binding Mode of Hoechst 33258 to d(GTGAATTCAC)₂ in the 1-1 Solution

- Complex as Determined by 2-Dimensional H-1 NMR. *Biochemistry* **30**, 11377-11388.
26. Embrey, K. J., Searle, M. S. & Craik, D. J. 1993 Interaction of Hoechst 33258 with the Minor Groove of the AT-Rich DNA Duplex d(GGTAATTACC)₂ Studied in Solution by NMR-Spectroscopy. *Eur. J. Biochem.* **211**, 437-447.
 27. Higgins, L. D. & Searle, M. S. 1999 Site-specificity of Bis-Benzimidazole Hoechst 33258 in A-tract Recognition of the DNA Dodecamer Duplex d(GCAAATTTTGC)₂. *Chem. Commun.*, 1861-1862.
 28. Pjura, P. E., Grzeskowiak, K. & Dickerson, R. E. 1987 Binding of Hoechst-33258 to the Minor Groove of B-DNA. *J. Mol. Biol.* **197**, 257-271.
 29. Teng, M., Usman, N., Frederick, C. A. & Wang, A. H. J. 1988 The Molecular-Structure of the Complex of Hoechst-33258 and the DNA Dodecamer d(CGCGAATTCGCG)₂. *Nucleic Acids Res.* **16**, 2671-2690.
 30. Carrondo, M., Coll, M., Aymami, J., Wang, A. H. J., Vandermarel, G. A., Vanboom, J. H. & Rich, A. 1989 Binding of a Hoechst Dye to d(CGCGATATCGCG)₂ and its Influence on the Conformation of the DNA Fragment. *Biochemistry* **28**, 7849-7859.
 31. Quintana, J. R., Lipanov, A. A. & Dickerson, R. E. 1991 Low-Temperature Crystallographic Analyses of the Binding of Hoechst-33258 to the Double-Helical DNA Dodecamer C-G-C-G-A-A-T-T-C-G-C-G. *Biochemistry* **30**, 10294-10306.
 32. Vega, M. C., Saez, I. G., Aymami, J., Eritja, R., Vandermarel, G. A., Vanboom, J. H., Rich, A. & Coll, M. 1994 3-Dimensional Crystal-Structure of the A-Tract DNA Dodecamer d(CGCAAATTTGCG)₂ Complexed with the Minor-Groove-Binding Drug Hoechst-33258. *Eur. J. Biochem.* **222**, 721-726.
 33. Spink, N., Brown, D. G., Skelly, J. V. & Neidle, S. 1994 Sequence-Dependent Effects in Drug-DNA Interaction – The Crystal-Structure of Hoechst-33258 Bound to the d(CGCAAATTTGCG)₂ Duplex. *Nucleic Acids Res.* **22**, 1607-1612.
 34. Clark, G. R., Squire, C. J., Gray, E. J., Leupin, W. & Neidle, S. 1996 Designer DNA-Binding Drugs: The Crystal Structure of a Meta-Hydroxy Analogue of Hoechst 33258 Bound to d(CGCGAATTCGCG)₂. *Nucleic Acids Res.* **24**, 4882-4889.
 35. Hud, N. V., Sklenar, V. & Feigon, J. 1999 Localization of Ammonium Ione in the Minor Groove of DNA Duplexes in Solution and the Origin of DNA A-tract Bending. *J. Mol. Biol.* **286**, 651-660.
 36. Hamdan, I., Skellern, G. G. & Waigh, W. D. 1998 Use of Capillary Electrophoresis in the Study of Ligand-DNA Interactions. *Nucleic Acids Res.* **26**, 3053-3058.

37. Pelton, J. G. & Wemmer, D. E. 1989 Structural Characterization of a 2-1 Distamycin A.d(CGCAAATTGGC)₂ Complex by Two-Dimensional NMR. *Proc. Natl. Acad. Sci. U. S. A.* **86**, 5723-5727.
38. Wemmer, D. E. & Dervan, P. B. 1997 Targeting the Minor Groove of DNA. *Curr. Opin. Struct. Biol.* **7**, 355-361.
39. Cantor, C. R., Warshaw, M. M., and Shapiro, H. 1970 Oligonucleotide Interactions III. Circular Dichroism Studies of the Conformation of Deoxyoligonucleotides. *Biopolymers* **9**, 1059-1077.
40. Baleja, J. D., Moulton, J. & Sykes, B. D. 1990 Distance Measurement and Structure Refinement with NOE Data. *J. Magn. Reson.* **87**, 375-384.
41. Koradi, R., Billeter, M. & Wuthrich, K. 1996 MOLMOL: A Program for Display and Analysis of Macromolecular Structures. *J. Mol. Graph.* **14**, 51-55.
42. Bostock-Smith, C. E., Laughton, C. A. & Searle, M. S. 1998 DNA Minor Groove Recognition by a Tetrahydropyrimidinium Analogue of Hoechst 33258: NMR and Molecular Dynamics Studies of the Complex with d(GGTAATTACC)₂. *Nucleic Acids Res.* **26**, 1660-1667.
43. Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. S., Cheatham III, T. E., Ferguson, D.M., Seibel, G. L., Singh, U. C., Weiner, P. K. & Kollman, P. A. 1995 AMBER 4.1, University of California, San Francisco.
44. York, D. M., Yang, W. T., Lee, H., Darden, T. & Pedersen, L. G. 1995 Toward the Accurate Modeling of DNA – The Importance of Long-Range Electrostatics. *J. Am. Chem. Soc.* **117**, 5001-5002.
45. Lavery, R. & Sklenar, H. 1989 Defining the Structure of Irregular Nucleic-Acids – Conventions and Principles. *J. Biomol. Struct. Dyn.* **6**, 655-667.
46. Bayly, C. I., Cieplak, P., Cornell, W. D. & Kollman, P. A. 1993 A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges - The RESP Model. *J. Phys. Chem.* **97**, 10269-10280.
47. Chazin, W. J., Wuthrich, K., Hyberts, S., Rance, M., Denny, W. A. & Leupin, W. 1986 H¹-Nuclear Magnetic-Resonance Assignments for d(GCATTAATGC)₂ Using Experimental Refinements of Established Procedures. *J. Mol. Biol.* **190**, 439-453.
48. Hunter, C. A. 1993 Sequence-Dependent DNA-Structure – The Role of Base Stacking Interactions. *J. Mol. Biol.* **230**, 1025-1054.
49. Burkhoff, A. M. & Tullius, T. D. 1988 Structural Details of an Adenine Tract That Does Not Cause DNA to Bend. *Nature* **331**, 455-457.
50. Katahira, M., Sugeta, H., Kyogoku, Y., Fujii, S., Fujisawa, R. & Tomita, K. 1988 One-Dimensional and Two-Dimensional NMR-Studies on the Conformation of

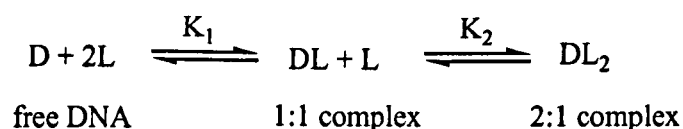
- DNA Containing the Oligo(Da)Oligo(Dt) Tract. *Nucleic Acids Res.* **16**, 8619-8632.
51. Nelson, H. C. M., Finch, J. T., Luisi, B. F. & Klug, A. 1987 The Structure of an Oligo(dA).Oligo(dT) Tract and its Biological Implications. *Nature* **330**, 221-226.
52. Dornberger, U., Flemming, J. & Fritzsche, H. 1998 Structure Determination and Analysis of Helix Parameters in the DNA Decamer d(CATGGCCATG)₂ – Comparison of Results from NMR and Crystallography. *J. Mol. Biol.* **284**, 1453-1463.
53. Baleja, J. D., Germann, M. W., Vandesande, J. H. & Sykes, B. D. 1990 Solution Conformation of Purine–Pyrimidine DNA Octamers Using Nuclear-Magnetic-Resonance, Restrained Molecular-Dynamics and NOE-Based Refinement. *J. Mol. Biol.* **215**, 411-428.
54. Sprou, D., Young, M. A. & Beveridge, D. L. 1999 Molecular Dynamics Studies of Axis Bending in d(G(5)-(GA(4)T(4)C)(2)-C-5) and d(G(5)-(GT(4)A(4)C)(2)-C-5): Effects of Sequence Polarity on DNA Curvature. *J. Mol. Biol.* **285**, 1623-1632.
55. Crothers, D. M., Haran, T. E. & Nadeau, J. G. 1990 Intrinsically Bent DNA. *J. Biol. Chem.* **265**, 7093-7096.
56. Leroy, J. L., Kochoyan, M., Huynhdinh, T. & Gueron, M. 1988 Characterization of Base-Pair Opening in Deoxynucleotide Duplexes Using Catalyzed Exchange of the Imino Proton. *J. Mol. Biol.* **200**, 223-238.
57. Nadeau, J. G. & Crothers, D. M. 1989 Structural Basis for DNA Bending. *Proc. Natl. Acad. Sci. U. S. A.* **86**, 2622-2626.
58. Fratini, A. V., Kopka, M. L., Drew, H. R. & Dickerson, R. E. 1982 Reversible Bending and Helix Geometry in a B-DNA Dodecamer – CGCGAATTCGCG. *J. Biol. Chem.* **257**, 4686-4707.
59. Leroy, J. L., Charretier, E., Kochoyan, M. & Gueron, M. 1988 Evidence from Base-Pair Kinetics for 2 Types of Adenine Tract Structures in Solution – Their Relation to DNA Curvature. *Biochemistry* **27**, 8894-8898.
60. Gueron, M., Kochoyan, M. & Leroy, J. L. 1987 A Single-Mode of DNA Base-Pair Opening Drives Imino Proton-Exchange. *Nature* **328**, 89-92.

3. Co-operative interaction of Hoechst 33258 within the DNA minor groove

3.1 Introduction

The origins of co-operativity in protein–DNA and drug–DNA complexes have generally become evident where structure determination has been performed. Typically, close contacts are observed in the 2:1 complex between the two ligand molecules. Thus, the binding of the second ligand to the 1:1 complex is associated with the formation of a greater number of favourable interactions than the binding of the first ligand to the free DNA. In the case of echinomycin (chapter 1), it is unlikely that co-operativity is mediated by direct contact between drug molecules, but that the drug-induced conformational changes at one site are propagated to the other, through effects on helical twist (helix unwinding) and minor groove width. In contrast, the oligopeptide antibiotic distamycin has demonstrated side-by-side anti-parallel binding to the DNA minor groove, with favourable π - π interactions stabilising the drug dimer (chapter 1). The 2:1 binding mode has been shown to be highly co-operative in the case of binding to AT rich sequences of DNA that do not have an intrinsically narrow minor groove (such as TATAT and AAGTT), where there is poor binding complementarity in the 1:1 complex. In the 2:1 complex, the groove width increases further to accommodate the drug dimer requiring sufficient DNA flexibility to optimise Van der Waals interactions between the drug and the walls of the groove. In this case, there is a clear shape complementarity requirement for co-operative binding that can be rationalised on the basis of a direct interaction between two bound ligand molecules.

For simple cases where two molecules of a ligand bind to the DNA target, a co-operativity index ($K_{\text{co-op}}$) can be defined (scheme 3.1) as the ratio of the equilibrium constants K_1 and K_2 for the two binding events in the formation of the 2:1 complex [ligand(L):DNA(D)]. If $K_{\text{co-op}} = 1$, then there is no interaction between sites where if $K_{\text{co-op}} > 1$, sites are co-operatively linked.



$$K_{\text{co-op}} = \frac{K_2}{K_1}$$

Scheme 3.1 Equilibria involved in the formation of the 2:1 ligand–DNA complex and definition of $K_{\text{co-op}}$.

3.2 Hoechst 33258 co-operative binding to DNA

In chapter 2, we described the titration of the DNA dodecamer duplex d(CTTTTGCAAAAG)₂ with H33258 that is followed by NMR. Signals of the free DNA are replaced by those of the 2:1 drug–DNA complex, without detection of any intermediate 1:1 complex. In terms of the equilibria involved (scheme 3.1) and the practical limits of NMR sensitivity, we have estimated a lower limit on the co-operativity index, $K_{\text{co-op}}$, of approximately 1000, which equates to a $\Delta\Delta G$ for the two binding events of -4 kcal mol^{-1} . Interestingly, the NMR-derived structures of the free DNA and 2:1 complex do not, in this case, give us any insight into the origins of co-operativity in this system. The structure of the complex shows that the ligands are not in contact and are oriented in the two AT-tracts in such a way that their positively charged piperidine rings face each other across the intervening GC step with the charged centres separated by $\sim 15 \text{ \AA}$. Restrained MD calculations on the 2:1

complex, using an explicit solvent model, indicate that the intervening groove is occupied by solvent and that water molecules may be involved in mediating electrostatic interactions with the floor of the groove, as well as screening the two positive charges from each other. The width of the minor groove in A_n sequences tends to decrease from their 5'-end, and this is observed in this complex, with the bulky piperazine ring of the drug located in the wider part of the groove close to the TpG step. Thus, the orientation of the ligands in this structure appears to be, at least in part, sterically driven.

In the present chapter, we present our efforts to establish co-operative binding of H33258 with other methods than NMR titration and also to investigate further the origins of co-operativity by studying the interaction of H33258 with other oligonucleotides containing adjacent binding sites. In the attempt to shed some light on the origins of co-operativity in this system, we have also used extended molecular dynamics simulations to study the free DNA, the 2:1 complex, and also the theoretical 1:1 complex. In collaboration with Dr. Sarah Harris, Dr. Charlie Laughton from the School of Pharmaceutical Sciences, University of Nottingham, we used the simulation data and we have been able to calculate thermodynamic quantities relating to the two binding events, which are not available using other methods. The thermodynamic properties of each system suggest an explanation for the observed co-operativity.

3. 3 Methods for detecting co-operativity in ligand–DNA recognition

3.3.1 NMR titration

Co-operativity in ligand–DNA recognition has been detected by a number of experimental methods. NMR spectroscopy is probably the most common technique

that has been used for observing co-operative binding of ligands to short oligonucleotides. NMR has also the advantage to allow further detailed structure determination of the resulting complex, which in most cases has been proven necessary to understand the origins of the co-operativity. However, NMR requires large amounts of the DNA and ligand and the method is limited to short oligonucleotides.

Figure 3.1 shows the 1D ^1H -NMR titration spectra of a distamycin analogue with the decamer $\text{d}(\text{CGTATATAGC})_2$ at 25 °C with 0.0, 0.5, 1.0, 1.5, 2.0 molar equivalents ligand added (1). Upon addition of the ligand, the complexity of the NMR spectrum increases as evidenced by doubling the number of the DNA resonances and appearance of a new set of resonances, indicating ligand complexation. As the drug:DNA ratio is raised, the intensities of the new resonances, bound DNA, increase (see peaks between 6.4–6.8 ppm) and the resonances from the free DNA decrease, until at drug–DNA ratio 2:1 only the resonances from the fully saturated complex are observed. Clearly one set of new resonances are maintained throughout the entire titration corresponding to one bound species, which is the 2:1 ligand–DNA complex. Even at low ligand:DNA stoichiometries such as 0.25:1, no intermediate 1:1 complex resonances were observed. The absence of any resonances of the 1:1 complex and the presence of only the 2:1 complex, from the beginning of the titration, demonstrates that two ligands bind to the duplex $\text{d}(\text{CGTATATAGC})_2$ with high co-operativity. The observation of the three thymine methyl peaks (1.4–1.7 ppm) in the NMR spectrum of the 2:1 complex ensures that ligands bind the DNA duplex symmetrically and the duplex keeps the dyad symmetry of the self-complementary sequence.

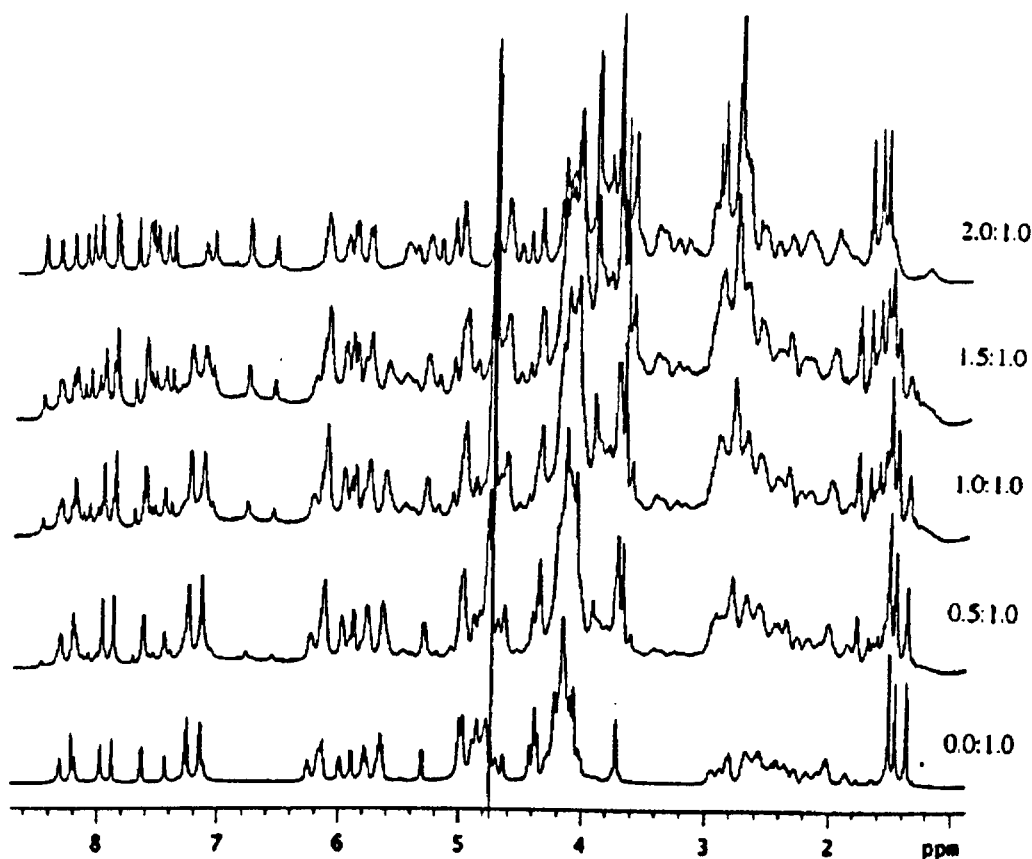


Figure 3.1 1D ¹H-NMR titration spectra of d(CGTATATAGC)₂ with tetra(N-methylpyrrole-2-carboxamide) carried out with 0.0, 0.5, 1.0, 1.5, 2.0 molar equivalents ligand added (1).

3.3.2 Circular dichroism titration

Circular dichroism (CD) can serve as a useful analytical tool for detecting co-operative binding in ligand–DNA recognition. CD spectral measurements have been used by Chen *et al* (2) to evidence co-operative binding of distamycin in 2:1 complex formation. As mentioned earlier, distamycin can form a 1:1 complex with duplexes containing binding sites with four base pairs, such as ATTA, and co-operative 2:1 complexes when binding sites contain five base-pairs such as AAGTT. On the basis of this observation, CD spectral titrations and comparison of spectral characteristics using these two types of oligonucleotides were made to provide evidence of differentiation of the two binding modes of distamycin. CD spectral measurements

were found to be successful since the dimer complex has a CD spectrum distinct from that of the monomer binding mode. This is due to the different extents to which the DNA duplex is distorted by the expansion of the minor groove.

Specifically, Chen *et al* (2) used oligomers containing ATTA and AAGTT binding sites to probe the possible spectral differences of the 1:1 and 2:1 binding modes. Distamycin is optically inactive when is free in solution, however in the presence of DNA, a CD spectrum can be induced upon complex formation. Figure 3.2 shows CD spectra for d(GCGATTAGCG) (A) and d(GCGAAGTTGCG) (B) in the absence and in the presence of increasing amounts of distamycin. It is obvious that CD spectral characteristics of the two binding modes are distinctly different. The CD spectral changes of these two binding modes can be seen clearly from the corresponding difference spectra having the DNA spectral contribution subtracted, in (C) and (D) respectively.

Binding of distamycin to the ATTA containing duplex results in a CD intensity enhancement of a band centring around 325 nm, some small changes near 250 and 230 nm are also observed (C). The rest of the spectrum remains almost the same. The maintenance of an isoelliptic point near 240 nm suggests a two-state process, with the distamycin forming a 1:1 complex. Distamycin binding to the AAGTT containing duplex induces two large positive bands near 265 nm and 330 nm and some small negative band centring near 290 nm (D). The isoelliptic point 241 nm appears again as with the ATTA containing duplex, however additional isoelliptic points at 284 nm and 306 nm emphasise again a two-state process. This time, the two-state process signifies the highly co-operative binding of distamycin molecules forming simultaneously the 2:1 complex. In summary, this CD study emphasises the distinctly different spectral characteristics of the binding of

distamycin in a 1:1 and co-operative 2:1 complex and illustrates the usefulness of CD titration as a sensitive method of detecting co-operativity in ligand–DNA recognition.

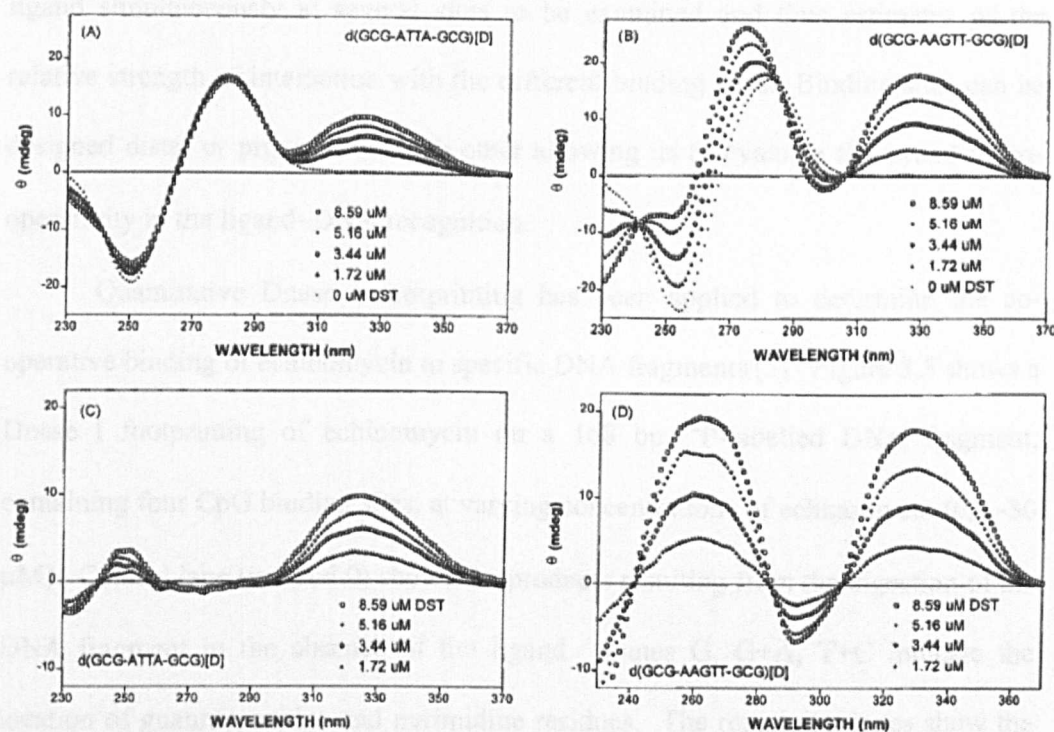


Figure 3.2 CD spectra titration of distamycin with d(GCGATTAGCG) (A) and d(GCGAAGTTGCG) (B). The corresponding ligand induced spectra, having the DNA spectral contributions subtracted, are shown in (C) and (D) (2).

3.3.3 Footprinting analysis

Footprinting refers to an assay in which the binding of a ligand inhibits the cleavage of the backbone of a nucleic acid by an enzymatic or chemical nuclease at specific sequences (3). The reaction produces radioactively labelled DNA fragments of varying length, which are separated by gel electrophoresis. The decrease in the intensities of certain bands identifies the position of the bound ligand, resulting in a drug footprint. Quantitative Dnase I footprinting was first used to estimate binding affinity and co-operativity with DNA-binding proteins (4). Later an analogous

procedure was developed to determine the binding affinity of low molecular weight drugs to specific nucleotide sequences (5). The advantage of this method is the use of longer DNA fragments than oligomers used in NMR and the potential of having multiple binding sites within a DNA fragment. In addition, it allows binding of a ligand simultaneously at several sites to be examined and thus estimates of the relative strength of interaction with the different binding sites. Binding sites can be designed distal or proximal of each other allowing us to evaluate the extent of cooperativity in the ligand–DNA recognition.

Quantitative Dnase I footprinting has been applied to determine the co-operative binding of echinomycin to specific DNA fragments (5). Figure 3.3 shows a Dnase I footprinting of echinomycin on a 168 bp ^{32}P -labelled DNA fragment, containing four CpG binding sites, at varying concentrations of echinomycin (0.1–30 μM). Control lane (marked 0) shows the products resulting from the digestion of the DNA fragment in the absence of the ligand. Lanes G, G+A, T+C indicate the location of guanine, purine and pyrimidine residues. The remaining lanes show the products of Dnase I digestion of the DNA fragment with the presence of increasing concentrations of echinomycin. Quantitative analysis of the concentration dependence of the footprints produces binding isotherms for each of the four CpG binding sites, present in the DNA fragment. Thus, the binding affinity for each binding site can be calculated. Co-operative binding can be detected when binding of the ligand to a specific site increases the binding affinity of the second ligand to an adjacent site.

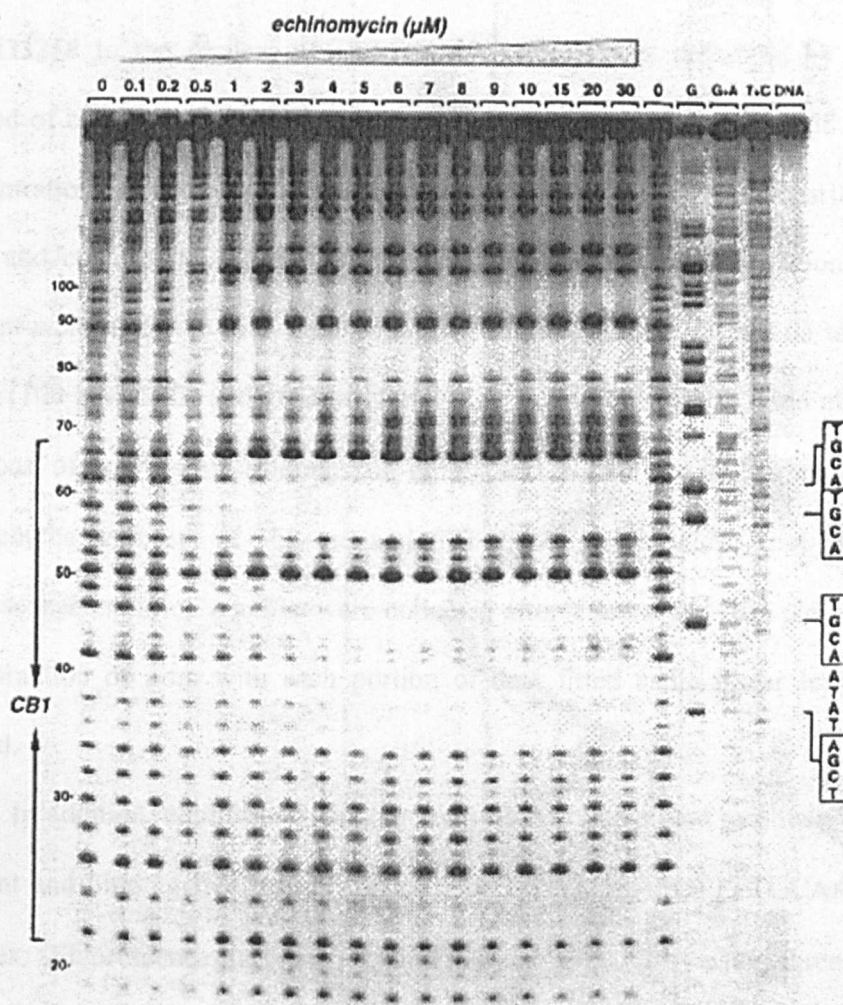


Figure 3.3 DNase I footprinting of echinomycin on the 168bp DNA fragment containing four CpG sites. The sequences on the right show the exact location of the ACGT and TCGA sites to which echinomycin binds (5).

3.4 Materials and methods

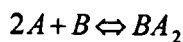
3.4.1 Fluorescence spectroscopy

For continuous variation and equilibrium binding analysis, stock solutions of the oligonucleotide at 500 μM and 50 μM , and stock solutions of H33258 at 300 μM , 30 μM and 3 μM were prepared. Solutions were prepared at pH 7.0 using a buffer solution of 100 mM NaCl and 10 mM NaH_2PO_4 . Fluorescence studies were performed at a temperature of 20 $^\circ\text{C}$ in 1cm path length polymethacrylate cuvettes,

using a Perkin Elmer luminescence spectrometer LS 50 B. Binding stoichiometry of the H33258 to the duplex d(CTTTTGCAAATG)₂ was measured by using the method of continuous variation analysis, described by Job (6). H33258 and DNA concentrations were varied but the sum concentration ($[DNA] + [Hoechst]$) was kept fixed and constant at 1 μ M. Different volumes of equimolar solutions of each reactant were mixed to give a mole fraction of drug ranging from 0.05 to 1 μ M in buffer (100 mM NaCl and 10 mM NaH₂PO₄). Final volume was fixed at 3ml with additions of appropriate volumes of buffer and H₂O. Excitation and emission wavelengths were set at 362 nm and 503 nm with slit width kept at 2.5 nm. Fluorescence emission maxima were collected after 4 scans and data plotted against mole fraction of drug with each portion of data fitted using linear least squares method.

In addition, equilibrium binding analysis was performed to obtain a binding constant and binding free energy for the Hoechst 33258–d(CTTTTGCAAATG)₂ complex. Fluorescence titrations were carried out at 20 °C by using three different (0.2, 0.5, 1.0 μ M) but fixed concentrations of Hoechst and varying the duplex concentration between 1nM and 1 μ M. The solution after each titration was stirred for 3 minutes in 1cm path length cuvette to establish homogeneity and equilibrium. Fluorescence emission maxima were collected and plotted against DNA concentration. A binding constant was obtained by fitting the data with the fitting equation 3.4 that is derived from the 2:1 binding mode of H33258 to the duplex.

In order to derive equation 3.4, we consider the 2:1 binding of the two drug molecules as a one step interaction; this is supported by observations from ¹H-NMR and fluorescence titrations that drug binding takes place in a highly co-operative mode. Thus, the 2:1 binding model is described by the following equilibrium:



where A is drug, B is DNA duplex and BA_2 is the doubly bound complex.

The equilibrium constant K_b is expressed as:

$$K_b = \frac{[BA_2]}{[B][A]^2} \quad \text{equation 3.1}$$

Also total concentrations of drug $[A]_{tot}$ and DNA duplex $[B]_{tot}$ are derived by the following equations:

$$[A]_{tot} = [A] + 2[BA_2] \Rightarrow [A] = [A]_{tot} - 2[BA_2] \quad \text{equation 3.2}$$

$$[B]_{tot} = [B] + [BA_2] \Rightarrow [B] = [B]_{tot} - [BA_2] \quad \text{equation 3.3}$$

The equation 3.4 is derived by substituting equations 3.2 and 3.3 into 3.1 and solving that equation for $[B]_{tot}$

$$[B]_{tot} = \frac{[BA_2]}{K_b([A]_{tot} - 2[BA_2])^2} + [BA_2] \quad \text{equation 3.4}$$

3.4.2 CD experiments

Circular dichroic (CD) spectra were measured at room temperature with an AVIV model 62 DS spectropolarimeter (Aviv Associates, Lakewood, NJ), using a 0.2 cm path length. CD spectral titrations were made by starting with oligonucleotide concentrations of 80 μ M in nucleotide followed by incremental additions of aliquots of Hoechst 33258 stock solution (1 mM). The oligomers and H33258 were prepared from stock solutions in 100 mM NaCl, 1 mM NaH_2PO_4 and 0.1 mM EDTA at pH=7.0 After each injection, the CD spectrum was recorded 5 times over the wavelength range 220–400 nm in 1.0 nm steps using a bandwidth of 4 nm, at 298 K. The CD data were baseline corrected by solvent subtraction, averaged and smoothly fitted.

3.4.3 NMR experiments

Stock solutions of the oligonucleotides and H33258 used in the following NMR experiments were synthesised, purified and prepared with exactly the same procedures as described in section 2.2.1. NMR titrations were carried out until the stoichiometry reaches the ratio 2:1 ligand:DNA. NMR data were collected and processed as described in section 2.2.3.

3.4.4 Molecular dynamics

All simulations were performed with the same parameters as described in section 2.2.5 using the AMBER 5.1 and AMBER 6 suites of programs. Starting structures for the free DNA and 2:1 complexes were taken from NMR restrained molecular dynamics simulation (chapter 2). Starting structure of the 1:1 complex was taken from the NMR structure of the 2:1 complex subtracting one drug molecule from the initial structure. The systems were electrically neutralised by addition of sodium counterions and immersed in a periodic box of around 40 x 40 x 60 Å water molecules, optimised, thermalised and equilibrated using our standard multistage protocol (section 2.2.5). The final equilibrated structures were then used to initiate three 5 ns unrestrained MD simulations at constant pressure ($P=1$ atm) and temperature ($T=298$ K). Shake was used to constrain all bonds, permitting a 2 fs time step for integration of Newton equations. Snapshots of the simulations collected every ps and the structures were determined to equilibrate by RMSD analysis. Helicoidal parameters were analysed with CURVES 5.1 and average structures and RMSDs were calculated over the equilibrated trajectories with CARNAL module of AMBER. Energy analysis was achieved using the MD implementation of the GB/SA method, configurational entropies and Molecular

Interaction Potential (MIP) calculations were calculated as described in the relevant publication (7).

3.5 Results and discussion

3.5.1 Fluorescence studies

The method of continuous variation analysis was used to obtain a Job plot for binding of H33258 with $d(\text{CTTTTGCAAAATG})_2$ at fixed H33258–DNA concentrations of $1\mu\text{M}$. Binding studies have used this method and successfully determined a single binding site (6) with different oligonucleotides, but have also shown multiple or degenerate binding stoichiometries with $\text{poly}(\text{dA–dT})_2$, $\text{poly}(\text{dG–dC})_2$ and calf thymus DNA (8). Figure 3.4 shows an intersection for the least squared fitted lines at 0.65 mole fraction of the drug, indicating a 2:1 stoichiometry that is consistent with the NMR study of the complex.

Binding constant and free energy was determined for the interaction of H33258 with the duplex $d(\text{CTTTTGCAAAATG})_2$. Fluorescence titrations were carried out for three different but fixed concentrations of H33258 at 20 °C and pH 7.0, figure 3.5. Binding constants were determined to be independent of the concentration in the range of $(0.2\text{--}1)\mu\text{M}$. Non-linear least-squared fitting determined a binding constant of $K_b = 0.4 \times 10^{15} \text{ M}^{-2}$ and a binding free energy of $\Delta G = -81.9 \text{ kJ mol}^{-1}$ at 20 °C. This is the first time that binding affinity of H33258 for a two binding site duplex has been reported and is in line with expectations. Fluorescence titrations have determined previously the binding constant of H33258 with a A_3T_3 (7) and a A_2T_2 (9) site in the range of $3.1\text{--}3.6 \times 10^8 \text{ M}^{-1}$. However, footprinting studies (10) of H33258 with a series of binding sites concluded that A_2T_2 and a A_3T_3 sites are unusually good binding sites and enable much tighter

binding than single TTTT or AAAA sites. It is important to note that these thermodynamic values are referred to the overall 2:1 interaction of Hoechst with the duplex. There is no way to determine the binding constant of the 1:1 interaction by fluorescence, since the process of 2:1 interaction is highly co-operative and fluorescence titrations can not separate the two binding sites.

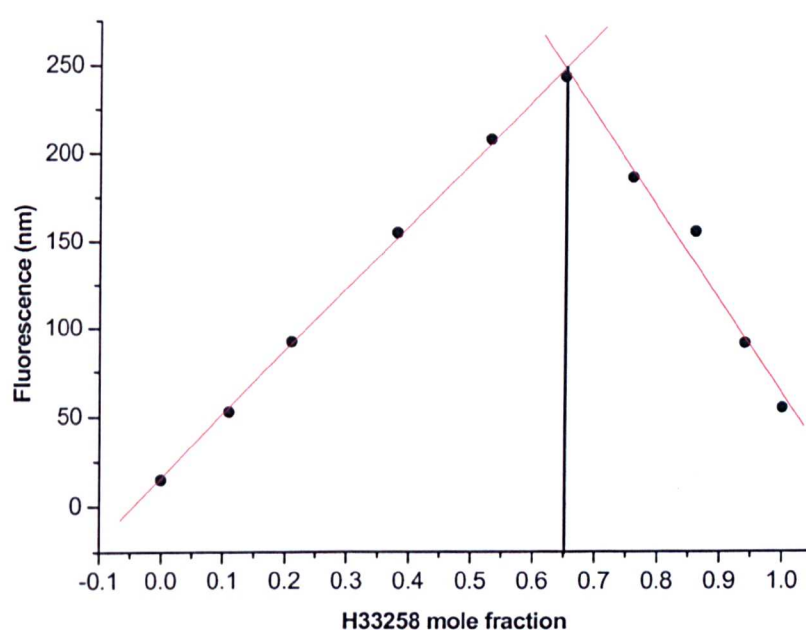


Figure 3.4 Job plot of Hoechst 33258 with d(CTTTTGCAAAAG)₂.

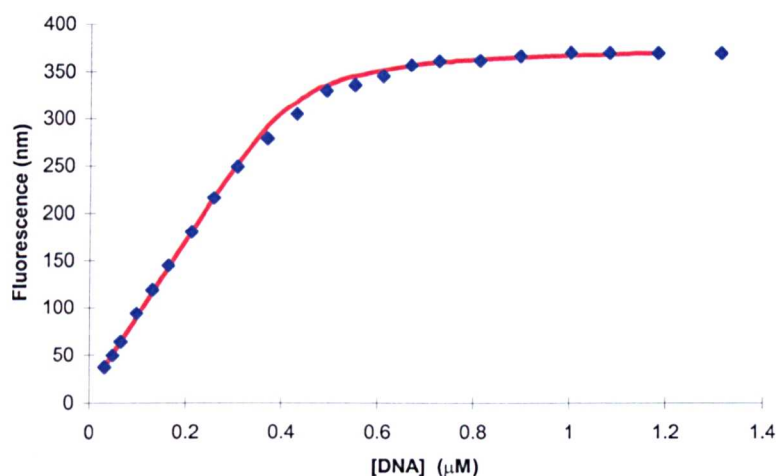


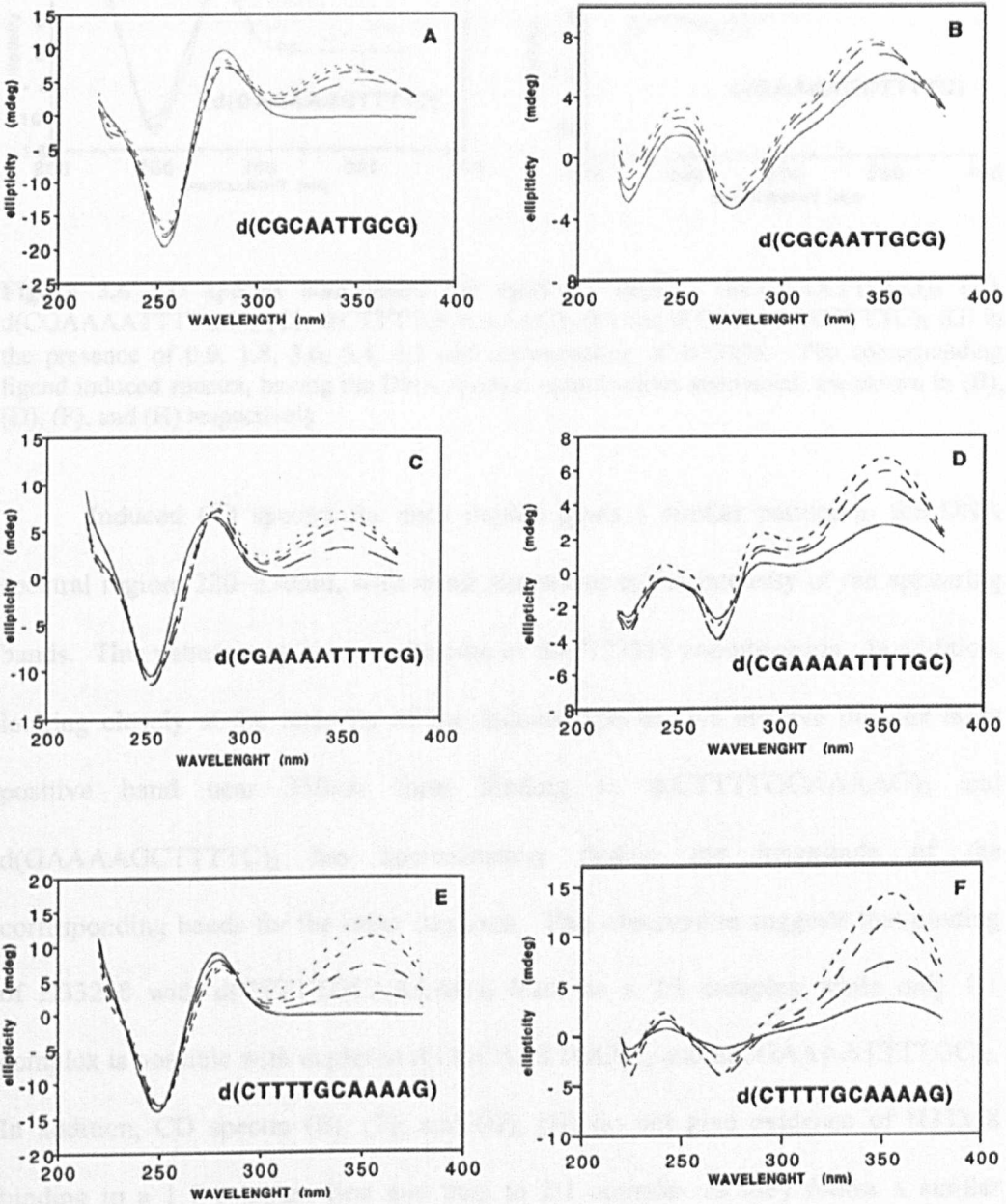
Figure 3.5 Equilibrium binding curve of H33258 with d(CTTTTGCAAAAG)₂.

3.5.2 CD analysis

CD spectral measurements successfully determined the detection of the co-operative 2:1 distamycin-binding mode. Therefore, we carried out an analogous study to probe the possible CD spectral characteristics of the 1:1, the 2:1 and also the co-operative 2:1 binding interaction of H33258. We have examined the interaction of H33258 to duplexes where only one H33258 molecule can bind such as d(CGCAATTGCG)₂ and where two H33258 molecules can bind such as d(CGAAAATTTTGC)₂, d(CTTTTGCAAAAG)₂, and d(GAAAAGCTTTTC)₂. The binding interaction of H33258 with the AATT site has been shown by NMR to lead to only a 1:1 complex (11). H33258 binding to the AAAATTTT site has been addressed by NMR and fluorescence titrations (12) and gives a single 1:1 complex, however, it is likely that two H33258 molecules can be accommodated to the two adjacent AAAA and TTTT binding sites. CD spectra titrations for d(CGCAATTGCG)₂ (A), d(CGAAAATTTTGC)₂ (C), d(CTTTTGCAAAAG)₂ (E), and d(GAAAAGCTTTTC)₂ (G) in the absence and presence of various amounts of

H33258 are shown in figure 3.6. Figure 3.6 shows also the induced CD spectra where the DNA contributions are subtracted in (B), (D), (F), and (H) respectively.

It is obvious that titration of H33258 to each oligonucleotide can induce a sizeable CD spectrum, showing complex formation. Apparently, binding of H33258 to each duplex induces a spectrum with different characteristics. Titration to the AATT containing duplex, (A) and (B), results in a CD enhancement of a band centering around 340 nm, with little spectral alterations except for some changes near 275 nm, 250 nm and 225 nm. Binding of H33258 to the AAAATTTT containing duplex, (C) and (D), induces a spectrum with a large positive band near 350nm, and also some small bands around 300 nm, 270 nm, 245 nm and 225 nm. In addition, (E) and (F) panels show that H33258 titration to the d(CTTTTGCAAAAG)₂ induces a CD spectrum with a large positive band near 350 nm and smaller changes at 270 nm, 245 nm and 225 nm. Finally, binding interaction to the d(GAAAAGCTTTTC)₂, (G) and (H), induces a spectrum with a large positive band near 350 nm and a smaller band around 280 nm. Isoelliptic points for the four CD titration spectra are not defined clearly probably due to instrumentation instability. However, there is evidence for isoelliptic point for the CD spectra of d(CGAAAATTTTGC)₂ (D), d(CTTTTGCAAAAG)₂ (F) and d(GAAAAGCTTTTC)₂ at 280 nm, 290 nm, and 260nm respectively. The CD spectrum of d(CGCAATTGCG)₂ (B), should also have an isoelliptic point around 290 nm, but this is not clear, largely due to small differences in intensity in the region 220–300 nm. Isoelliptic points provide strong evidence that binding of H33258 to the three duplexes is a two-state process. This specifically indicates that binding of H33258 to the two binding sites of the d(CTTTTGCAAAAG)₂ and d(GAAAAGCTTTTC)₂ appears to be a highly co-operative process.



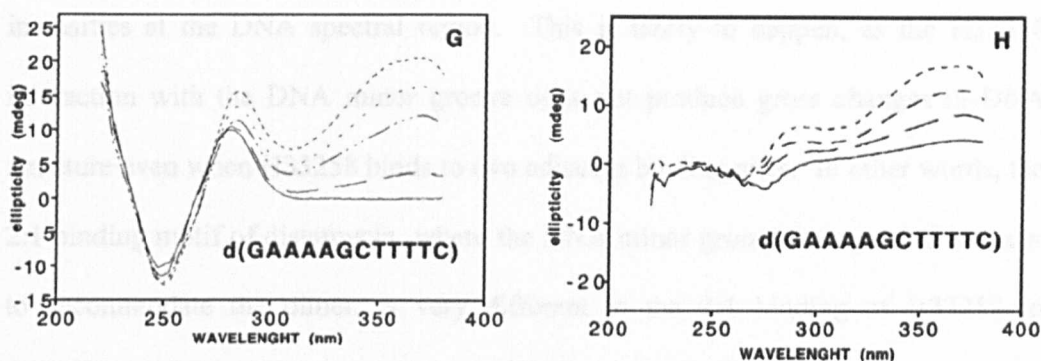


Figure 3.6 CD spectra comparison for 4 μ M (in duplex) d(CGCAATTGCG)₂ (A), d(CGAAAATTTTGC)₂ (C), d(CTTTTGCAAAAG)₂ (E) and d(GAAAAGCTTTTC)₂ (G) in the presence of 0.0, 1.8, 3.6, 5.4, 8.2 μ M concentration of H33258. The corresponding ligand induced spectra, having the DNA spectral contributions subtracted, are shown in (B), (D), (F), and (H) respectively.

Induced CD spectra for each duplex gives a similar pattern in the DNA spectral region, 220–330 nm, with small alterations in the intensity of the appearing bands. This pattern may be characteristic of the H33258 complexation. In addition, looking closely at the intensity of the induced spectra, we observe that the large positive band near 350 nm from binding to d(CTTTTGCAAAAG)₂ and d(GAAAAGCTTTTC)₂ has approximately double the magnitude of the corresponding bands for the other duplexes. This observation suggests that binding of H33258 with d(CTTTTGCAAAAG)₂ leads to a 2:1 complex while only 1:1 complex is possible with duplexes d(CGCAATTGCG)₂ and d(CGAAAATTTTGC)₂. In addition, CD spectra (E), (F), and (G), (H) do not give evidence of H33258 binding to a 1:1 complex first and then to 2:1 complex as they follow a similar pattern with the other spectra, suggesting again the presence of co-operative binding.

In summary, these results have provided CD spectral characteristics of the H33258–DNA recognition with different duplexes containing one and two binding sites. CD spectral titration provided some evidence of differentiation when H33258 binds to a single site or to two sites, however, induced CD spectra have similar band

intensities at the DNA spectral region. This is likely to happen, as the H33258 interaction with the DNA minor groove does not produce gross changes in DNA structure even when H33258 binds to two adjacent binding sites. In other words, the 2:1 binding motif of distamycin, where the DNA minor groove is expanded in order to accommodate the dimer, is very different to the 2:1 binding of H33258 to d(CTTTTGCAAAG)₂, where minor differences in DNA structure are observed (13). In addition, isoelliptic points in the CD spectrum provide us with evidence to suggest co-operative binding of H33258 with d(CTTTTGCAAAG)₂ and d(GAAAAGCTTTTC)₂.

3.5.3 NMR binding studies of H33258

In chapter 2, we described NMR titration studies of the binding of H33258 to the dodecamer duplex d(CTTTTGCAAAG)₂. NMR titration, leads us to conclude that two H33258 molecules bind co-operatively to d(CTTTTGCAAAG)₂ and further NMR-MD analysis suggested that the duplex has time-averaged sequence-dependent features that account for the bound orientation of the two drug molecules. On the basis of these results, we aimed to investigate the effect of the distance between the binding sites on the co-operative binding of H33258. Our initial oligonucleotide contains two TTTT binding sites separated by an intervening GC base step, which isolates the two ligands by approximately 15 Å (13). We have now examined binding of H33258 to d(CTTTTGGCCAAAAG)₂ in which the central GC base step have been extended to GGCC base step and the binding sites has been further separated. This duplex retains the TpG (CpA) junctions at the end of the binding site, which appear to restrict the orientation of H33258. The duplex should have analogous sequence-dependent features to those observed for the dodecamer

d(CTTTTGCAAAAG)₂, however the distance between the sites is increased and this may change the degree of co-operative binding if additional steps play a role in transmitting the “information” between the binding sites.

Titration of d(CTTTTGGCCAAAAG)₂ with H33258 was carried out at 25 °C in 0.25 molar equivalents per ligand addition and followed by ¹H-NMR. Figure 3.7 shows the 1D NMR spectra of the free duplex and at 1:1 and 2:1 ratios of complex. Upon addition of the ligand, the NMR spectrum increases in complexity as a new set of signals corresponding to bound DNA appear. Looking closely at the methyl region (1.3–1.8 ppm), it is evident, that the intensity of the methyl peak, corresponding to the four thymine methyl protons of the free DNA, decrease while only four new resonances increase their intensity over the range 0.25:1 to 2:1 drug:DNA ratio. Resonances of the complex formed can also be observed in previously unoccupied spectral regions (6.5–6.0 ppm).

Clearly, the same set of new resonances persisted up to a ligand to DNA ratio of 2:1 and a single 2:1 H33258–d(CTTTTGGCCAAAAG)₂ complex was formed through the entire titration as there is no evidence for resonances from an intermediate 1:1 complex, even at low drug–DNA stoichiometries such as 0.25:1. The fact that only one set of DNA resonances are observed with the 2:1 complex are also confirmed by 2D NMR, figure 3.8. This indicates that the self-complementary duplex keeps the C₂ symmetry in the complex and H33258 molecules bind again in

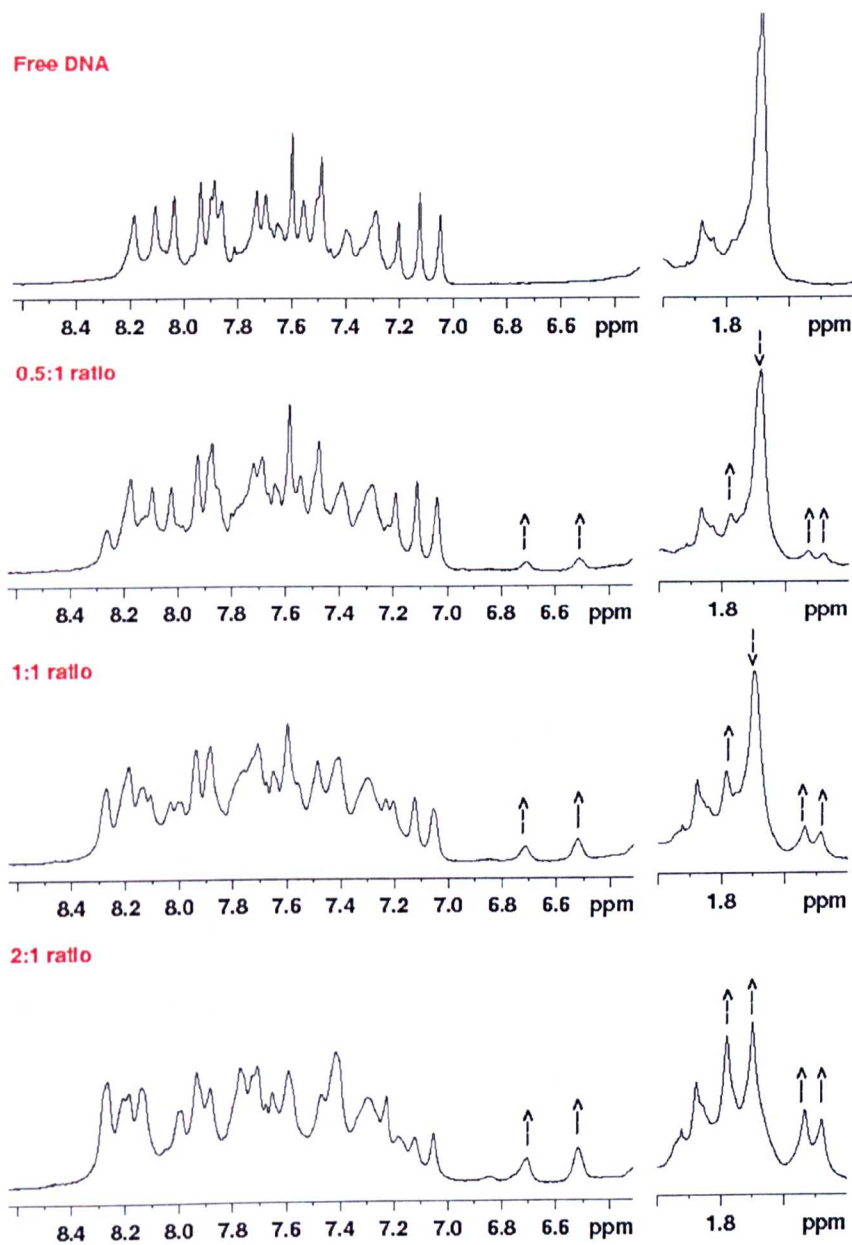


Figure 3.7 1D ^1H -NMR titration spectra of $d(\text{CTTTTGGCCAAAAG})_2$ with H33258 showing 0.0, 0.5, 1.0, 2.0 molar equivalents ligand added.

symmetrically-related orientations. To summarise, the NMR titration follows an analogous trend with the titration of the dodecamer $d(\text{CTTTTGCAAAG})_2$ and shows that even when the intervening GC base step has been extended to GGCC, the binding of H33258 at the two sites appears to be co-operative with a similar degree of co-operativity.

We were surprised to observe again a high degree of co-operative binding with the duplex $d(\text{CTTTTGGCCAAAAG})_2$ as we expected that the increased distance between sites would at least change the degree of co-operativity and it would be possible able to detect any 1:1 complex. The resonances of the 2:1 complex have been assigned by a combination of 2D NMR techniques DQF-COSY, TOCSY and NOESY. The pattern of the NOESY spectra (figure 3.8) and the volumes of the cross peaks, after integration, of the 2:1 complex with the duplex $d(\text{CTTTTGGCCAAAAG})_2$ are very similar with the NOESY spectra of the respective dodecamer duplex. The same applies for the number and the volume of the intermolecular NOEs, indicating that H33258 molecules again are bound within the TTTT tracts with the positively charged piperazine rings facing each other towards the central GGCC steps.

The structure and dynamics of the central steps GC or GGCC of both the duplexes should be expected to be very similar. An interesting feature of this type of purine-pyrimidine intervening sequence predicted from extensive computational studies (14), is that it is expected to have minimal motion and low flexibility. We have suggested earlier that H33258 binding does not induce significant changes in the structure of the dodecamer duplex; NMR analysis shows similar findings for the binding of H33258 with the duplex $d(\text{CTTTTGGCCAAAAG})_2$. This allows us for the first time to consider that overall flexibility of the DNA molecule can be

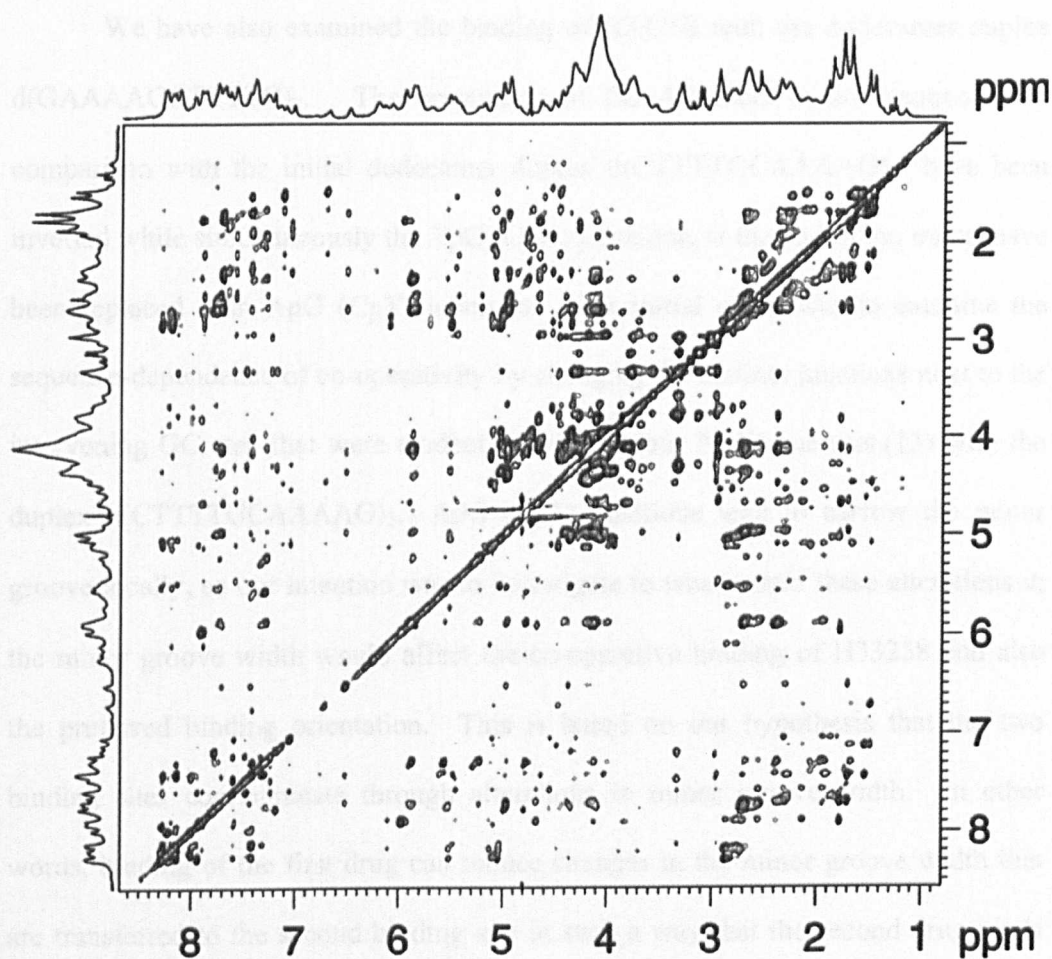


Figure 3.8 NOESY spectrum of the 2:1 H33258–d(CTTTTGGCCAAAAG)₂ complex in D₂O solution, recorded at 298 K with mixing time 250 ms.

affected by H33258 binding and this may be important for co-operative binding. Thus, we decided to carry out extended molecular dynamics simulations to study the free DNA, the 2:1 complex and also the theoretical 1:1 complex of the duplex d(CTTTTGCAAAAG)₂ (see below) and explore all the possible factors that could be responsible for the co-operativity.

We have also examined the binding of H33258 with the dodecamer duplex d(GAAAAGCTTTTC)₂. The orientation of the AT-tracts in this sequence, by comparison with the initial dodecamer duplex d(CTTTTGCAAAAG)₂ have been inverted while simultaneously the TpG (CpA) junctions, at the end of the tracts, have been replaced with ApG (CpT) junctions. Our initial effort was to examine the sequence-dependence of co-operativity by changing the distinct junctions next to the intervening GC step that were evident in our previous NMR analysis (13) with the duplex d(CTTTTGCAAAAG)₂. ApG (CpT) junctions tend to narrow the minor groove locally, so our intention was to investigate to what extent these alterations in the minor groove width would affect the co-operative binding of H33258 and also the preferred binding orientation. This is based on our hypothesis that the two binding sites communicate through alterations in minor groove width. In other words, binding of the first drug can induce changes in the minor groove width that are transferred to the second binding site in such a way that the second drug binds more favourably.

1D NMR titration spectra of H33258 with d(GAAAAGCTTTTC)₂, recorded at 25 °C in 0.25 molar equivalents per ligand addition, are shown in figure 3.9. NMR titration shows that the complexity of the spectrum increases upon ligand addition and at a ratio of drug to duplex of 2:1, we see a large number of broad new resonances. Specifically, observing the methyl region (1.5–1.8 ppm) and the

aromatic region (7.0–8.4 ppm), it is evident that peaks corresponding to the protons of the free DNA are decreased and surrounded gradually with additional resonances, while at the 2:1 ratio the total number of resonances can not account for a single species in solution. This is consistent with drugs bound in a number of different orientations, forming more than one 2:1 complex. Possible complexes arising from the orientational disorder of H33258 with d(GAAAAGCTTTTC)₂ are shown schematically in figure 3.10. The complexity of the spectrum does not permit us to resolve the mixture of the bound species in solution. 2D NOESY spectra show an excess of cross peaks in the aromatic region, consistent with a mixture of bound species, figure 3.11. The ApG junctions do not appear to result in distinct structural features capable of influencing H33258 minor groove recognition, in terms of its binding orientation. To conclude, this study adds no further information about co-operative binding between adjacent sites. Co-operative binding may still be present with the duplex d(GAAAAGCTTTTC)₂ as the CD spectral titrations suggested, however it is H33258 orientational disorder that makes spectral analysis complicated.

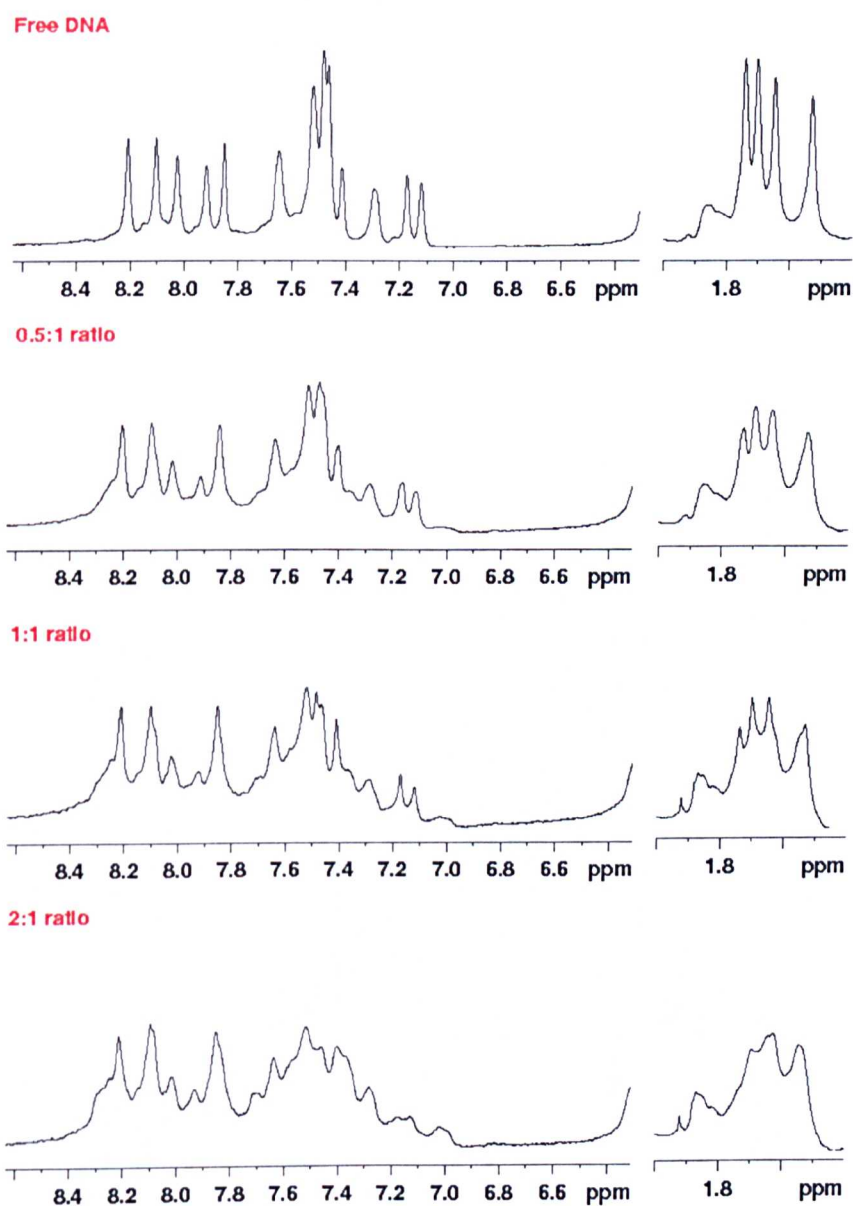


Figure 3.9 1D ^1H -NMR titration spectra of $\text{d}(\text{GAAAAGCTTTTC})_2$ with H33258 showing 0.0, 0.5, 1.0, 2.0 molar equivalents ligand added.

3.3.4 Footprinting analysis

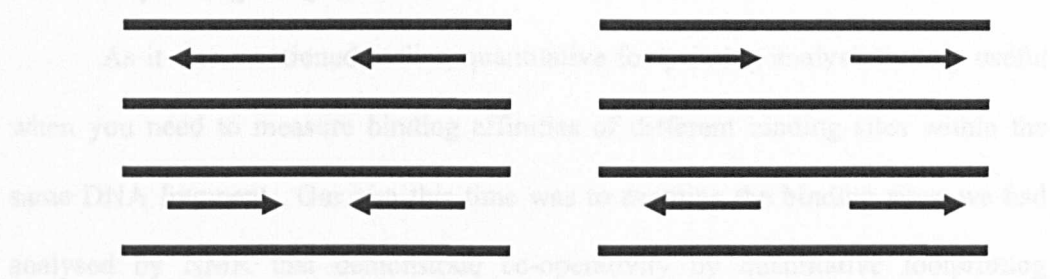


Figure 3.10 Representation of the possible types of complexes from the binding of two drug molecules along the duplex $d(GAAAAGCTTTTC)_2$. The head of the arrow represents the bulky N-methyl piperazine ring of H33258.

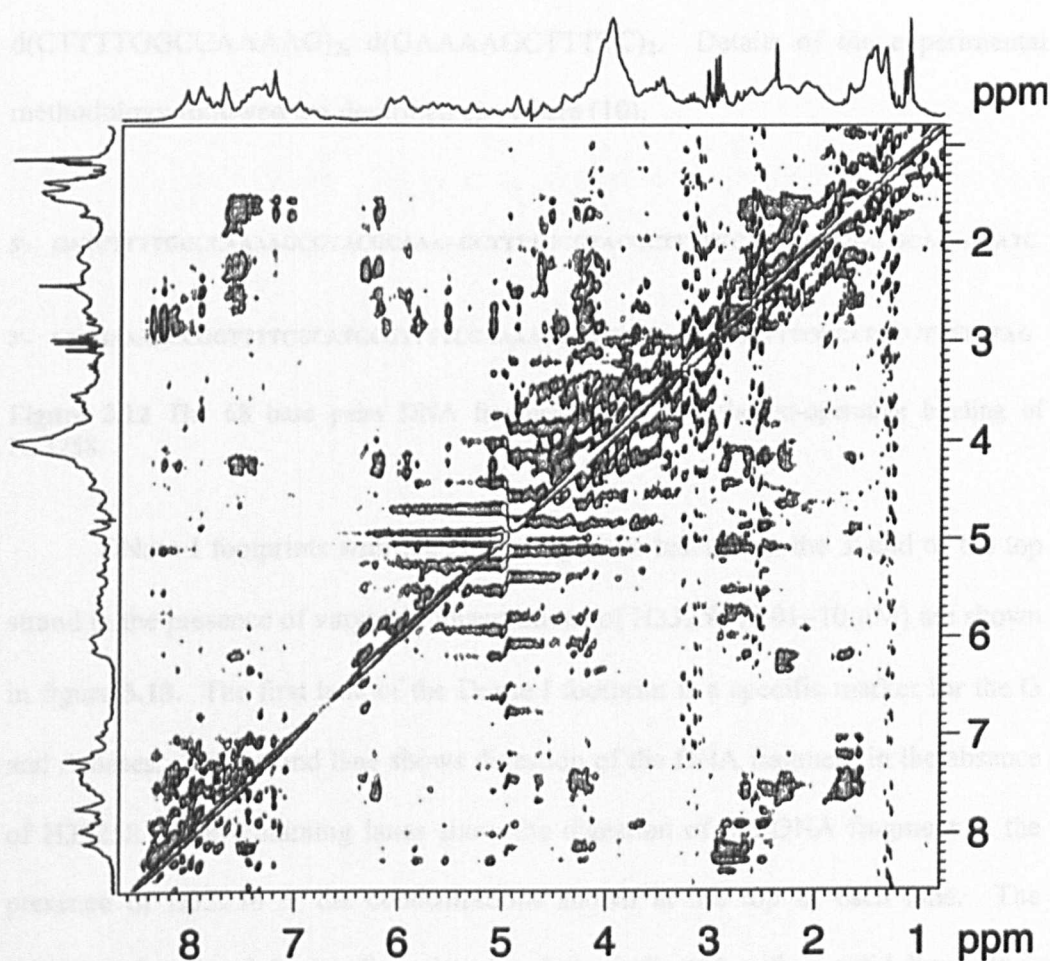


Figure 3.11 NOESY spectrum of the 2:1 H33258- $d(GAAAAGCTTTTC)_2$ complex in D_2O solution, recorded at 298 K with mixing time 250 ms.

3.5.4 Footprinting analysis

As it was mentioned earlier, quantitative footprinting analysis is very useful when you need to measure binding affinities of different binding sites within the same DNA fragment. Our aim this time was to examine the binding sites, we had analysed by NMR that demonstrate co-operativity by quantitative footprinting analysis and evaluate the degree of co-operativity between these sites. To this end, we collaborated with Professor Keith R. Fox, University of Southampton, and we designed a synthetic 68 base pair DNA fragment (figure 3.12) that contains the sequences we have so far analysed by NMR d(CTTTTGCAAAAG)₂, d(CTTTTGGCCAAAAG)₂, d(GAAAAGCTTTTC)₂. Details of the experimental methodology followed are described elsewhere (10).

5'- GATCTTTTGGCCAAAAGCGTACGCAAAAGCTTTTGGTACGCTTTTGCAAAAGCGTACGCAAAAGATC
3'- CTAGAAAACCGGTTTTTCGCATGCGTTTTCGAAAACGCATGCGAAAACGTTTTTCGCATGCGTTTTCTAG

Figure 3.12 The 68 base pairs DNA fragment used to probe co-operative binding of H33258.

DNase I footprints with the DNA fragments labelled at the 3'-end of the top strand in the presence of varying concentrations of H33258 (0.01–10 μ M) are shown in figure 3.13. The first lane of the DNase I footprint is a specific marker for the G and A bases. The second lane shows digestion of the DNA fragment in the absence of H33258. The remaining lanes show the digestion of the DNA fragment in the presence of H33258 at the concentrations shown at the top of each lane. The location of each of the binding sites has been indicated with a solid line. It is immediately apparent that DNase I footprints can be seen at each of the expected TTTT binding site, however, visual inspection of the footprints show a similar

degree of concentration dependence for all the binding sites. This is also confirmed by quantitative analysis of the concentration dependence of each footprint.

This is achieved by measuring the intensity of the bands within each of the footprinting sites using a microdensitometer. The bands are then normalised relative to a band in the cleavage pattern, which is not affected by the ligand, and these normalised values are plotted against ligand concentration. Footprint plots (figure 3.14) are then fitted by a simple binding curve (equation 3.5), which gives an indication of the binding affinity K .

$$I_c = I_o(K/(L+K)) \quad \text{equation 3.5}$$

I_c is the band intensity in the presence of the ligand, I_o is the band intensity in the control and L is the ligand concentration. The binding affinity for each binding site is indicated on the plots. Footprinting plots show a similar degree of H33258 binding affinity for each of the various binding sites, which are not significantly different to strongly indicate co-operativity.

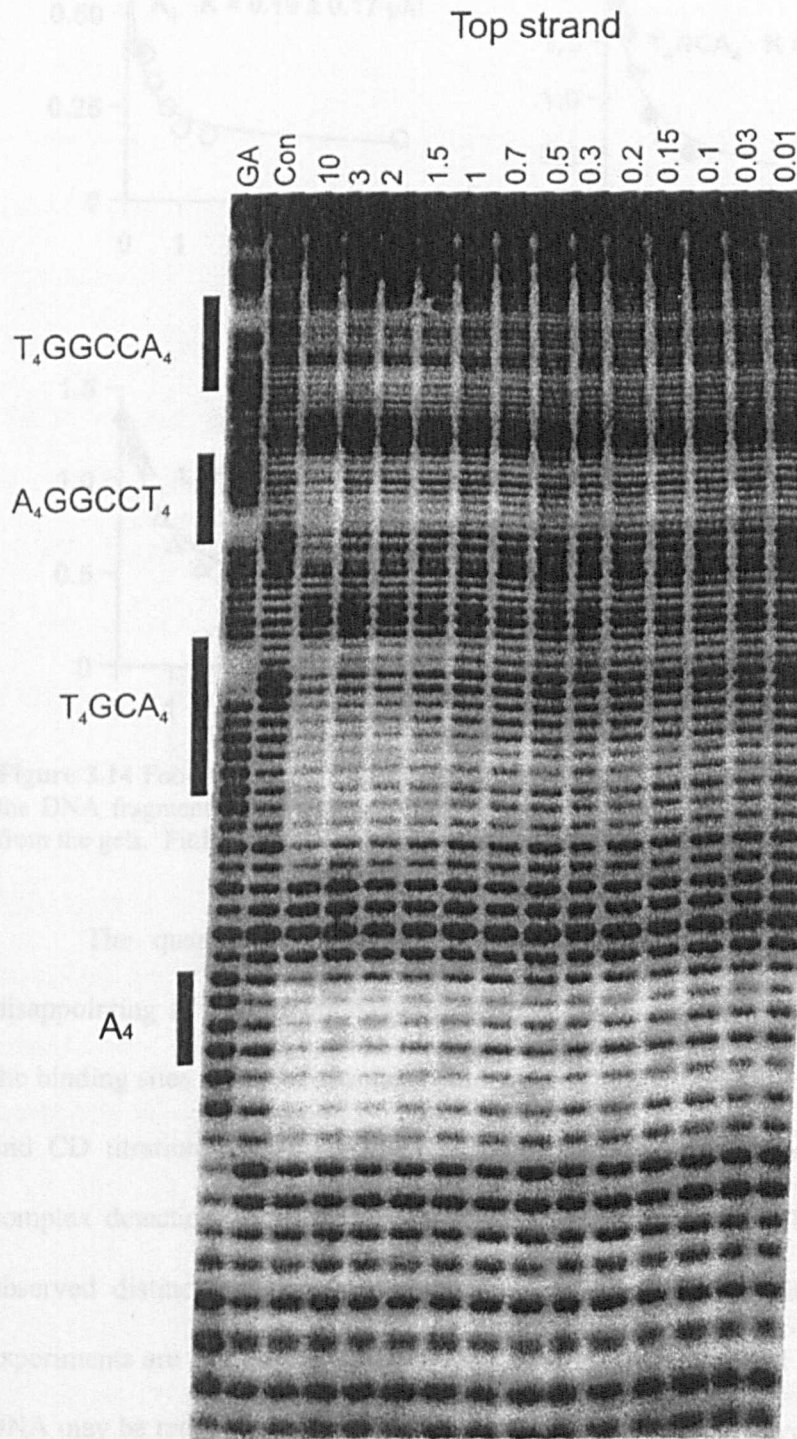


Figure 3.13 Dnase I footprinting of H33258 on the 68 base pairs DNA fragment at various concentrations. Numbers on the top of each lane refer to the concentration of H33258 (μM) and sequences on the left show the exact location of the binding sites to which H33258 binds.

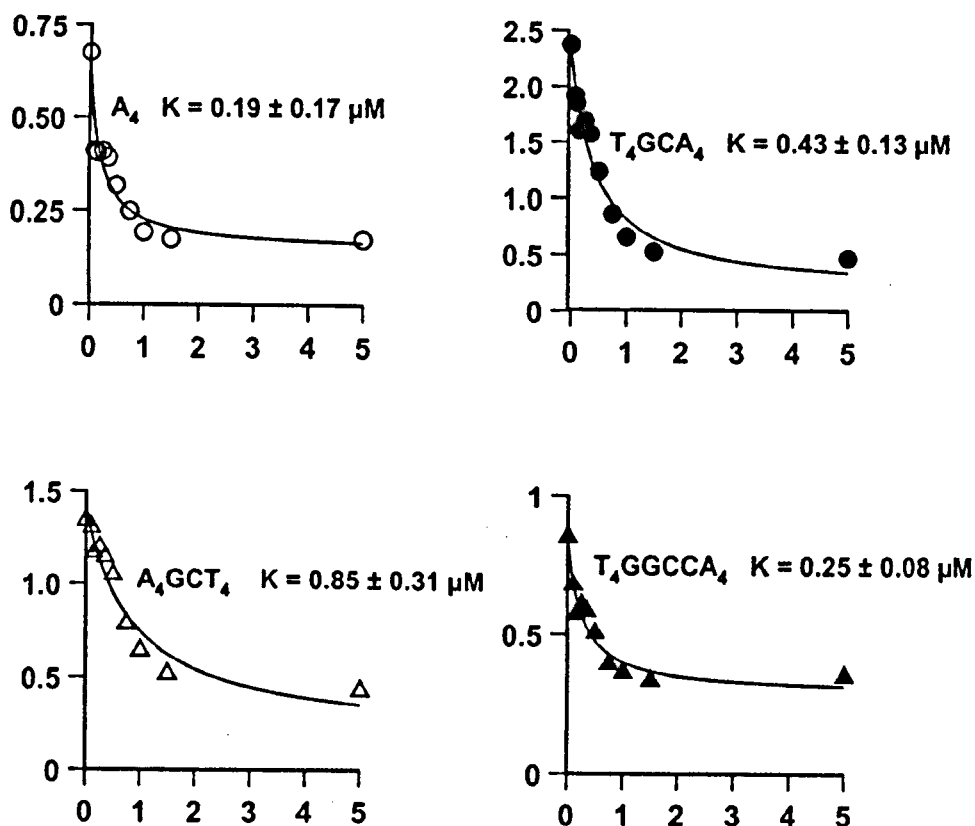


Figure 3.14 Footprinting plots showing the interaction of H33258 with various sites within the DNA fragment. Data points were obtained from measuring the intensity of the bands from the gels. Fitting data with equation 3.5 results to binding affinities shown at each plot.

The quantitative footprinting studies presented here have been rather disappointing as regards the demonstration of co-operative binding of H33258 with the binding sites that NMR studies demonstrated. This seems confusing, since NMR and CD titrations showed strong evidences of co-operative binding with no 1:1 complex detection. It is possible that co-operative binding of H33258 cannot be observed distinctly at the low concentrations at which quantitative footprinting experiments are carried out. In addition, co-operative binding of H33258 to duplex DNA may be reduced and disappear when the DNA molecule is very long compared to short oligonucleotides studied by NMR and CD spectroscopy here in these studies. If this is the case, then the origins of the co-operativity are affected somehow by the length of the sequence and this should be kept in mind.

3.5.5 Molecular dynamics analysis

Long molecular dynamics simulations of the free d(CTTTTGCAAAAG)₂, the 2:1 complex and the theoretical 1:1 complex were performed for 5 ns under the same conditions. All three simulations relaxed quickly from their initial conformations and remained stable over the 5 ns simulation periods, judging from RMSD analysis (not shown). Time-averaged structures were generated from the final 4 ns of each simulation. The average structures of the free DNA and 2:1 complex were found to be in excellent agreement with the average NMR-derived structures. Heavy-atom RMSD between the modelled and experimentally derived structures were 1.98 Å for the free DNA and 1.05 Å for the 2:1 complex. This was reduced to 1.65 Å and 1.03 Å respectively if terminal bases were excluded. Excellent agreement between experimental (NMR) and theoretically derived (MD simulation) helical parameters was also observed (figure 3.15; for comparison see figure 2.14 chapter 2). The MD simulations predict that the conformation of the DNA in the 1:1 complex does not differ greatly from that of the free DNA or the 2:1 complex. It shows the expected narrowing of the minor groove for the occupied A-tract, while the width of the unoccupied A-tract remains close to its free DNA value. This suggests that co-operativity does not relate to any dramatic conformational changes.

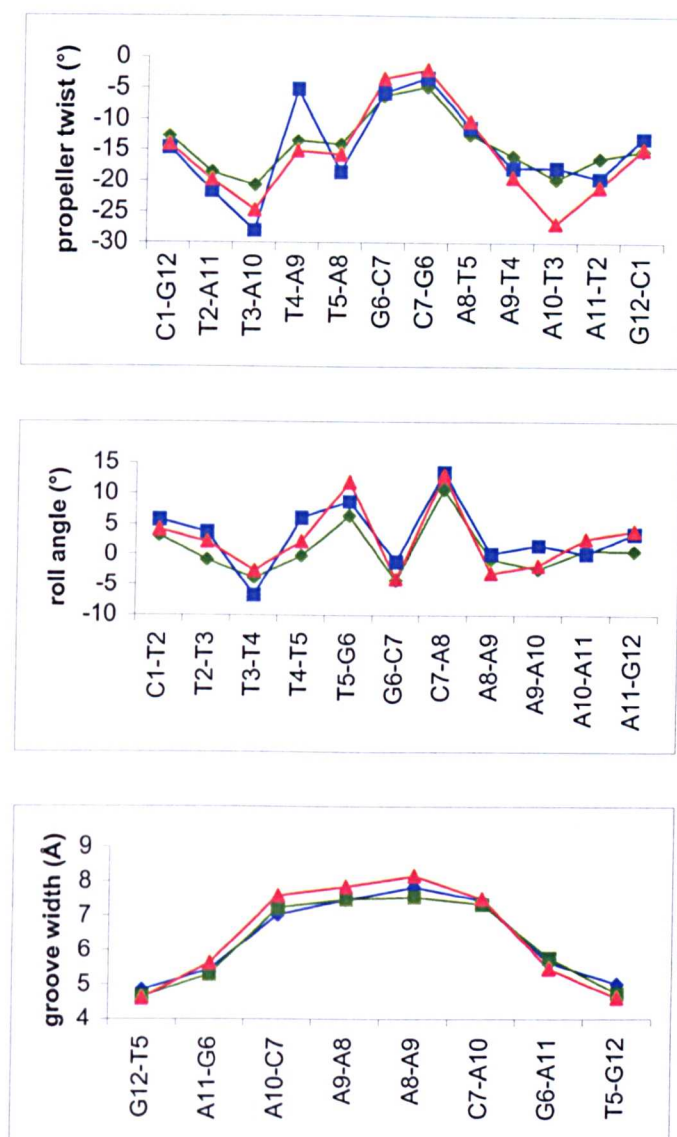
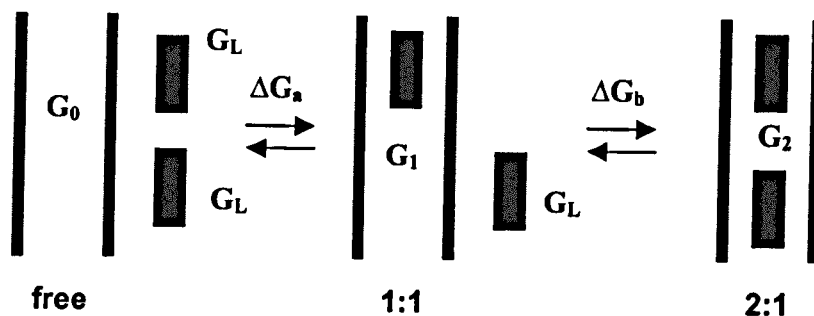


Figure 3.15 Propeller twist, roll and minor groove widths for the time-averaged structures of the free DNA (diamonds), 1:1 complex (squares) and 2:1 complex (triangles).

The nature of the system and the equilibria involved allows us to calculate the difference between the free energy changes associated with the first and second binding events ($\Delta\Delta G$) directly from our simulations without having to characterise the thermodynamics of the unbound drug (scheme 3.2). This work was carried out in

collaboration with Dr. Sarah Harris and Dr. Charlie Laughton from the School of Pharmaceutical Sciences, University of Nottingham.



$$\Delta G_a = (G_1 + G_L) - (G_0 + 2G_L)$$

$$\Delta G_b = G_2 - (G_1 + G_L)$$

$$\Delta\Delta G = \Delta G_b - \Delta G_a$$

$$= G_0 + G_2 - 2G_1$$

Scheme 3.2 Partitioning of the free energy terms. G_0 : free energy of the free DNA; G_L : free energy of the free ligand; G_1 : free energy of the 1:1 drug–DNA complex; G_2 : free energy of the 2:1 drug–DNA complex.

It is useful to expand G , the total free energy of the system (including water and counterions) as shown in equation 3.6:

$$G = E^{\text{intra}} + G^{\text{solv}} - TS^{\text{intra}} \quad \text{equation 3.6}$$

The first term, E^{intra} , is the internal energy of the solute (DNA or DNA–drug complex). The second term, G^{solv} , is the free energy of solvation of the solute and the third term, S^{intra} , is the configurational entropy of the solute. Using scheme 3.2 and equation 3.6, the free energy difference that constitutes co-operativity can be computed as shown in equation 3.7. Note that G^{solv} includes both enthalpic and entropic terms related to the reorganisation of the solvent (water and counterions).

$$\Delta\Delta G = (E^{\text{intra}} + G^{\text{solv}})_0 + (E^{\text{intra}} + G^{\text{solv}})_2 - (E^{\text{intra}} + G^{\text{solv}})_1 - T(S^{\text{intra}}_0 + S^{\text{intra}}_2 - 2 S^{\text{intra}}_1) \quad \text{equation. 3.7}$$

Trajectories were analysed using the GB/SA approach (see below) and the term $(E^{\text{intra}} + G^{\text{solv}})$ computed, though we cannot unambiguously separate out the two contributions. Then, the trajectories were analysed in terms of i) internal energy and solvation components $(E^{\text{intra}} + G^{\text{solv}})$, and ii) configurational entropy components (S^{intra}) . Our aim is to elucidate which one, or more, of these terms is responsible for the co-operativity observed in this system.

The internal energies, with solvation correction, of the free DNA, the 1:1 complex and the 2:1 complex in each snapshot from the equilibrated portions of the trajectories were calculated using the GB/SA method. This method is implemented in AMBER 6 and its validity was tested on nucleic acids molecular dynamics (16). The resulting estimates (table 3.1) gives $\Delta\Delta(E^{\text{intra}} + G^{\text{solv}}) = 3.3 \pm 0.4 \text{ kcal mol}^{-1}$. Thus, on the basis of enthalpy considerations alone (including a solvation correction), the interaction of H33258 with this DNA sequence is predicted to be somewhat anti-co-operative.

Configurational entropies of the free DNA, the 1:1 complex and the 2:1 complex were calculated from the dynamics data via Principal Component Analysis, using the method of Schlitter (17). Entropy values obtained in this way are sensitive to the simulation length, therefore S was calculated for various sample window widths, and S_∞ estimated by fitting the empirical equation 3.8:

$$S(t) = S_\infty - \frac{a}{t^{2/3}} \quad \text{equation 3.8}$$

Table 3.1 Thermodynamic parameters calculated from the MD simulations of the free DNA, 1:1 drug–DNA complex and 2:1 drug–DNA complex. All values are in kcal mol⁻¹ ± standard errors, for T=300K.

System	$E^{\text{int}}+G^{\text{solv}}$	$\Delta(E^{\text{int}}+G^{\text{solv}})$	TS_{∞} (system)	$T\Delta S_{\infty}$	TS_{∞} (DNA only)	$T\Delta S_{\infty}$
Free DNA	-4375.6 ± 0.2		830.6 ± 0.5		830.6 ± 0.5	
		29.0		-24.3		25.6
1:1 Complex	-4404.6 ± 0.2		854.9 ± 0.3		805.0 ± 0.5	
		25.7		-34.8		13.1
2:1 Complex	-4430.3 ± 0.2		889.7 ± 0.1		791.9 ± 0.5	

From the resulting values (table 3.1) $\Delta\Delta TS$ was calculated at 300K to be 10.4 ± 0.7 kcal mol⁻¹, figure 3.16. This implies that the binding of the first drug molecule to the DNA is associated with a considerably greater entropic penalty than the binding of the second. These calculations assume that changes in translational and rotational entropy can be ignored. By statistical mechanics, the absolute translational entropy of a molecule is dependent on its mass and its rotational entropy on the moments of inertia. Ligand binding will have a very small effect on these quantities. Indeed, quartz crystal microbalance experiments on related drug–DNA systems (18) indicate that ligand binding is associated with no change in the effective mass of the molecule, because an equivalent mass of tightly bound water is displaced from the minor groove in the process. Combined with the value of $\Delta\Delta(E^{\text{intra}}+G^{\text{solv}})$ previously obtained, $\Delta\Delta G$ for this system was calculated to be -7.1 ± 0.8 kcal mol⁻¹. The result is in good agreement with the NMR titration estimates and we predict that, in this case, co-operativity is the result of the balance of entropic factors, which over-ride the small intrinsically anti-co-operative nature of the enthalpic terms involved.

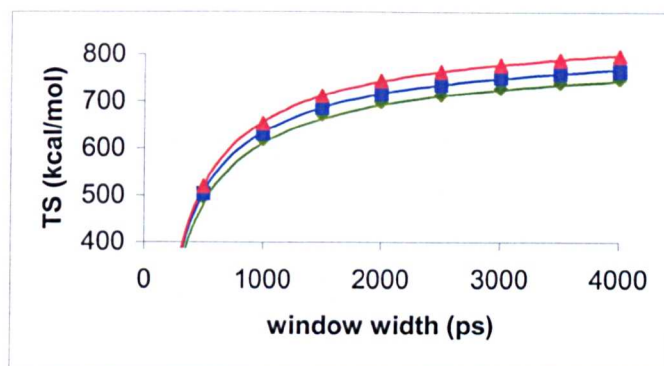


Figure 3.16 Effect of sample window width on the values of the configurational entropies, calculated by the method of Schlitter (17). The symbol coding is same as in figure 3.15. The points are the experimental values, the lines are the results of the least squares fit to the equation 3.8.

The GB/SA approach does not allow the energetics of the system to be reliably decomposed into individual intramolecular energy and solvation terms. However, we have obtained some qualitative insight into these through examination of molecular interaction potential (MIP) maps (19) and hydration density maps (20). Figure 3.17 shows the MIP map obtained for the time-averaged structures of the free DNA and the 1:1 complex, when a molecule of H33258 is used as the probe. For the free DNA, we see clear areas of density in both A-tracts. The density obtained for the unoccupied site in the 1:1 complex is very similar, thus, there is no obvious change in the affinity of the second site for H33258 once the first molecule of drug is bound in the other site. Figure 3.18 shows the MIP maps obtained for the same structures when a water molecule is used as the probe. Clear “spines of hydration” are predicted in the unoccupied A-tracts of both structures, and again there is no evident difference between the free DNA and 1:1 complex. This argues against cooperativity being the result of easier displacement of water from the second binding site, once the first molecule of drug is bound.

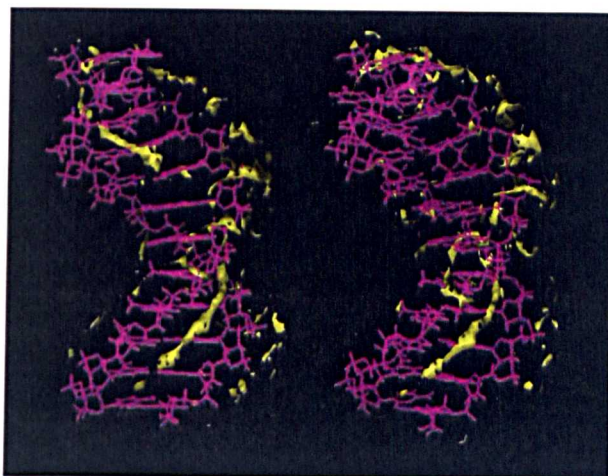


Figure 3.17 MIP plots for the time-averaged structures of the free DNA (left), and 1:1 complex (right) obtained using H33258 as the probe. The map is contoured at a favourable interaction potential of $-32 \text{ kcal mol}^{-1}$.

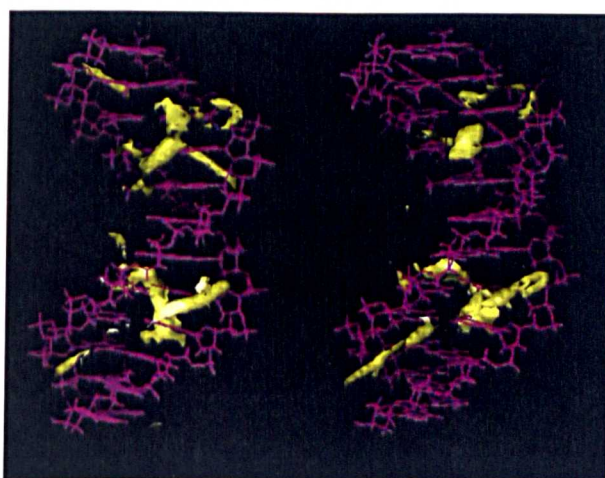


Figure 3.18 MIP plots for the time-averaged structures of the free DNA (left), and 1:1 complex (right) obtained using a water molecule as the probe. The map is contoured at a favourable interaction potential of $-10 \text{ kcal mol}^{-1}$.

The MIP maps only relate to the enthalpic components of probe-target recognition, but hydration density maps reflect, qualitatively, free energies of solvation. The hydration density maps for the free DNA and 1:1 complexes are shown in figure 3.19. Again, it is clear that, in areas not masked by the presence of a

molecule of the drug, the hydration patterns for the DNA in the two situations are very similar. These qualitative examinations suggest that the overall modestly positive value of $\Delta\Delta(E^{\text{intra}}+G^{\text{solv}})$ is not the result of a near cancelling out of individual $\Delta\Delta E^{\text{intra}}$ and $\Delta\Delta G^{\text{solv}}$ terms that are large in magnitude but opposite in sign. This adds further support for our argument that we can regard the configurational entropy term as being the critical one.

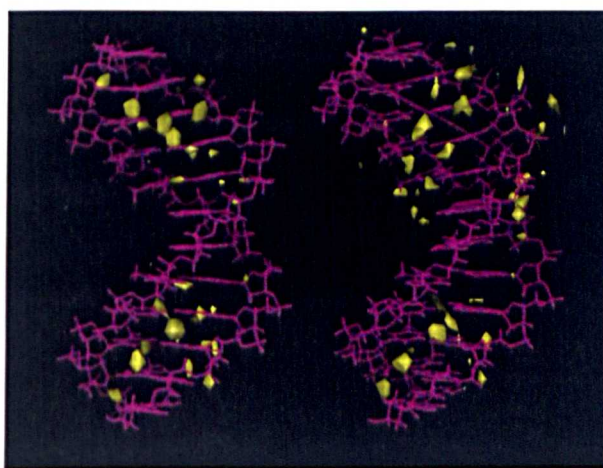


Figure 3.19 Hydration density maps for the free DNA and 1:1 complex, calculated by integrating water occupancies over the full 4 ns of the equilibrated trajectories. The map is contoured at a level corresponding to twice the density of the bulk solvent.

To gain further insight, the configurational entropy of the DNA component alone was calculated in the 1:1 and 2:1 complexes and these values compared with that for the free DNA (table 3.1). For the binding of the first drug, we find $T\Delta S$ to be $-25.5 \pm 0.8 \text{ kcal mol}^{-1}$, while for the binding of the second drug it is calculated to be $-13.1 \pm 1.0 \text{ kcal mol}^{-1}$. For the DNA alone then, $T\Delta\Delta S$ is estimated to be $12.4 \pm 1.1 \text{ kcal mol}^{-1}$. This is somewhat larger than that previously calculated for the whole drug–DNA system ($10.4 \pm 0.7 \text{ kcal mol}^{-1}$), and suggests therefore that when the second drug binds, though the DNA itself is not greatly further restrained, the

previously-bound drug molecule loses some of its residual flexibility. To investigate this further, the trajectories were examined to see if the longitudinal motion of the drug in the A-tract minor groove was reduced by the introduction of the new positive charge further along this groove. However, analysis of selected drug–DNA distances sensitive to such motion revealed no significant difference between the 1:1 and 2:1 complexes. In addition, a general reduction in the flexibility of the drugs was evident from the calculation of simple coordinate RMS fluctuations. In the 1:1 complex, the average RMS fluctuation of the drug atoms is 0.28 angstroms, while for the 2:1 complex it is 0.23 angstroms (averaged over both drug molecules).

In any system of this type, where the binding sites for the two ligand molecules are physically separated, co-operativity relies on the ability of the receptor to pass information regarding the occupancy or otherwise of one site to the other. To examine this, we calculated the configurational entropies for each half of the DNA separately, in each of the free, 1:1 and 2:1 complexes, table 3.2. First, we see that in the free DNA, the two halves of the DNA are calculated to have very similar entropies – a good test of the adequacy of the lengths of our simulations and the sampling. Binding of the first drug molecule is accompanied by a large reduction in the configurational entropy of that half of the DNA, as expected, but the “information” regarding occupancy is clearly also passed to the second, unoccupied half of the DNA for which the configurational entropy of the DNA is noticeably reduced. Binding of the second drug molecule to this site results in further conformational restriction, but interestingly results in a “message” being passed back to the first occupied site that results in an increase in its configurational entropy. The calculated entropies for the two half sites in the 2:1 complex are in close agreement, giving us some insight into the reliability of the approach. It is important to note that

the entropy components calculated here cannot be summed to equate to the total values calculated previously, since this approach neglects the configurational entropy due to the relative motion of the two halves of the DNA.

Table 3.2 Configurational entropies for top and bottom halves of the DNA in the free, 1:1 and 2:1 drug–DNA complexes. Values were calculated over the full 4 ns trajectories but not corrected to S_{∞} . All values are in kcal mol^{−1}, with T=300 K.

System	Top		Bottom	
	TS	TΔS	TS	TΔS
Free DNA	366.3		368.4	
		−24.6		−4.6
1:1 complex	341.7		363.8	
		6.4		−16.9
2:1 complex	348.1		346.9	

3.6 Conclusions

Several methods have been used in order to elucidate the co-operative interaction of H33258 with duplexes containing two adjacent sites. Using NMR methods, we have previously provided evidence of co-operative binding of H33258 with d(CTTTTGCAAAG)₂ and we also showed that co-operativity is present with the similar duplex d(CTTTTGGCCAAAAG)₂. In either NMR titration no evidences of an intermediate 1:1 complex have been recorded. NMR spectral complexity raised by H33258 orientational disorder did not allow us to evidence co-operativity with d(GAAAAGCTTTTG)₂ as we inverted the binding sites of the initial duplex and changed the junctions of the A-tracts. Quantitative DNase I has been used to probe the co-operative binding of H33258 with binding sites, within the same DNA

fragment, that had been analysed by NMR. Binding affinities calculated for each binding site are not significantly different and did not provide strong evidence of co-operativity. In addition, CD spectral titrations of H33258 have been carried out and provided evidence of differentiation between the 1:1 and 2:1 complex formation, and also confirmed the co-operative binding of H33258 within the (CTTTTGCAAAAG)₂ and also d(GAAAGCTTTTC)₂.

We have also used extended molecular dynamics simulations to study the free DNA, the 2:1 complex, and also the theoretical 1:1 complex and been able to calculate thermodynamic quantities relating to the two binding events, which were not accessible using any other method. The sequence-dependent interaction of DNA with molecules that bind in the minor groove involves a delicate interplay between enthalpic and entropic components of the recognition process. In general, examples of co-operativity in DNA recognition appear to owe this characteristic to enthalpic factors, which are generally fairly evident, at least in qualitative form, from structure determinations. These typically reveal close physical contacts between the two ligands, and/or a major structural deformation of the DNA that requires both ligands to stabilise it. In this case, NMR structure determination has shown that neither of these factors is operating. The MD studies reported here lead us to conclude that in this case, co-operativity is largely the result of the overall rigidity of the system. Binding of the first ligand restricts the flexibility of the DNA well beyond the actual binding site. Binding of the second ligand has little further effect. Both sites are already structurally fairly well predisposed towards ligand binding, the small adjustments required bear a modest enthalpic penalty, and though anti-cooperative, are outweighed by the entropic term.

The observation that ligand binding can sometimes induce changes in conformational flexibility in sites of the structure distant from the binding site has been demonstrated in the co-operative binding of two calcium ions to the EF-hand protein calbindin D_{9k} (21). Calcium ion binding to either site results in a slight decrease in second site flexibility. In contrast, binding of the second calcium ion has relatively small changes in flexibility.

In our system, the molecular origins of co-operativity in the absence of conformational change cannot be understood unless the dynamic properties of the system are taken into account. The results presented in this thesis illustrate the power of molecular simulation methods to investigate such phenomena, and highlight the general importance of flexibility in determining the properties of biomolecules such as DNA. Yet an element of rigidity, as well, is the key to the ability of this dodecamer to transmit “information” between the two drug binding sites. In ongoing investigations, the modulation of co-operativity by the sequence and length of the intervening DNA is being examined.

3.7 References

1. Yang, Y. W., Chen, Y. H., Pon, R. T. & Lown, J. W. 1996 Studies on Cooperative Binding of an Extended Distamycin A Analogue in the Minor Groove of DNA by NMR Spectroscopy. *Biochem. Biophys. Res. Commun.* **222**, 764-769.
2. Chen, F. M. & Sha, F. 1998 Circular Dichroic and Kinetic Differentiation of DNA Binding Modes of Distamycin. *Biochemistry* **37**, 11143-11151.
3. Petri, V. & Brenowitz, M. 1997 Quantitative Nucleic Acids Footprinting: Thermodynamic and Kinetic Approaches. *Curr. Opin. Biotechnol.* **8**, 36-44.
4. Brenowitz, M., Senear, D. F., Shea, M. A. & Ackers, G. K. 1986 Quantitative Dnase Footprint Titration – A Method for Studying Protein-DNA Interactions. *Method Enzymol.* **130**, 132-181.
5. Bailly, C., Hamy, F. & Waring, M. J. 1996 Cooperativity in the Binding of Echinomycin to DNA Fragments Containing Closely Spaced CpG Sites. *Biochemistry* **35**, 1150-1161.
6. Job, P. 1928 Formation and Stability of Inorganic Complexes in Solution. *Ann. Chim. (Paris)* **9**, 113-203.
7. Harris, S. A., Gavathiotis, E., Searle, M. S., Orozco, M. & Lughton, C. A. 2001 Cooperativity in Drug-DNA Recognition: A Molecular Dynamics Study. *J. Am. Chem. Soc.* **123**, 12658-12663.
8. Haq, I., Ladbury, J. E., Chowdhry, B. Z., Jenkins, T. C. & Chaires, J. B. 1997 Specific Binding of Hoechst 33258 to the d(CGCAAATTTGCG)₂ Duplex: Calorimetric and Spectroscopic Studies. *J. Mol. Biol.* **271**, 244-257.
9. Loontjens, F. G., Regenfuss, P., Zechel, A., Dumortier, L. & Clegg, R. M. 1990 Binding Characteristics of Hoechst-33258 with Calf Thymus DNA, Poly[d(A-T)], and d(CCGGAATTCCGG)₂ – Multiple Stoichiometries and Determination of Tight-Binding with a Wide Spectrum of Site Affinities. *Biochemistry* **29**, 9029-9039.
10. Abu-Daya, A. & Fox, K. R. 1997 Interaction of Minor Groove Binding Ligands with Long AT Tracts. *Nucleic Acids Res.* **25**, 4962-4969.
11. Fede, A., Labhardt, A., Bannwarth, W. & Leupin, W. 1991 Dynamics and Binding Mode of Hoechst 33258 to d(GTGGGAATTCCAC)₂ in the 1:1 Solution Complex as Determined by 2-Dimensional ¹H-NMR. *Biochemistry* **30**, 11377-11388.
12. Higgins, L. D. & Searle, M. S. 1999 Site-Specificity of Bis-Benzimidazole Hoechst 33258 in A-tract Recognition of the DNA Dodecamer Duplex d(GCAAAATTTTGC)₂. *Chem. Commun.* **18**, 1861-1862.

13. Gavathiotis, E., Sharman, G. J. & Searle, M. S. 2000 Sequence-Dependent Variation in DNA Minor Groove Width Dictates Orientational Preference of Hoechst 33258 in A-tract Recognition: Solution NMR Structure of the 2:1 Complex with d(CTTTTGCAAAAG)₂. *Nucleic Acids Res.* **28**, 728-735.
14. Packer, M. J., Dauncey, M. P. & Hunter, C. A. 2000 Sequence-dependent DNA Structure: Dinucleotide Conformational Maps. *J. Mol. Biol.* **295**, 71-83.
15. Cheatham, T. E. & Kollman, P. A. 2000 Molecular Dynamics Simulation of Nucleic Acids. *Annu. Rev. Phys. Chem.* **51**, 435-471.
16. Tsui, V. & Case, D. A. 2000 Molecular Dynamics Simulations of Nucleic Acids with a Generalized Born Solvation Model. *J. Am. Chem. Soc.* **122**, 2489-2498.
17. Schlitter, J. 1993 Estimation of Absolute and Relative Entropies of Macromolecules Using the Covariance-Matrix. *Chem. Phys. Lett.* **215**, 617-621.
18. Pope, L. H., Allen, S., Davies, M. C., Roberts, C. J., Tendler, S. J. B. & Williams, P. M. 2001 Probing DNA Duplex Formation and Drug-DNA Interactions by the Quartz Crystal Microbalance Technique. *Langmuir* **17**, 8300-8304.
19. Orozco, M., Jorgensen, W. L. & Luque, F. J. 1993 Comparison of 6-31*-Based MST/SCRF and FEP Evaluations of the Free-Energies of Hydration for Small Neutral Molecules. *J. Comput. Chem.* **14**, 1498-1503.
20. Soliva, R., Sherer, E., Luque, F. J., Laughton, C. A. & Orozco, M. 2000 Molecular Dynamics Simulations of PNA Center Dot DNA and PNA Center Dot RNA Duplexes in Aqueous Solution. *J. Am. Chem. Soc.* **122**, 5997-6008.
21. Akke, M., Skelton, N. J., Kordel, J., Palmer, A. G., III. & Chazin, W. J. 1993 Effects of Ion Binding on the Backbone Dynamics of Calbindin D_{9K} Determined by ¹⁵N NMR Relaxation. *Biochemistry* **32**, 9832-9844.

4. DNA minor groove recognition by a bis-benzimidazole analogue of Hoechst 33258

4.1 Introduction

Rational design of small molecules capable of recognising selectively the DNA minor groove of AT-rich or mixed sequences needs our clear understanding of the interactions involved. A number of NMR and X-ray crystal structures (1) have been reported for minor groove drugs complexed with short oligonucleotides containing AT-rich sequences, however, the relative importance of the various factors contributing to DNA minor groove recognition is still a matter of discussion. Generally, these structures show that all drugs bind in the narrow minor groove of the AT-rich sequences allowing optimal Van der Waals contacts with the floor and the walls of the minor groove, though, hydrogen-bonding interaction between the ligand and the base pair acceptors on the floor of the groove has been frequently observed (2). Thus, sequence selectivity of groove binders is more likely to be controlled by their ability to recognise variation in groove width.

The bis-benzimidazole family of ligands have been used as a starting point for rational design of new anticancer agents because of their synthetic accessibility and high binding affinities (1), figure 1.1. In chapter 2, it is described that some crystal structures locate the sterically bulky, non-planar N-methylpiperazine ring of H33258 **1** in the wider minor groove associated with GC base pairs at the end of the narrower AT-tract. The crystal structure of a H33258 analogue **2**, in which the N-methylpiperazine has been replaced by the smaller, more planar 4,5-dihydro-3H-

imidazol-1-ium (imidazoline) ring, bound with d(CGCGAATTCGCG)₂ duplex, results in a narrower minor groove that reflects the decrease in steric bulk (3). Although the imidazoline ring can be accommodated in a narrower AT minor groove, the geometry of the ring prevents it from hydrogen bonding with the groove floor. The increase in stability, observed by melting experiments, of the complex compared to H33258 complex with the same duplex is attributed to the better Van der Waals complementarity rather than hydrogen bonding. The NMR study of the complex of Hoechst 43254 with d(GGTAATTACC)₂ (4) duplex, in which the N-methylpiperazine ring of H33258 has been replaced by a 1, 3, 4, 5-tetrahydropyrimidin-1-ium ring **3**, shows that the drug prefers to bind in the narrower AT minor groove as a result of both reduced steric bulk and incompatibility of the hydrogen-bond donor of the pyrimidinium ring with the 2-amino group of guanine. However, the increase in thermal stability in this complex is suggested to be dominated by the additional hydrogen bonding from the pyrimidinium ring to the floor of the groove.

The crystal structure of a tris(benzimidazole) analogue **4** of H33258, called TRIBIZ, bound to the d(CGCAAATTTGCG)₂ duplex has been determined (5). This study examined the relative importance of isohelicity, a measure of Van der Waals complementarity, and phasing of the ligand hydrogen-bonding groups with the base pairs along the binding site. The structure shows that the ligand adopts a highly twisted conformation, with an overall twist 50°, that enables it to follow the helical curvature and maintain hydrogen-bonding complementarity with the floor of the groove. The phasing of the ligand is also ensured by the DNA unwinding slightly across the AT-tract. The ligand fails to have an optimal isohelical fit to the DNA

curvature suggesting that ligand isohelicity is less important than phasing for stabilisation of the TRIBIZ ligand.

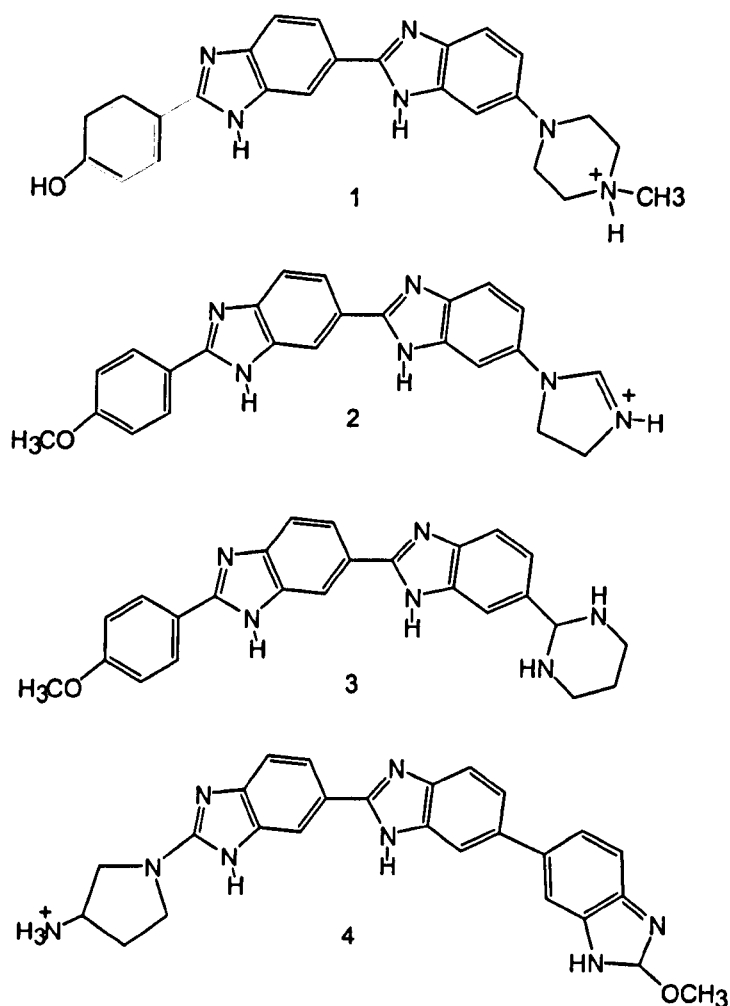


Figure 4.1 Chemical structures of Hoechst analogues 1–4.

In order to examine the relative importance of isohelicity and phasing of the minor groove binders, we selected an analogue of H33258, namely Hoechst 10 (H10), in which the N-methylpiperazine ring is linked to the phenyl end of the molecule rather than to the benzimidazole ring (figure 4.2) giving the molecule a

more linear arrangement. The overall shape of the ligand is less isohelical with the curvature of the minor groove, although, the phasing of the ligand hydrogen-bonding groups with the floor of the groove should remain the same as in H33258. Fluorescence titration studies using the DNA decamer d(GCAAATTTGC)₂ showed that H10 and H33258 bind with an indistinguishable binding affinity, $\Delta\Delta G \sim 0.3 \text{ kJ mol}^{-1}$ (6). Thus, the intrinsic difference in curvature between H10 and H33258 seems to have a little impact on DNA binding affinity. In this context, we studied the interaction of H10 with the DNA decamer d(GCAAATTTGC)₂ by NMR and employed a restrained molecular dynamics protocol that allows us to examine the key structural features of the complex.

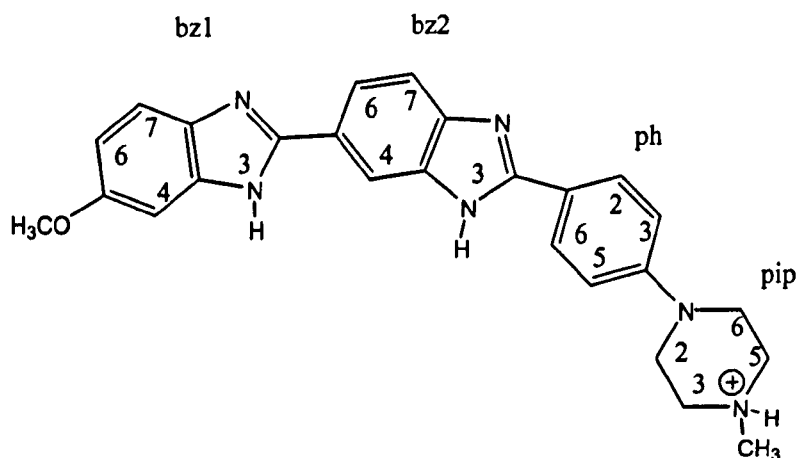


Figure 4.2 Chemical structure of Hoechst 10.

4.2 Materials and methods

4.2.1 Sample preparation and analysis

The DNA decamer d(GCAAATTTGC)₂ was synthesised and purified as described in section 2.2.1. The bis-benzimidazole analogue H10 was provided by

Hoechst, Germany and was used without any further purification after checking its purity by ^1H NMR. The drug–DNA complex was formed by titrating small aliquots of the drug solution in D_2O (~10 mM) into a solution of the oligonucleotide in 100mM NaCl and 10mM NaD_2PO_4 . Titration was monitored by ^1H NMR and stopped when the resonances of the free DNA were fully replaced by those of the 1:1 complex. NMR data were collected on a 2mM solution of the complex at pH=7.0 and 298 K as described in section 2.2.3. NMR restraints in the 1:1 complex were derived by integration of NOEs assigned from 75, 100 and 125 ms NOESY spectra in D_2O , and 200 and 100 ms NOESY data in H_2O solutions and calibration to a number of fixed reference distances: sugar $\text{H}2' - \text{H}2''$, cytosine $\text{H}5 - \text{H}6$, thymine $\text{CH}_3 - \text{H}6$, Bz1 $\text{H}6 - \text{H}7$, Bz2 $\text{H}6 - \text{H}7$ following the methodology described in section 2.2.4. A total of 316 distance restraints were used for structure calculations of the 1:1 H10–d(GCAAATTTTCG) $_2$ complex including a set of 33 drug–DNA NOEs. Tables with the list of the NOE restraints are found in appendix 2.4.

4.2.2 Structure calculations

Energy minimisations and restrained molecular dynamics of the 1:1 H10–d(GCAAATTTTCG) $_2$ complex were carried out using the SANDER module of AMBER 6 following the same procedure described in section 2.2.6. Force field parameters of H10 atoms were determined by using AMBER 95 force field parameters and partial charges (appendix 3.3) calculated using the HF/6-31G(d)/RESP methodology in order to maintain consistency with the AMBER force field charges. The structure of the decamer d(GCAAATTTGC) $_2$ was built initially as a canonical B–DNA with 18 Na^+ ions added to neutralise negative charges of the system and solvated by a periodic box of 216 TIP3P waters to a minimum distance of

5 Å around the solute. The periodic box contained 3407 water molecules resulting in the size of 45 Å x 50 Å x 60 Å. The H10 molecule was docked in three different positions to avoid bias of the binding sites (section 2.2.6), and one of the three complexes was selected based on the fitness of the drug–DNA restraints for a further of 1 ns restrained molecular dynamics. Snapshots of the simulations collected every ps and the structures were determined to equilibrate by RMSD analysis. Structures calculated after equilibration satisfied adequate the set of 316 restraints and a minimised structure has 43 distance violations that correspond to a 27.67 kcal mol⁻¹ energy penalty. The maximum distance violation was 0.38 Å, 9 restraints were violated by > 0.2 Å and 20 restraints by > 0.1 Å. Helicoidal parameters were analysed with CURVES 5.1 and average structure was calculated over the equilibrated trajectory with CARNAL module of AMBER 6.

4.3 Results and discussion

4.3.1 Structural analysis by NMR

The NMR experiments show that drug binding lifts the C_{2v} symmetry of the DNA decamer such that the two strands are not equivalent in the complex. This is clearly determined in the thymine methyl region, where six methyl resonances are identified, suggesting that the drug binds in a single high affinity binding site. Preliminary NMR analysis of the complex has been previously described (7). NOESY spectra (figure 4.3) of the 1:1 complex contain a number of chemical exchange cross peaks between corresponding protons in the non-equivalent strands of the duplex, indicating a slow dissociation of the drug from the complex and rebinding in a 180° related orientation with respect to each of the strands.

DNA chemical shift resonances have been identified following previously described methods in section 2.3.1 from a combination of NOESY and TOCSY spectra in D₂O and H₂O solutions. A full list of chemical shifts assignments for the complex is found in appendices 1.5 and 1.9. Part of the 250 ms NOESY spectrum (figure 4.4) shows the NOE sequential intra- and internucleotide connectivities between H1'–H8/H6 protons for each strand. A number of drug–DNA NOEs is also illustrated. The oligonucleotide numbering scheme adopted is as follows: d(5-G1–C2–A3–A4–A5–T6–T7–T8–G9–C10)·d(5'C11–G12–A13–A14–A15–T16–T17–T18–G19–C20). Drug protons are labelled according to the numbering scheme shown in figure 4.2. The assignment pathways in figure 4.4 reveal that H8–H1' intranucleotide cross peaks for the symmetry related base pairs G1/G11 and G10/G20, and also H6–H1' cross peaks of C2/C12 and C9/C19 are very similar. In contrast, the larger chemical shift differences are detected for the complementary adenine and thymine pairs at the centre of the sequence. This implies that the ligand binds towards the center of the duplex and not close to the strand ends.

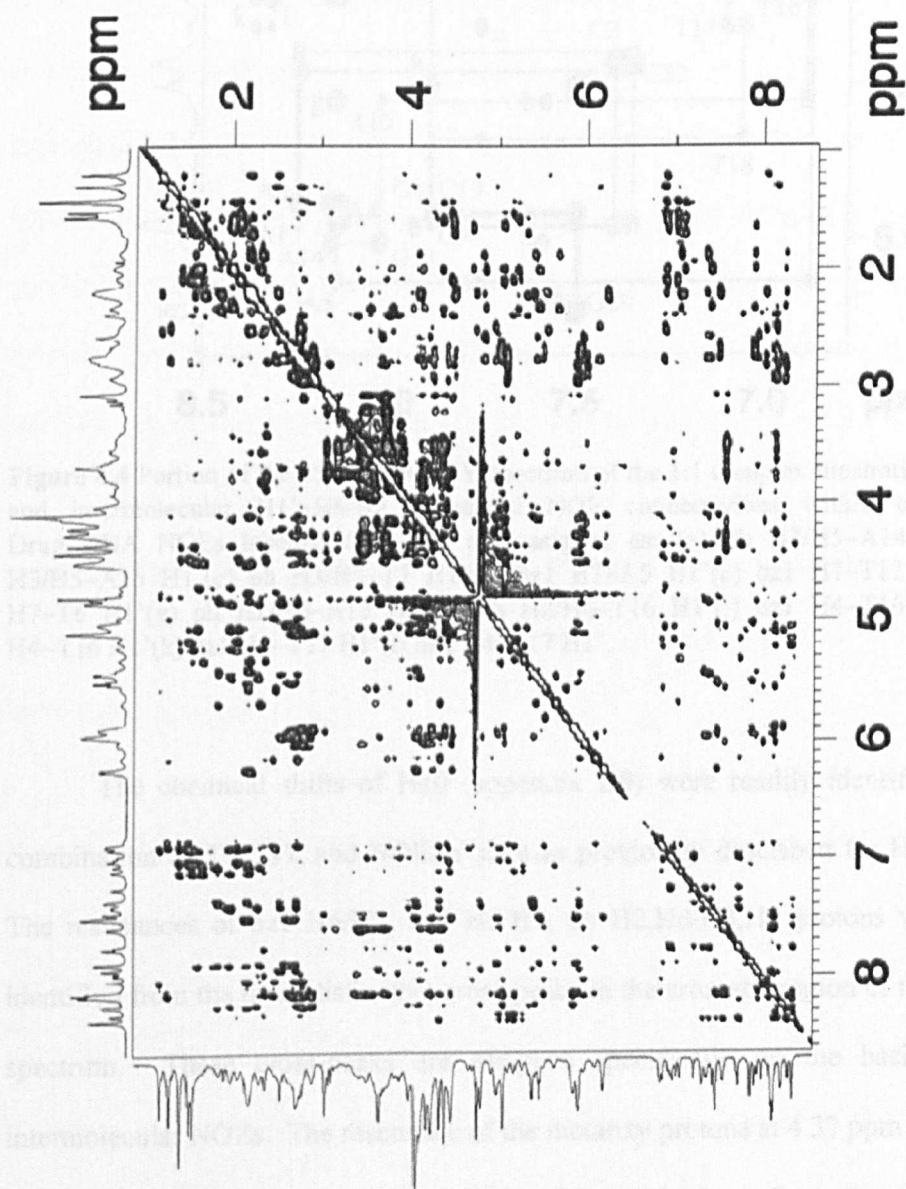


Figure 4.3 NOESY spectrum of the 1:1 H10-d(GCAAATTTCG)₂ complex in D₂O solution, recorded at 298 K with mixing time 250 ms.

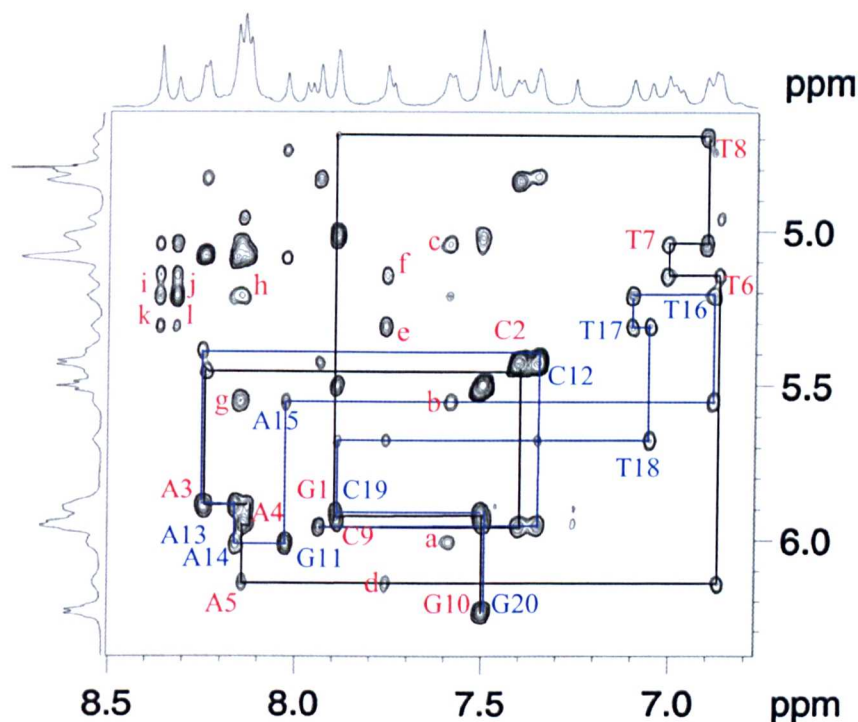


Figure 4.4 Portion of the 250 ms NOESY spectrum of the 1:1 complex illustrating the intra- and intermolecular H1'–H6/H8 sequential NOE connectivities within each strand. Drug–DNA NOEs labelled (a) – (l) are assigned as: (a) ph H3/H5–A14 H1' (b) ph H3/H5–A15 H1' (c) ph H3/H5–T7 H1' (d) bz1 H7–A5 H1' (e) bz1 H7–T17 H1' (f) bz1 H7–T6 H1' (g) ph H2/H6–A15 H1' (h) ph H2/H6–T16 H1' (i) bz1 H4–T16 H1' (j) bz2 H4–T16 H1' (k) bz1 H4–T17 H1' (l) bz2 H4–T17 H1'.

The chemical shifts of H10 (appendix 1.9) were readily identified from a combination of TOCSY and NOESY data as previously described for H33258 (8). The resonances of bz1 H6/H7, bz2 H6/H7, ph H2, H6/H3, H5 protons were easily identified from the three distinctive cross peaks in the aromatic region of the TOCSY spectrum. These cross-peaks are assigned specifically on the basis of their intermolecular NOEs. The resonance of the methoxy protons at 4.37 ppm has a NOE with bz1 H6/H7 protons allowing us to distinguish them from the bz2 H6/H7 protons. Only a single resonance is observed for the symmetrical H2, H6 and H3, H5 protons of the phenyl ring, which suggests a rapid conformational averaging between the two sides of the ring, consistent with 180° flip around the bz2–ph bond.

This rapid internal dynamic motion of the drug is in a faster time scale than the rate of dissociation of the drug from the duplex. The ph H3/H5 protons are distinguished from the ph H2/H6 because of their strong NOE to the pip H2/H6 protons. The methylene protons of the N-methylpiperazine ring are assigned on the basis of the NOE effects observed from the N-methyl protons, identified at 3.08 ppm. The benzimidazole NH proton resonances, identified in 90% H₂O, were assigned specifically by the observation of the NOE between bz2 NH-ph H2/H6. The benzimidazole NH proton assignment allowed for the identification of the bz1 H4 and bz2 H4. In addition, latter assignments are confirmed from the strong interaction between the bz1 H4 proton and the methoxy protons. The aromatic region of the NOESY spectrum (figure 4.5) highlights intramolecular drug NOEs such as bz1 H6/H7, bz2 H6/H7 and ph H2,H6/H3,H5 and a number of drug-DNA NOEs involving base H2 atoms of the adenine residues. NOEs between base H6/H8 protons are also illustrated and indicate good stacking interactions between the corresponding bases.

The binding site of H10 in the complex is identified with some precision from the set of 37 intermolecular NOEs (table 4.1) of the drug protons and DNA protons found on the walls and the floor of the groove. Figure 4.6 summarises schematically the intermolecular NOEs that establish unambiguously the orientation of the drug in the groove. Many of the NOEs are highlighted in the portions of the NOESY spectra in figure 4.4 and 4.5. For instance, NOE signals are established from the benzimidazole methoxy group to the base H2 protons of A3 and A4, and deoxyribose H1' protons of A3, A4 and T17 nucleotides and from the N-methylpiperazine ring to the base H2 of A14 and A13 and deoxyribose H1' protons of the A14, A13 and T8 nucleotides. In addition, the drug-induced perturbations of the DNA chemical shifts

are significant only for the deoxyribose H1' and base H2 protons of the A₃T₃ tract. There is no interaction between protons at the edges of the ligand with GC bases and clearly the intermolecular NOEs locate the bound ligand relative centrally within the A₃T₃ tract. This indicates that the sterically bulky N-methylpiperazine is located close to the end of the A₃T₃ tract but does not have any preference for GC recognition.

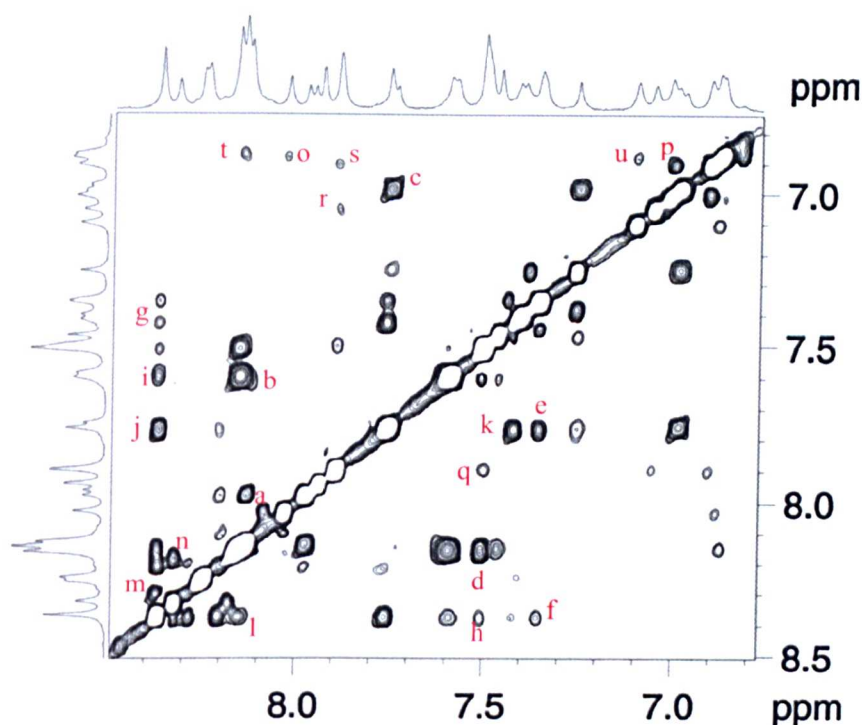
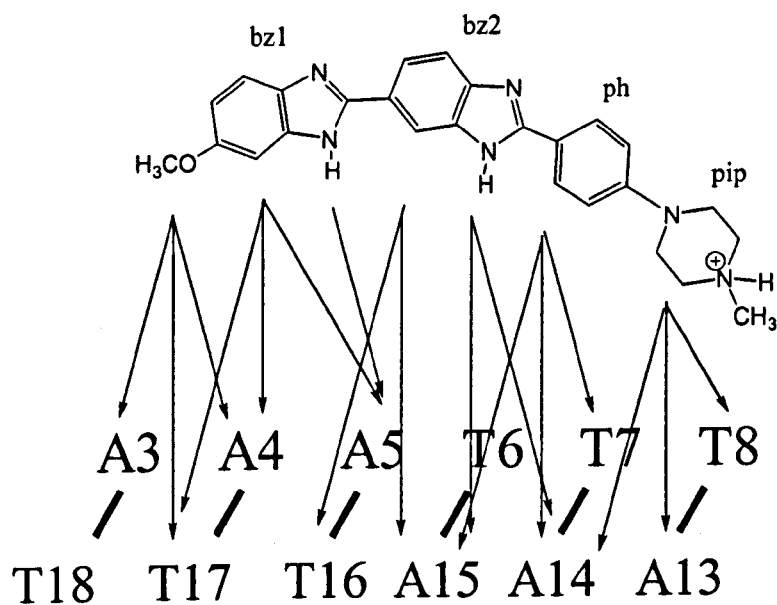


Figure 4.5 Portion of the 250 ms NOESY spectrum of the 1:1 complex illustrating several intramolecular and intermolecular NOEs. NOEs labelled (a) – (u) are assigned as: (a) bz2 H6–bz2 H7 (b) ph H2/H6–ph H3/H5 (c) bz1H6–bz1 H7 (d) ph H2/H6–A14 H2 (e) bz1 H7–A4 H2 (f) bz1 H4–A4 H2 (g) bz1 H4–A5 H2 (h) A14 H2–A15 H2 (i) ph H3/H5–A15 H2 (j) bz1 H4–bz1 H7 (k) bz1 H7–A5 H2 (l) ph H2/H6–A15 H2 (m) bz2 H4–A15 H2 (n) bz2 H4–bz2 H7 (o) A15 H8–T16 H6 (p) T7 H6–T8H6 (q) G9 H8–C10 H6, G19 H8–C20 H6 (r) T18 H8–C19 H6 (s) T8 H8–G9 H8 (t) T6 H6–A5 H8 (u) T17 H6–T16 H6.

Table 4.1 Intermolecular NOEs in the 1:1 H10-d(GCAAATTTTCG)₂ complex.

Drug protons	DNA protons
ph H2/H6	A14 H2, A15 H1', A15 H2, T16 H1'
ph H3/H5	A14 H2, A15 H1', A15 H2, T7 H1', A14 H1'
pip H2/H6	A14 H1', A14 H2
pip H3/H5	A14 H1', A14 H2
bz1 H7	A5 H1', T16 H1', T17 H1', A5 H2, A4 H2
bz1 H4	A5 H2, T16 H1', T17 H1', A4 H2
bz2 H4	T16 H1', T17 H1', A15 H2
O-Me	A5 H2, A4 H2, A4 H1', T17 H1', A5 H1', A3H2
N-Me	A13H2, A13H1', T8H1'
bz1 H3	A5H2
bz2 H3	A15H2, A14H2

**Figure 4.6** Schematic representations of some of the intermolecular NOEs observed in the 1:1 complex. Only the central base pairs of the sequence are illustrated.

4.3.2 Structure of the 1:1 H10–d(GCAAATTTGC)₂ complex

The NMR restrained molecular dynamics simulation produced stable structures within the 1 ns of the run with RMS deviation from the mean structure over all heavy atoms 1.10 ± 0.16 Å. The last 500 ps of the simulation was selected for structural analysis and the average structure of the 1:1 complex from this part was minimised and shown in figure 4.7. The agreement between the experimental and calculated distances, specifically between the intermolecular distances, is consistent with the drug binding at a single site. The drug is bound in a single location within the minor groove of the 5'-AATTT sequence with the N-methylpiperazine ring spanning towards the 3'-end of it. Similarly, X-ray crystal structures of d(CGCAAATTTGCG)₂ complexed with H33258 determined the location of the drug within AATTT (9) and ATTTG (10) sequences with the N-methylpiperazine ring close to the GC regions.

There are a number of close Van der Waals contacts between the ring subunits of the drug and the walls of the minor groove, which involve hydrogen atoms attached to the C1', C4' and C5' on the phosphodiester backbone. The methoxy group laying deep in the groove and is interacting with groove walls, C1' and O4' atoms of A5 and also with the floor of the groove with C2 atoms of A4 and A5 and O2 of T17. The benzimidazole 1 ring has close contacts with the C1', O4' and C4' atoms of T17 from one strand and C1' and O4' atoms of T6 and C5' atom of T7 from the second strand. The benzimidazole 2 ring has close contacts with the C1', O4', C4' and C5' atoms of T16 from one strand and C1', O4', C4' and C5' atoms of T7 from the second strand. Close contacts between the phenyl ring and the C1' atom of A15 and C1', O4' and C4' atoms of T8 from the second strand are also

observed. The bulky N-methylpiperazine ring, which is accommodated where the minor groove is wider at the 3'-end of the binding site (see below), has preferably a chair conformation that is involved in Van der Waals contacts (including the methyl group) with the C1', O4' and C4' atoms of both A15 and A14. The carbon-carbon distance of these interactions with the walls of the groove is between 3.3 and 4.1 Å. In addition, each subunit of the drug except the N-methylpiperazine ring has close contacts ~3.5 Å with atoms at the floor of the groove like C2 atoms of the adenines. Thus, H10 in this complex has a number of intermolecular non-bonded contacts with the minor groove of the 5-AATTT sequence and in agreement with binding studies (6), the more linear shape of the molecule seems not to be crucial factor for strong binding affinity.

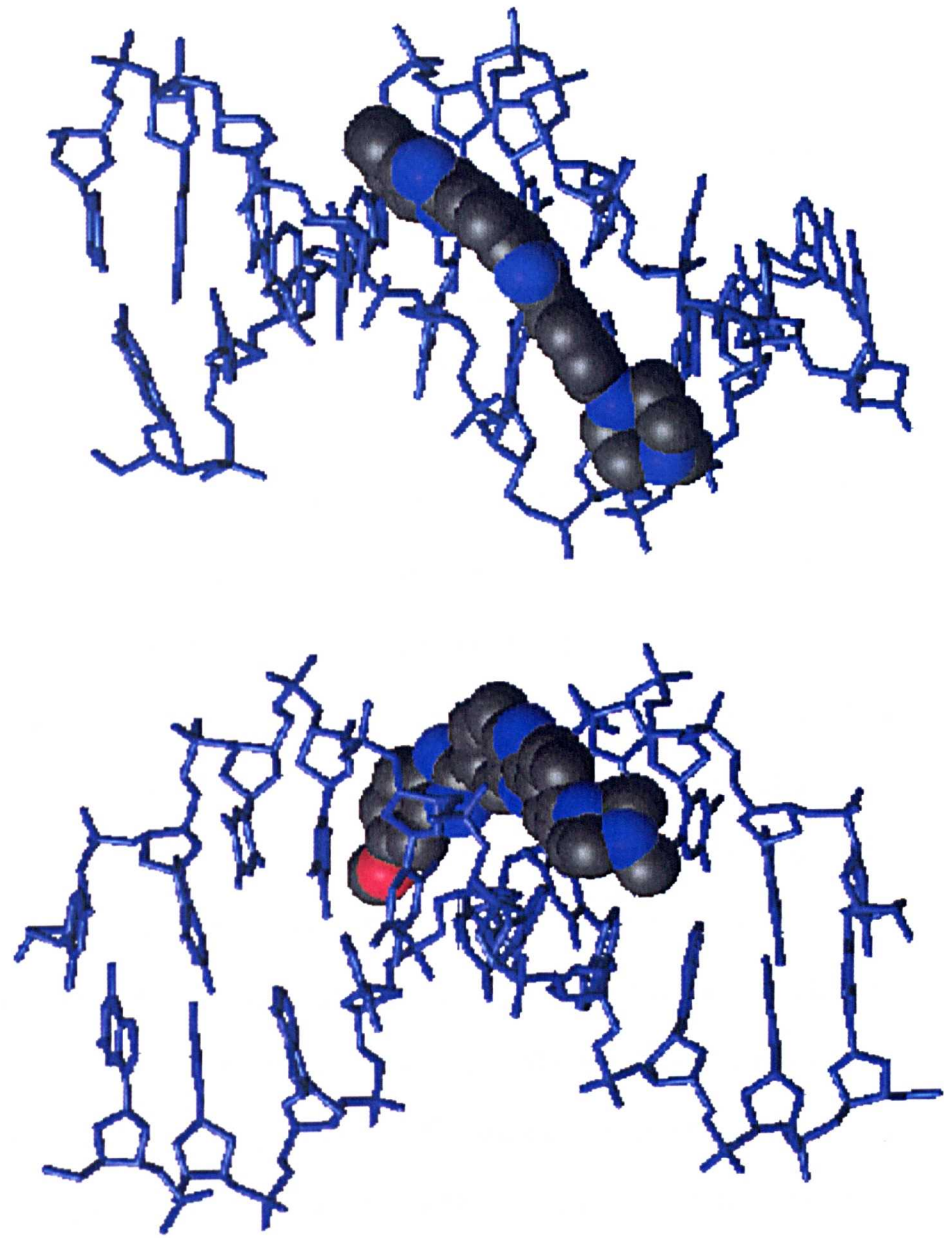


Figure 4.7 Two views of the average minimised structure of the 1:1 H10-d(GCAAATTTCG)₂ complex taken from the last 500 ps.

In contrast with the TRIBIZ complex (5), there is no need for H10 to adopt a highly twisted conformation in order to follow the helical curvature of the minor groove. Specifically, the benzimidazole–benzimidazole (bz1–bz2) and the benzimidazole–phenyl (bz2–ph) torsion angles have average values of 18.8 ± 9.5 and -1.2 ± 10.1 , respectively over the last 500 ps. The torsion angle between the phenyl and the piperazine ring (ph–pip) has an average value of 90.6 ± 12.1 . The torsion angle values of the bz1–bz2 and bz2–ph rings in this complex are within the range reported for X-ray and NMR structures of H32258 complexes (4). Hydrogen-bond contacts between H10 and the DNA decamer are apparent, table 4.2. The NH atom of benzimidazole 1 forms three-centre hydrogen bonds to O2 atoms of thymine T6 and T16 and the NH atom of benzimidazole 2 forms hydrogen bond to O2 of thymine T7. The average hydrogen-bond distance between bz2 NH proton and N3 of adenine A15 is 3.5 ± 0.2 Å but with an acceptable angle of $127.4 \pm 9.6^\circ$, which suggests that this is rather a weak interaction. However, most of the three-centre hydrogen bonds on biological small molecules exhibit unsymmetrical geometry with major and minor component (2). Therefore, it is not uncommon for the minor components of three-centre hydrogen bond to show H---A distance of > 2.9 Å. H33258 benzimidazole moieties in crystal structures (9,10) with d(CGCAAATTTGCG)₂ duplex found to form hydrogen bonds with the same adenine and thymine bases. Thus, the weak interaction between bz2 NH proton and A15 N3 atom is likely due to a NOE restraint effect and not the inherent ability of the ligand to hydrogen bond. In addition, the piperazine ring is not involved in any direct hydrogen bond with the DNA. Figure 4.8 shows H10 within the binding site and highlights the atoms that contribute to hydrogen bonds.

Table 4.2 Mean intermolecular hydrogen-bonding distances and angles, in the 1:1 complex with d(GCAAATTTGC)₂ taken over the final 500 ps of restrained molecular dynamics simulations.

Hydrogen-bond distance (H---X, Å) and angles [N-H---A(N3)/T(O2), degrees]		
H10		
Bz1 NH	T6 O2 2.1 (0.2) [144.0 (7.6)]	T16 O2 2.7 (0.3) [127.0 (10.4)]
Bz2 NH	A15 N3 3.5 (0.2) [127.4 (9.6)]	T7 O2 2.1 (0.2) [142.3 (11.9)]

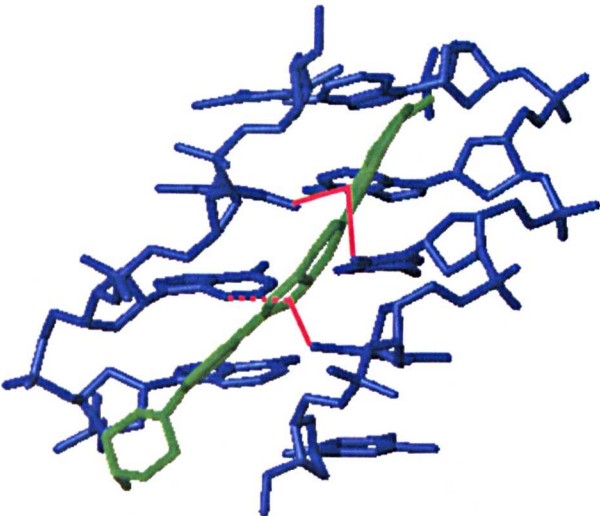


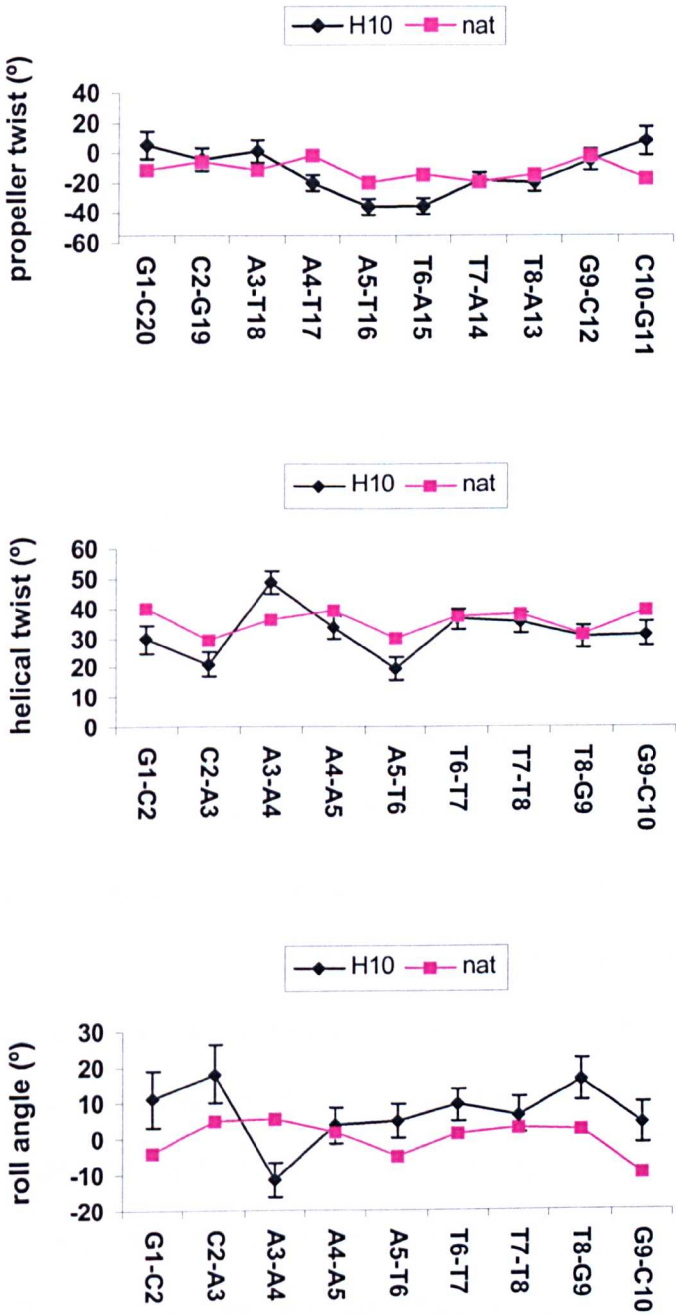
Figure 4.8 H10 within the binding site of the 1:1 H10–d(GCAAATTTGC)₂ complex. Red lines highlight the hydrogen-bond contacts with the DNA adenine N3 and thymine O2 hydrogen-bond acceptors.

The DNA helical parameters have been analysed over the final 500 ps of the dynamics simulation and average values along with standard deviations calculated, table 4.3. Table 4.3 also includes values of the helical parameters from the native crystal structure d(CGCAAATTTTCGC)₂ (11). We have also calculated the variation of propeller twist, helical twist, roll angle and minor groove along the sequence

during the last 500 ps of the simulation and plotted them (figure 4.9) along with the relative values from the native crystal structure d(CGCAAATTTTCGC)₂. Overall, the structure of the duplex is within the range of B-DNA conformation. The structural parameters of the AAATTT sequence in the 1:1 complex are not significant different from those in the native structure, however, sequence-dependent effects in propeller twist and helical twist are evident in the complex.

Table 4.3. Average values and standard deviations for main helical structural features of d(GCAAATTTTCG)₂ (A) calculated from the final 500 ps of the 1 ns restrained simulations of the 1:1 complex. Corresponding values from the native crystal structure of d(CGCAAATTTTCGC)₂ (B) are also given.

	A		B
	average	sd	X-ray
Pucker (°)	132.2	31.6	138.6
Twist (°)	31.7	8.5	35.5
Roll (°)	6.8	8.6	1.3
Rise (°)	3.4	0.4	3.4
Tilt (°)	2.6	8.6	5.1
Slide (°)	-0.6	0.7	-0.1
Shift (°)	0.3	0.8	0.4
Propeller (°)	-12.3	15.9	-12.2
Inclination (°)	3.6	2.7	0.5
X-disp (Å)	-1.1	0.2	-0.3
P20-P4	9.4	1.1	8.9
P19-P5	7.0	1.0	6.9
P18-P6	7.7	0.7	5.5
P17-P7	5.8	0.5	6.2
P16-P8	5.5	0.4	5.8
P15-P9	6.4	0.5	6.9
P14-P10	8.0	0.7	8.6



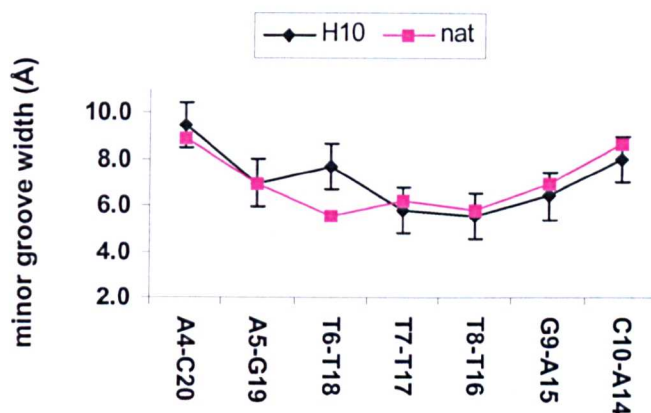


Figure 4.8 Sequence-dependent analysis of propeller twist, helical twist, roll angle and minor groove width (P_i-P_{j+3} distance minus 5.8 Å) in the 1:1 complex of H10 with $d(\text{GCAAATTTGC})_2$ (black markers), and for the native duplex $d(\text{CGCAAATTTGCG})_2$ (magenta markers); errors bars represent one standard deviation from the mean value over the final 500 ps of the 1 ns dynamics simulation.

Propeller twist is similar across the sequence in the free and bound state with values typical of AT rich sequences (12), but the central A5–T16 and T6–A15 base pairs of the complex have significantly negative ($\sim -36^\circ$) twist in the structure. The higher propeller twist for the central base pairs in the AAATTT tract compared to the native structure is not correlated with narrower minor groove in the structure of the complex. The plot of minor groove width shows a general trend of narrowing towards the centre of the AAATTT tract. Drug binding causes only a local increase in the groove width suggesting that the minor groove of the duplex has an optimum width for H10 interaction. However, the variation from the average values suggests that a degree of minor groove breathing is important for adjusting the intermolecular interactions.

The helical twist of the duplex in the 1:1 complex has an average value of 31.7° , typical of B–DNA duplexes (12), and the average helical twist of the AAATTT sequence is 34° and 36° for the complex and the native structures,

respectively. However, the helical twist shows sequence-dependent effect at the A5–T6 base-pair step, which has significantly low average value of $\sim 19^\circ$. The low helical twist of the A5–T6 step indicates a local unwinding of the duplex, which may be correlated with the local increase in minor groove width. In contrast, exceptionally high twist of the helix is observed at the A3–A4 ($\sim 49^\circ$) step, which compensates for the low value of the A5–T6 step. Such high twist values are also evident in ApT steps of the TRIBIZ complex with a non-complementary duplex containing a AAATAT sequence (13). Interestingly, the average values of the roll angle of the 1:1 complex structure, except at the A3–A4 step, are positively increased compared to the native structure. However, the increase in roll angle, at least for this complex structure, is not associated with changes to the minor groove width.

In the 1:1 complex, H10 binds strongly within the AATTT sequence with a number of non-bonded Van der Waals interactions with the walls and the floor of the groove. Bifurcated hydrogen bonds from the benzimidazoles to DNA acceptors on the floor of the groove are also present in the complex, but one of them appears not optimum. The N-methylpiperazine group due to steric requirements binds where the minor groove starts to widen, while hydrogen bonding between the benzimidazole moieties is possible with a number of AT bases. Thus, the position of the drug is determined by optimal fit with the groove. This also has been seen in a number of minor groove complexes including H33258 and TRIBIZ complexes with AAATTT sequences (9,5). Isohelicity of the ligand is optimal when there are close Van der Waals contacts between the C2 atoms of adenines and atoms at the concave surface of the ligand. In our complex, such contacts are evident from all the aromatic ring subunits of H10. Although the shape of H10 is less isohelical with the curvature of

the minor groove, isohelicity and hydrogen bonding are ensured without any significant distortion of the drug conformation. H10 binding induces some changes to the DNA conformation, best described by propeller twist, helical twist and roll angle. The structural analysis show that the DNA sequence has an optimum minor groove width for H10 binding and is also flexible enough to adopt the geometrically optimal conformation compatible with the necessary binding interactions. Phasing of H10 within the minor groove appears to be achieved partially through induce-fit.

4.4 Conclusions

Binding studies by fluorescence titration showed previously that both H10 and H33258 have almost the same binding affinity with the DNA decamer d(GCAAATTTGC)₂. The structure of the 1:1 H10–d(GCAAATTTGC)₂ complex studied here revealed that even though the drug is less isohelical than H33258, the drug forms a number of non-bonded Van der Waals interactions while phasing with the hydrogen-bond acceptors of the bases-pairs is also observed. The structure of d(GCAAATTTGC)₂ is shown to be flexible enough to adopt a conformation for increasing intermolecular interactions, showing the important role of DNA structural flexibility in drug–DNA recognition. Isohelicity and phasing have a role in drug design of minor groove agents, however, when sequences with more than four AT bases are targeted then DNA flexibility and sequence-dependent structure must also be considered.

4.5 References

1. Neidle, S. 2001 DNA Minor-Groove Recognition by Small Molecules. *Nat. Prod. Rep.* **18**, 291-309.
2. Tabernero, L., Bella, J. & Aleman, C. 1996 Hydrogen-Bond Geometry in DNA-Minor Groove Binding Drug. *Nucleic Acids Res.* **24**, 3458-3466.
3. Wood, A. A., Nunn, C. M., Czarny, A., Boykin, D. W. & Neidle, S. 1995 Variability in DNA Minor-Groove Width Recognized by Ligand. *Nucleic Acids Res.* **23**, 3678-3684.
4. Bostock-Smith, C. E., Laughton, C. A. & Searle, M. S. 1998 DNA Minor Groove Recognition by a Tetrahydropyrimidinium Analogue of Hoechst 33258: NMR and Molecular Dynamics Studies of the Complex with d(GGTAATTACC)₂. *Nucleic Acids Res.* **26**, 1660-1667.
5. Clark, G. R., Gray, E. J., Neidle, S., Li, Y. H. & Leupin, W. 1996 Isohelicity and Phasing in Drug-DNA Sequence Recognition: Crystal Structure of a Tris(benzimidazole)-Oligonucleotide Complex. *Biochemistry* **35**, 13745-13752.
6. Bostock-Smith, C. E. & Searle, M. S. 1999 DNA Minor Groove Recognition by Bis-Benzimidazole Analogues of Hoechst 33258: Insights into Structure-DNA Affinity Relationships Assessed by Fluorescence Titration Measurements. *Nucleic Acids Res.* **27**, 1619-1624.
7. Bostock-Smith, C. E. 1998 DNA Structure, Dynamics and Recognition. Ph.D Thesis, University of Nottingham.
8. Searle, M. S. & Embrey, K. J. 1990 Sequence-Specific Interaction of Hoechst-33258 with the Minor Groove of an Adenine-Tract DNA Duplex Studied in Solution by ¹H-NMR Spectroscopy. *Nucleic Acids Res.* **18**, 3753-3762.
9. Spink, N., Brown, D. G., Skelly, J. V. & Neidle, S. 1994 Sequence-Dependent Effects in Drug-DNA Interaction. *Nucleic Acids Res.* **22**, 1607-1612.
10. Vega, M. C., Saez, I. G., Aymami, J., Eritja, R., Vandermarel, G. A., Vanboom, J. H., Rich, A. & Coll, M. 1994 3-Dimensional Crystal-Structure of the A-Tract DNA Dodecamer d(CGCAAATTTGCG)₂ Complexed with the Minor-Groove-Binding Drug Hoechst-33258. *Eur. J. Biochem.* **222**, 721-726.
11. Edwards, K. J. Brown, D. G., Spink, N., Skelly, J. V. & Neidle, S., 1992 Molecular Structure of the B-DNA Dodecamer d(CGCAAATTTGCG)₂. *J. Mol. Biol.* **226**, 1161-1173.
12. Neidle, S. 1998 *Oxford Handbook of Nucleic Acid Structure*, Oxford University Press, New York.

13. Aymani, J., Nunn, C., M. & Neidle, S. 1999 DNA Minor Groove Recognition of a Non-Self-Complementary AT-rich Sequence by a Tris-Benzimidazole Ligand. *Nucleic Acids Res.* **27**, 2961-2968.

5. Investigations into the structure of quadruplex DNA

5.1 Introduction

Nucleic acid molecules can adopt several different conformations from the well established A, B and Z duplex forms to recently studied structures such as triplexes, i-DNA formed by cytosine-rich sequences and G-quadruplex formed by guanine-rich sequences (1). There has been a great interest in studying the structure, stability and regulating factors of formation of G-quadruplex DNA since it was found that telomeric DNA (2–5), gene promoter regions (6–7) and immunoglobulin switch regions (8) have continuous repeats of G-rich sequences. Such DNA sequences have the ability to form G-quadruplex structures and in many cases this is well demonstrated under laboratory conditions (9–11). In addition, a number of proteins have been identified that exhibit specific recognition of parallel-stranded G-quadruplexes (12–13) or promote G-quadruplex formation (14). Although such DNA sequences have a strong propensity to form G-quadruplex structures, evidence of a functional role of the G-quadruplex structure *in vivo* remains elusive.

Telomeres are found at the ends of chromosomes and contain tandem repeats of guanine-rich DNA sequences of several kilobases that protect them from recombination, nuclease degradation and end-to-end fusions (2–3). Examples of such sequence repeats are TTAGGG, TTGGGG, TTTTGGGG, TTTAGGG and TTTTAGGG, which are found in telomeres of human, *Tetrahymena*, *Oxytricha*, *Arabidopsis*, *Chrorella* and *Chlamydomona*, respectively (3). Telomeric sequences

have the potential to form structures held together by guanine tetrads by either the folding back of a repetitive sequence or formation of a hairpin dimer. Thus, G-quadruplex formation may be involved in capping the chromosome end with a structure resistant to nucleases or to association of chromosomes. Recent efforts suggest that G-quadruplex structure plays an important role in the regulation of telomerase action suggesting a potential role as a therapeutic target for telomerase inhibition in cancer therapy (15–16).

The basic unit of the G-quadruplex is the G-tetrad, which has a geometry formed by a hydrogen-bonded network of square, planar aligned guanines, figure 5.1. The adjacent guanines found in the G-tetrads are hydrogen-bonded through their Watson-Crick and Hoogsteen edges with their carbonyl groups directed towards the interior of the G-tetrad. One of the amino protons of the guanines is involved in hydrogen bonding with the adjacent guanine and the other is exposed to the solvent. The G-quadruplex formation requires the presence of a monovalent cation such as Na^+ or K^+ , with the K^+ cations forming the most stable structure (17). The precise localisation of Na^+ is well defined in the high resolution crystal structure (0.95 Å) of d(TG₄T). Co-ordination sites associated with the O6 atoms of guanine are found midway between G-tetrad planes and involve octahedral co-ordination of Na^+ or K^+ (18), figure 5.1. An NMR approach for direct observation of interaction between NH_4^+ and the quadruplex d(G₄T₄G₄)₂ located the cation binding site equidistant from each G tetrad while no binding site was observed between the T₄ loops (19).

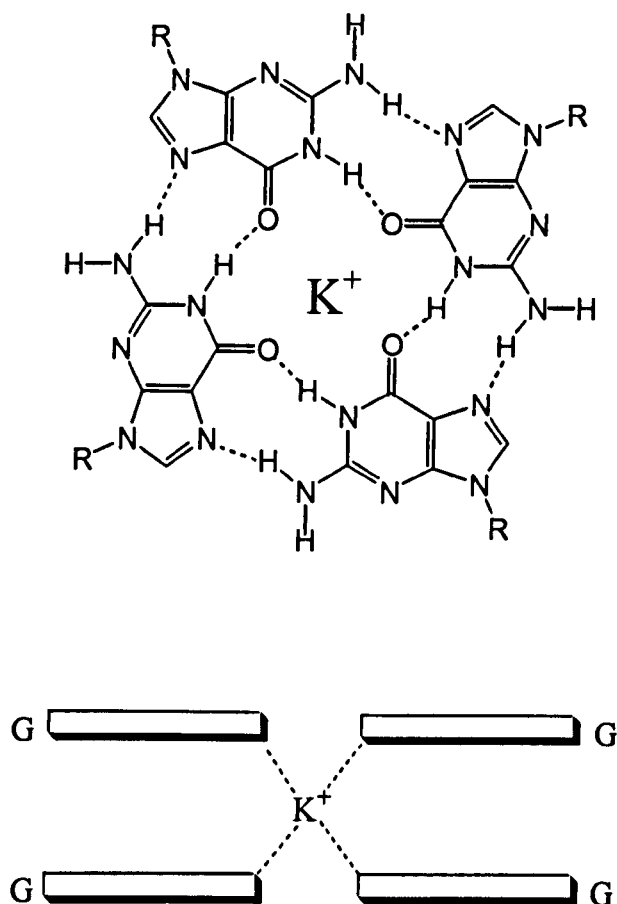


Figure 5.1 Hydrogen-bonding alignment of guanine nucleotides in a guanine tetrad and coordination scheme of K^+ midway between G-tetrad planes.

Extensive structural studies by X-ray and NMR approaches on telomeric DNA sequences indicate that they form different G-quadruplex folding topologies depending on the nature of the sequence, the length of the sequence and the presence of different cations in solution (11). The quadruplex structures can adopt parallel or antiparallel strand orientation and the glycosidic torsion angles of guanine nucleotides can vary between *syn* and *anti* conformation. There are three general types of quadruplex structure based on the structures resolved so far. These are classified by the relative directionality of individual strands and the conformation of

glycosidic torsion angles of guanine nucleotides within the G-tetrads. First, the quadruplex structure formed by association of four parallel strands has all the guanine nucleotides with *anti* conformation. Second, the *syn-anti-syn-anti* type structure has the guanine residues alternating between *syn* and *anti* conformation within G-tetrads and adjacent strands having antiparallel directionality. Finally, the structure can have guanines within each tetrad that alternate *syn-syn-anti-anti*, with adjacent strands having both parallel and antiparallel directionalities, figure 5.2.

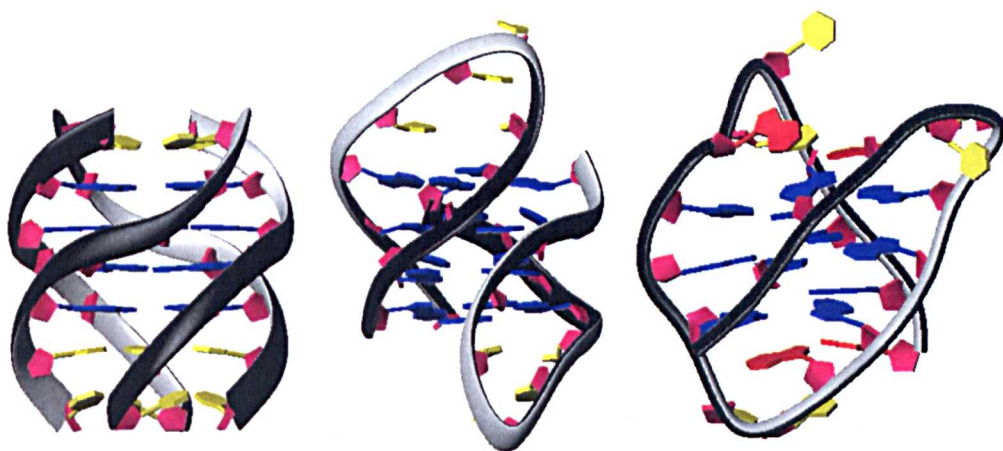


Figure 5.2 Examples of the different folding topology of the quadruplex structures: (a) parallel $d(\text{TTGGGGT})_4$ (20), (b) antiparallel dimer $d(\text{GGGGTTTGGGG})_2$ (21), (c) parallel-antiparallel intramolecular structure $d(\text{AGGG}(\text{TTAGGG})_3$ (22). G-tetrads are coloured in blue.

The quadruplex structures of the last two categories that contain antiparallel strands and have a combination of *syn* and *anti* glycosidic conformations are formed exclusively from dimeric hairpins (figure 5.2b) or unimolecular structures (figure 5.2c). The *Oxytricha* telomeric sequence $d(\text{GGGGTTTGGGG})_2$ forms a quadruplex structure in a crystalline state by association of two hairpin sequences, with each strand having an antiparallel neighbouring strand where the glycosidic

torsion angles of the guanines alternate *syn-anti-syn-anti* in each G-tetrad. The human telomeric sequence d(AGGG(TTAGGG)₃ forms an intramolecular quadruplex structure in Na⁺ solution, with each strand having one antiparallel and one parallel neighbouring strand and *syn-syn-anti-anti* glycosidic torsion angles conformation in each G-tetrad. The significant difference between these quadruplex structures and the parallel structure is the existence of loop connectivities, which usually consist of adenine and thymine residues. The orientation of loop connectivities is dependent on the sequence and plays a role in the general folding topology of the quadruplex structure.

Parallel-stranded G-quadruplex structures of the following sequences d(T₂G₄T)₄ in K⁺ (20), d(T₄G₄)₄ in K⁺ (21), d(TG₄T)₄ in Na⁺ (22) have been solved in solution by a combined NMR and molecular dynamics approach. These structures consist of four right-handed helices with all the residues adopting *anti* glycosidic torsion angles and predominantly C2'-endo sugar pucker conformations. Each G-tetrad in these studies is well defined, adopting a coplanar alignment and stacking well between the two adjacent G-tetrads with minimal overlap of guanine base planes. However, the thymine residues are less well defined, sampling multiple conformations in solution and presenting no evidence of T-tetrad formation. The crystal structure of the d(TG₄T) sequence (22) at 0.95 Å resolution shows qualitatively similar conformation with the solution structure, except that the thymine residues do not contribute to the four-stranded conformation and project away from the core quadruplex structure. This study also resolved the position of sodium ions between the tetrad planes in co-ordination with the O6 of the guanine residues and defined water molecules within the tetraplex helical grooves clustering around the

backbone phosphates, suggesting a role of water molecules in the stacking behaviour of the guanines.

A closely related RNA quadruplex structure of d(UGGGGU)₄ (23) in K⁺ solution shows significant stacking of the terminal uracils with the guanine tetrads. Also, at low temperature the 3' terminal uracils appear to form a U-tetrad as uracil imino protons are observed to be protected from solvent exchange. This is not observed with the terminal thymines of the previously mentioned studies because of lack of NOEs to define the conformation of the terminal residues, even though thymine residues are capable of forming stable T-tetrads when sandwiched between guanine residues (24). Recently, a parallel-stranded solution structure of d(TGGTGGC)₄ in K⁺ solution (25) revealed a stable T-tetrad formation in the center of the quadruplex despite underwinding of the right-handed helix and poor stacking across the T4–G5 step. Thymines at the 5' terminal also form a stable T-tetrad, but only when the pH is shifted from 7.0 to 4.8.

The subject of tetrad formation and stacking interactions with guanine tetrads by the neighbouring nucleotides in parallel-stranded quadruplexes has been explored with the NMR quadruplex structures of d(AGGGT)₄ and d(TAGGGT)₄ in the presence of K⁺ ions (26). This comparison study revealed that although the G₃ segments in the quadruplexes have largely similar structures, only the adenine residues in the d(AGGGT)₄ structure form an A-tetrad and have good stacking with the adjacent G-tetrad. This was the first time that adenine residues demonstrated the potential to be held in a plane by hydrogen bonds, although they are positioned at the end of the sequence. Surprisingly, thymine and adenine residues in the d(TAGGGT)₄ quadruplex structure sample a range of conformations resulting in poor stacking and formation of A-tetrads. It is obvious that there is a certain degree of mobility of the

terminal residues in the quadruplex structures that prevents good stacking with the core G-quadruplex structure, which usually affects the stability and conformation of the adjacent residues. Thus, formation of tetrad square alignment by other residues except guanines is highly dependent on the content of the sequence and their location within it.

Intramolecular G-quadruplexes, which contain repeats of guanine-rich DNA sequences, have attracted much interest in the last decade, as they are more likely to be formed at the 3' overhang of telomeric DNA. Intermolecular G-quadruplexes may have less biological significance, but are likely to play a role *in vivo* in phenomena such as recombination and end to end pairing. In addition to the strong interest in the structure and stability of G-quadruplexes, there is also the need to explore the role of the guanine and other neighbouring residues in the G-quadruplex structure and stability and how adjacent residues are perceived in molecular recognition by proteins and small ligands. Thus, we have carried out a careful NMR-molecular dynamics approach to determine the details of the quadruplex structure d(TTAGGGT)₄, which contains the human telomeric repeat TTAGGG with an extra T at the 3' end of the sequence. This extra thymine residue prevents the possible formation of multiple conformations and aggregates in solution through end to end stacking between G-tetrads of different molecules. This study presents the experimental evidence for formation of the parallel quadruplex structure d(TTAGGGT)₄ and explores the stability and dynamics of the structure by NMR and MD studies. Such information from this quadruplex structure is valuable in order to understand the molecular basis of ligand recognition.

5.2 Materials and methods

5.2.1 DNA samples

The oligonucleotide d(TTAGGGT) was synthesized and purified as described in Section 2.2.1, and shown to be > 95% pure and in the quadruplex form by ^1H NMR spectroscopy. The oligonucleotide sample was quantified spectrophotometrically using a calculated extinction coefficient at 260 nm determined by the nearest-neighbour method (27). The NMR sample of d(TTAGGGT) was prepared at a concentration of single strand of 6.4 mM (or 1.6 mM in quadruplex) in 0.6 ml in 90% H_2O , 10% D_2O solution, having a final salt concentration of 100 mM KCl and 10 mM K_2HPO_4 . The quadruplex was lyophilised again and dissolved in D_2O solution (0.6 ml) for the observation of non-exchangeable proton resonances.

5.2.2 NMR experiments

Standard phase-sensitive 2D NMR pulse sequences were used to record NOESY, TOCSY and DQF-COSY spectra. One-dimensional NMR experiments were recorded over 16384 data points in t_2 for each 512 points in t_1 with spectral width of 20 ppm and a delay time of 1.5 s, at a temperature of 5–65 °C. NOESY spectra at various mixing times 300 ms, 200 ms, 150 ms, 100 ms, 70 ms, were recorded over 2048 data points in t_2 and 512 points in t_1 with 64 transients for each, with a spectral width of 10000 Hz and a delay time of 1.5 s, at a temperature of 25 °C. TOCSY spectra using a 75 ms spin lock and DQF-COSY data were recorded over 2048 data points in t_2 and 512 points in t_1 with 48 transients for each with a spectral width of 10 ppm and a delay time of 1.5 s, at a temperature of 25 °C. WATERGATE-NOESY spectra at 300 ms, 200 ms, 100 ms and 75 ms mixing times

were acquired using a sample in 90% H₂O and 10% D₂O solution over 2048 data points in t_2 and 512 points in t_1 with 64 transients for each, and employed a 20 ppm spectral width and a delay time of 1.5 s, at temperatures between 5 °C and 25 °C. Two-dimensional data were zero filled to 4k x 2k prior to Fourier transformation, apodised typically with a 90° shifted sine bell squared function in both dimensions and also treated with automatic baseline correction.

5.2.3 NMR restraints

Interproton distances of the non-exchangeable and exchangeable protons were derived from integration of their NOE cross peak volumes from NOESY experiments in 150 ms, 100 ms, 75 ms in D₂O and 200 ms and 100 ms in H₂O solutions. Distances were determined by extrapolation to zero mixing time (28), using the thymine H6–Me fixed distance (3.0 Å) for NOEs involving base protons, and the sugar H2'–H2'' fixed distance (1.85 Å) for NOEs involving only sugar protons. For well resolved non-exchangeable cross peaks the distances were given upper and lower error bounds of 15% of the calculated distance, for partially overlapped non-exchangeable cross peaks the distances were given upper and lower error bounds of 20%, while for the exchangeable cross peaks 25% upper and lower error bounds were used. NOEs that involve the guanine NH₂ protons were quite broad and the NOE restraints for these protons were defined with wider bounds, 2.5–5.0 Å. Hydrogen-bond restraints were included for atoms involved in the ideal hydrogen-bonding geometry of the G-tetrad. The heavy atom-heavy atom distance restraints for O6–N6 and N7–N7 distances were set to 2.85 ± 0.10 Å.

Distance restraints were checked for large geometrical inconsistencies by comparing visually with the distances of the energy-minimised quadruplex structure

derived from unrestrained molecular dynamics simulation using MolMol software (29). In order to distinguish ambiguities in NOE restraints between interstrand and intrastrand NOE interactions resulting from the symmetry of the quadruplex structure the energy-minimised structure was considered. Interstrand restraints were manually determined by matching their calculated distance with the distances of the initial energy-minimised structure. Those distance restraints that included violations over 0.5 Å were adjusted after restrained energy minimisation was carried out to derive a set of restraints that will be consistent with the geometrical limits of the structure. A total set of 728 restraints was used for energy minimizations and restrained molecular dynamics simulations. Tables with the lists of the NOE restraints are found in appendix 2.5.

5.2.4 Structure calculations

Energy minimisations and restrained molecular dynamics calculations were performed on an Origin 200 Silicon Graphics Server using the AMBER 6 suite (30) of programs employing the AMBER 94 force field with modifications (31) and the Particle Mesh Ewald (PME) (32) method for the treatment of long-range electrostatics. The initial co-ordinates for the starting model of the quadruplex were taken from the NMR structure of the d(TTGGGGT)₄ quadruplex (20) (Protein Data Bank accession entry number 139d), choosing randomly one of the deposited structures. The starting model of the d(TTAGGGT) quadruplex was generated by replacing the guanine nucleotides in the third position of the sequence from the 5' end with adenine nucleotides using the LEAP module of AMBER 6. The DNA molecule was solvated in a periodic TIP3 water box of approximate dimensions 60 Å x 60 Å x 60 Å, which extended to a distance of 10 Å from any solute atom and

contained 5030 water molecules. Two internal potassium ions were manually positioned in the central channel between adjacent G-quartets using standard parameters for the AMBER force field. The potassium ions were positioned equidistant from two adjacent G-quartets to allow octahedral coordination with guanine carbonyl oxygen O6 atoms. The quadruplex system was neutralised externally requiring 22 potassium ions placed at the most negative locations using Coloumbic potential terms with the LEAP module. All the potassium ions including those placed in the central channel were treated as part of the solvent.

Energy minimisations and restrained molecular dynamics were carried out using the SANDER module of AMBER 6. Calculations with SANDER were performed with a 2 fs time step, with the SHAKE algorithm (tolerance 0.00005 Å) applied to all bonds to remove bond stretching, and a 9 Å cut off to the Lennard Jones interactions. The restrained molecular dynamics were performed at 300K and a constant pressure of 1.0 atm with isotropic position scaling utilising the Berendsen algorithm for temperature coupling. Translational and rotational motions were removed every 100 fs. All calculations were carried out with the PME method using a 9 Å cut-off for direct space non-bonded calculations and a 0.00001 Ewald convergence tolerance for the inclusion of long-range electrostatics in our calculations.

The quadruplex system was allowed to equilibrate fully before the molecular dynamics calculations, following the same protocol as in section 2.2.5. Minimisation was performed with 50 steps of steepest descent and 5000 steps of conjugate gradient to first the water and counterions, with the DNA coordinates frozen, followed by a further 5000 steps on all the components of the system. Next, 10 ps unrestrained molecular dynamics were run at 100K on the water alone with the DNA and

potassium ions constrained, followed for another 10 ps to allow the potassium ions to move. In the following 5 ps of dynamics the temperature of the system was increased from 100K to 300K. The next runs, each of them of 10 ps dynamics, the DNA force constant is gradually reduced from 100 to 50, 25, 10, 5 and 2.5 kcal mol⁻¹Å⁻². The equilibration step ends with 100 ps of dynamics on the whole fully unrestrained system. The system now is fully equilibrated and NOE restraints can be applied to the quadruplex system. Distance restraints were introduced gradually on the system over the first 10 ps of 100 ps MD run with the temperature stable at 300K and PME on. All NOE restraints were introduced in the form of square well potentials with a force constant of 50 kcal mol⁻¹ Å⁻¹ for the hydrogen-bond restraints and 30 kcal mol⁻¹ Å⁻¹ for all the other NOE distance restraints. A total of 1000 ps simulation was performed under the same conditions, after which the system was energy minimized also with NOE restraints.

Calculated structures satisfied the majority of the NOE restraints from the set of 728 restraints. The average minimised structure had no restraint violation > 0.3 Å, while there were 18 restraint violations between 0.2–0.3 Å, 41 restraints between 0.1–0.2 Å and 29 restraints < 0.1 Å that contributed to a 48.22 kcal mol⁻¹ energy penalty. Snapshots of each picosecond were extracted from the whole simulation and the structures were determined to be equilibrated on the basis of RMSD analysis. Time-averaged structures were calculated with the CARNAL module of AMBER. Helicoidal structural properties, sugar pucker and backbone torsion angles have been analysed using CURVES (33). The helicoidal parameters were measured individually for each strand of the G-quadruplex and the four strands were found to have very similar values.

5.2.5 Analysis of thermodynamic parameters

The melting transition of d(TTAGGGT) was investigated by NMR. A two state transition model for the association of d(TTAGGGT) to the quadruplex form was assumed from the pattern of the peaks in 1D ^1H NMR spectra at various temperatures (278–338 K) in which the two species (single strand \rightleftharpoons quadruplex) are in reversible exchange. Concentrations of the single- and four-stranded species of the d(TTAGGGT) were determined by integration of peaks (adenine H8 proton and thymine H6 proton) associated with the two forms in the aromatic peak region of the 1D ^1H NMR spectra. The equilibrium constant for this system is expressed as:

$$K_{\text{eq}} = \frac{[\text{quad}]}{[\text{single}]^4} = \frac{f_{\text{q}} \cdot c/4}{[f_{\text{s}} \cdot c]^4} \quad \text{eq. 5.1}$$

f_{q} and f_{s} are the fractions of quadruplex and single-stranded species, respectively, and C_{t} is the total concentration of the oligonucleotide in single-stranded form. Fractions of the two species in the solution calculated from integration data included an uncertainty of 5–10%. The K_{eq} value was determined for each temperature (278–338 K), which enable us to calculate the temperature dependence of the change in free energy of association $\Delta G^\circ = -RT \ln(K_{\text{eq}})$ (eq. 5.2). The linearity of the temperature dependence of the ΔG° confirms the two-state transition model for the association of d(TTAGGGT) and enables us to calculate the change in enthalpy ΔH° and entropy ΔS° at 298 K by linear fitting of the equation $\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ$ (eq. 5.3), with Kaleidagraph Software (Synergy, Inc.). Enthalpy and entropy values were calculated from the slope and y-intercept of the plot of ΔG° versus T.

5.3 Results and Discussion

5.3.1 NMR spectral features

The 1D ^1H -NMR spectra (6–12 ppm) of the $\text{d}(\text{T TAGGGT})_4$ in K^+ , H_2O solution, showing the exchangeable imino and non-exchangeable base protons at different temperatures are plotted in figure 5.3. Three distinct exchangeable imino protons are observed in the region of 10–11.5 ppm, two exchangeable amino protons are observed at 9.3 and 7.1 ppm, and non-exchangeable base protons are detected between 7.0–8.5 ppm. The exchangeable imino protons persist up to very high temperatures indicating high stability, as observed from the NMR temperature dependence study in figure 5.3. The imino protons also appear in the spectrum after dilution of the sample in D_2O solution, especially the imino proton that resonates at 10.5 ppm, showing slow exchange with the solvent. High melting temperature and low solvent accessibility are characteristic features of the guanine imino protons that form G-quadruplex structure, as a number of previous NMR studies have established (22,26,34,35). The NMR spectra in the base region at low temperatures show two sets of resonances, with one set corresponding to a major conformation, the quadruplex conformation. Increase of the temperature generates two distinct sets of resonances as observed from clear duplication of the base protons, which are replaced with a single set of resonances at very high temperature, figure 5.3. The former indicates the melting of the quadruplex conformation, seen at low temperature, into the single strand $\text{d}(\text{T TAGGGT})$. In addition, NMR spectra of $\text{d}(\text{T TAGGGT})$ at low temperature show base protons of the single strand in slow equilibrium with the quadruplex structure. The number of peaks in the base region shows that the intermolecular G-quadruplex structure formed by $\text{d}(\text{T TAGGGT})$ is

highly symmetrical and the four strands are equivalent, thus there is only one resonance observed for each base and sugar proton, figure 5.3.

The melting temperature T_m of the quadruplex $d(\text{T TAGGGT})_4$ (single stranded and quadruplex equally populated) is $\sim 60^\circ\text{C}$ at a strand concentration of 6.4 mM. The melting temperature T_m of quadruplex structures is concentration dependent (38) and increases with increases of strand concentration, however a similar T_m has been found from NMR melting studies for the quadruplex $d(\text{AGGGT})_4$ at similar concentration (26). The presence of resonances from both the single- and the four-stranded forms of the $d(\text{T TAGGGT})_4$ that are detected in the same spectrum, enables us determine the change in the population of the folded and unfolded states of the quadruplex $d(\text{T TAGGGT})_4$ as a function of temperature. The aromatic peaks of the A3 and T7 residues were used for calculating the populations. The uncertainty of integration data in the temperature range 278–293 K is greater due to signal broadening; however, the results of the fitting are very similar whether these data are taken into account or not. The equilibrium constant K_{eq} (eq. 5.1) and free energy of association ΔG° (eq. 5.2) are calculated for each temperature and then the relationship of the change in ΔG° with temperature T is plotted, figure 5.4. The linear relationship of the ΔG° with T is consistent with ΔH and ΔS being independent of temperature ($\Delta C_p=0$); fitting the data to eq. 5.3 can provide a full thermodynamic characterisation of the quadruplex folding. Thus, at 298 K we determined the following thermodynamic profile for the formation of $d(\text{T TAGGGT})_4$: $\Delta H^\circ = -256 \text{ KJmol}^{-1}$, $\Delta S^\circ = -763 \text{ JK}^{-1}\text{mol}^{-1}$, $\Delta G^\circ = -28.9 \text{ KJmol}^{-1}$. In terms of G-tetrad formation (table 5.1) these values agree very well with previous calorimetric experiments with the parallel four-stranded quadruplex of $d(\text{TGGGT})_4$ in K^+ solution (38).

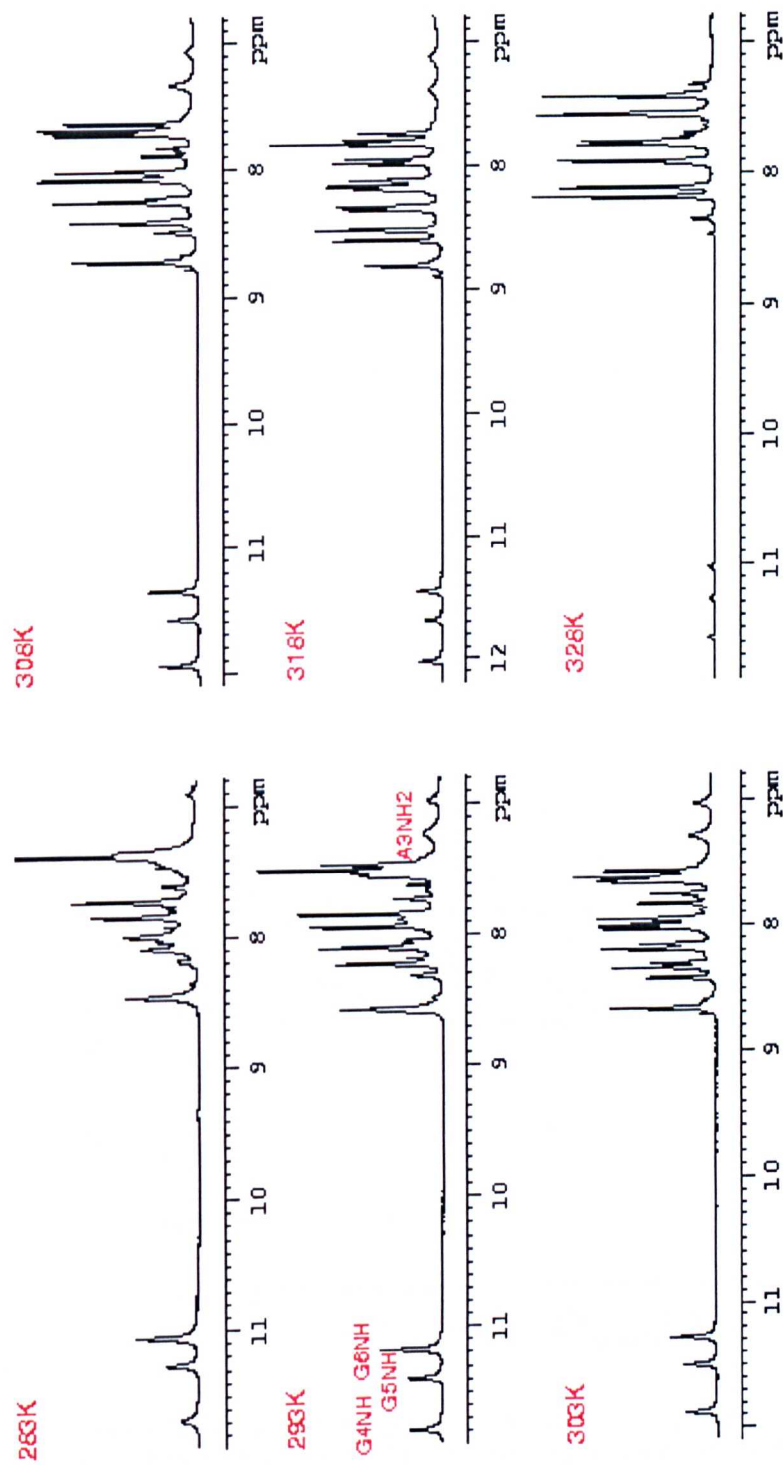


Figure 5.3 ^1H -NMR spectra (6–12 ppm) of the $d(\text{TTAGGGT})_4$ in 100 mM K^+ in H_2O solution at various temperatures. Guanine imino and adenine amino proton resonances are assigned on the spectrum at 293 K.

The resulting values establish that at 298 K the intermolecular folding of d(TTAGGGT)₄ to quadruplex formation is enthalpy driven with a negative entropic factor, resulting from the cost of ordering the flexible backbone and in bringing four strands together in the complex. The data are consistent with the majority of the enthalpic driving force being derived from G-quadruplex formation (stacking and K⁺ complexation) with a small contribution from adjacent nucleotides (A and T).

Table 5.1 Thermodynamic parameters per G-tetrad of the d(TTAGGGT)₄, d(TGGGT)₄ quadruplexes determined at 298 K. *Data from (38).

	ΔH° KJmol ⁻¹	ΔS° JK ⁻¹ mol ⁻¹	ΔG° KJmol ⁻¹
d(TTAGGGT)	-85.3	-256	-9.6
d(TGGGT)*	-87.8	-259	-9.6

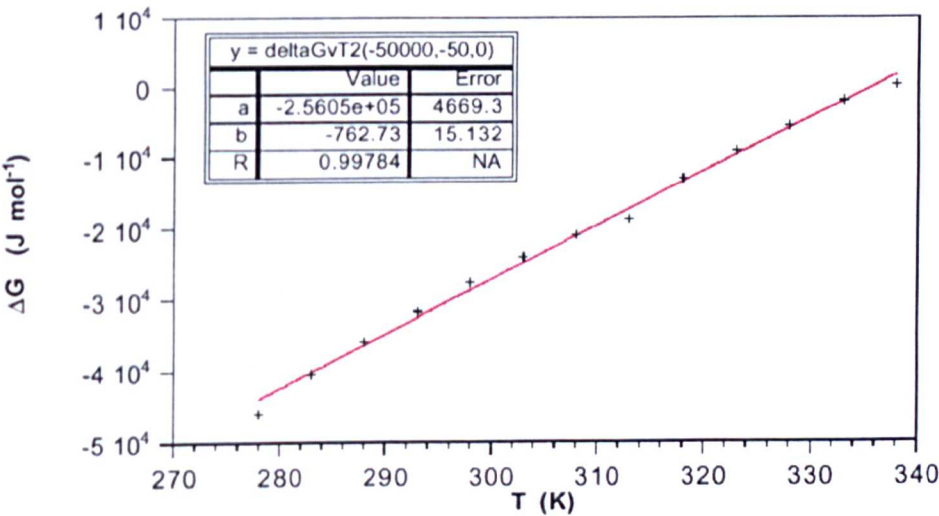


Figure 5.4 Temperature dependence of the free energy of association ΔG° determined from the relative populations of the folded and unfolded states by ¹H NMR peak intensities. Linear fitting of the data to eq 5.3 resulted in a correlation coefficient R = 0.99.

5.3.2 Exchangeable protons assignments

Resonance assignments of the exchangeable and non-exchangeable protons were based on the NOE information from NOESY spectra (figure 5.5) at various mixing times 70–400 ms and on through-bond scalar coupling interactions observed in TOCSY and DQF-COSY spectra. The chemical shifts of all protons of d(TTAGGGT) were obtained readily following standard assignment procedures for nucleic acids protons, see appendix 1.6. For assignment purposes nucleotides are numbered sequentially from 5' to 3' end, as d(T1–T2–A3–G4–G5–G6–T7).

Expanded regions A, B and C of the NOESY spectrum of d(TTAGGGT)₄ in H₂O solution, recorded at 300 ms mixing time and 293 K, are plotted in figure 5.6. NOEs between adjacent guanine imino protons in the sequence, G4NH–G5NH and G5NH–G6NH, are detected in figure 5.6A. In addition, guanine imino protons show NOEs with their own base protons (a–b cross peaks) and also to their 5' flanking base protons (c–i cross peaks) in the A3–G4–G5–G6 part of the d(TTAGGGT)₄ quadruplex structure, figure 5.6B. Thus, the 11.0 ppm imino proton of G6 exhibits NOEs with the base protons of G6H8, G5H8 and the 10.7 ppm imino proton of the G5 exhibits NOEs with the base protons of G5H8 and G4H8. We also observe NOEs between the base protons A3H8, A3H2, G4H8 and the imino proton of G4 at 11.0 ppm. These NOEs are evidence of the interstrand interactions between the guanine residues that are involved in G-tetrad formation, as they cannot be accounted for by NOEs within the same strand. In addition, such NOEs provide evidence of good stacking interactions between adjacent G-tetrads formed by G6, G5, G4 residues and also show stacking interactions of the G4-tetrad with the A3 residues within the d(TTAGGGT)₄ quadruplex structure.

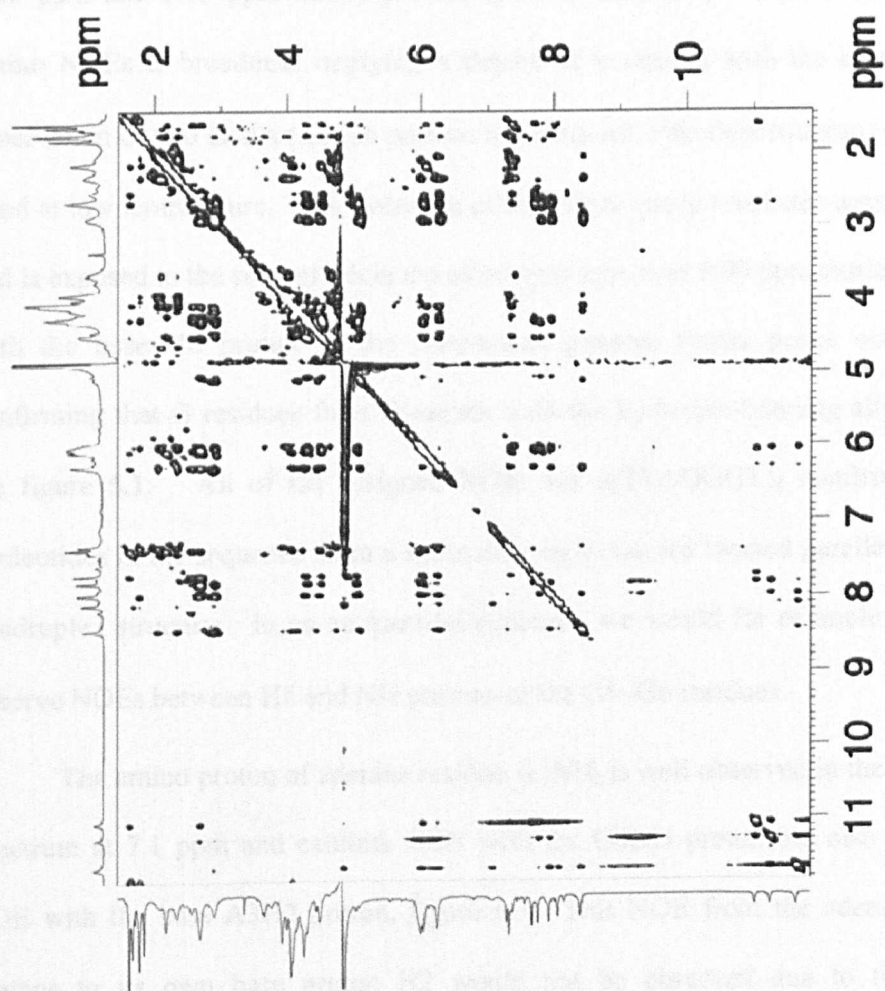
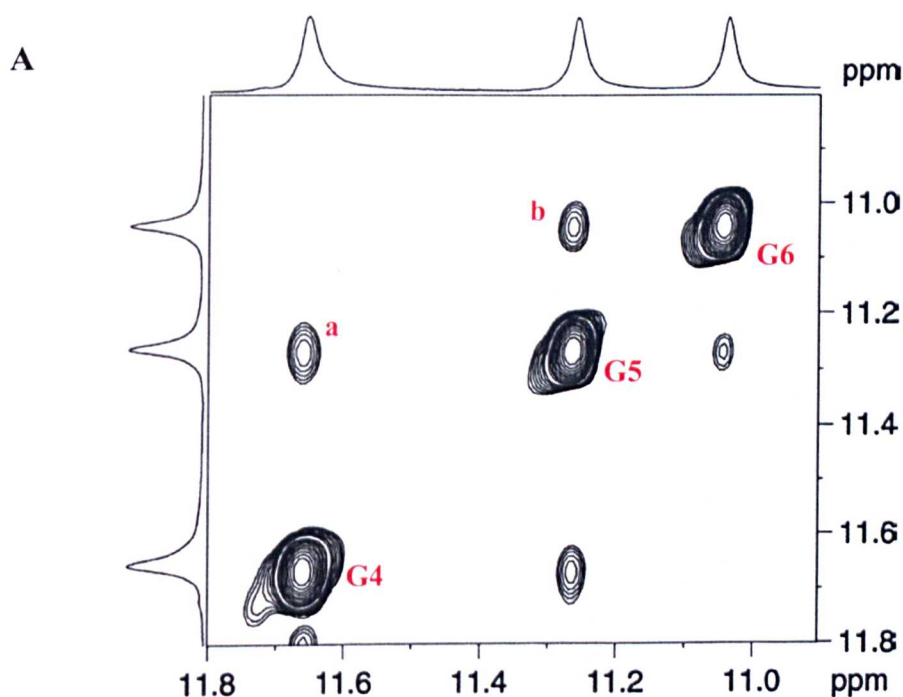


Figure 5.5 NOESY spectrum of d(TTAGGGT)₄ in H₂O solution, recorded at 300 ms mixing time and 293 K.

The expanded region of the NOESY spectrum in figure 5.6C includes NOEs between each guanine imino proton and its own guanine amino protons that resonate at two distinct chemical shifts. The G4 imino proton shows NOEs with its 9.15 ppm and 5.70 ppm amino protons (j, k cross peaks), the G5 imino proton with its 9.10 ppm and 5.90 ppm amino protons (l, m cross peaks) and the G6 imino proton with its 9.00 ppm and 5.75 ppm amino protons (n, o cross peaks). Characteristic of the amino NOEs is broadness implying a degree of exchange with the solvent. The observation of two distinct amino protons is consistent with slow rotation of the C–N bond at low temperature. One hydrogen of the amino group resonates near 6.00 ppm and is exposed to the solvent while the other hydrogen near 9.00 ppm exhibits a NOE with the base H8 proton of the inter-tetrad guanine (cross peaks not shown), confirming that G residues form G-tetrads with the hydrogen-bonding alignment in the figure 5.1. All of the assigned NOEs for d(TTAGGGT)₄ confirm that the nucleotides in the sequence form a symmetric right-handed twisted parallel-stranded quadruplex structure. In an antiparallel structure, we would for example expect to observe NOEs between H8 and NH protons of the G4–G6 residues.

The amino proton of adenine residue A3NH₂ is well observed in the 1D NMR spectrum at 7.1 ppm and exhibits NOE with the G4NH proton but also shows an NOE with the base A3H2 proton, figure 5.7. This NOE from the adenine amino protons to its own base proton H2 would not be observed due to their large separation unless it is attributed to an intertetrad interaction between A residues. In addition, the persistence of the ANH₂ resonance at high temperature, figure 5.3, similar to the guanine imino protons, shows evidence of higher stability than could arise from the formation of stable A–A interactions in a A-tetrad structure, figure 5.8. Similar findings involving adenine nucleotides are reported for the d(AGGGT)₄

quadruplex structure, but surprisingly not with the d(TAGGGT)₄ structure (26). The thymine imino protons exchange fast with the solvent and we do not observe NOEs between them and other protons in the NOESY spectrum of H₂O solution, even at low temperature. There is no direct evidence of hydrogen-bond pairing of the thymine residues that would suggest the formation of a hydrogen-bonded T-tetrad analogous to the G-tetrad (figure 5.1). Thus the NMR data suggests that only the purine residues of the sequence form tetrads in the quadruplex structure and the thymine residues of each strand do not have a well-defined conformation.



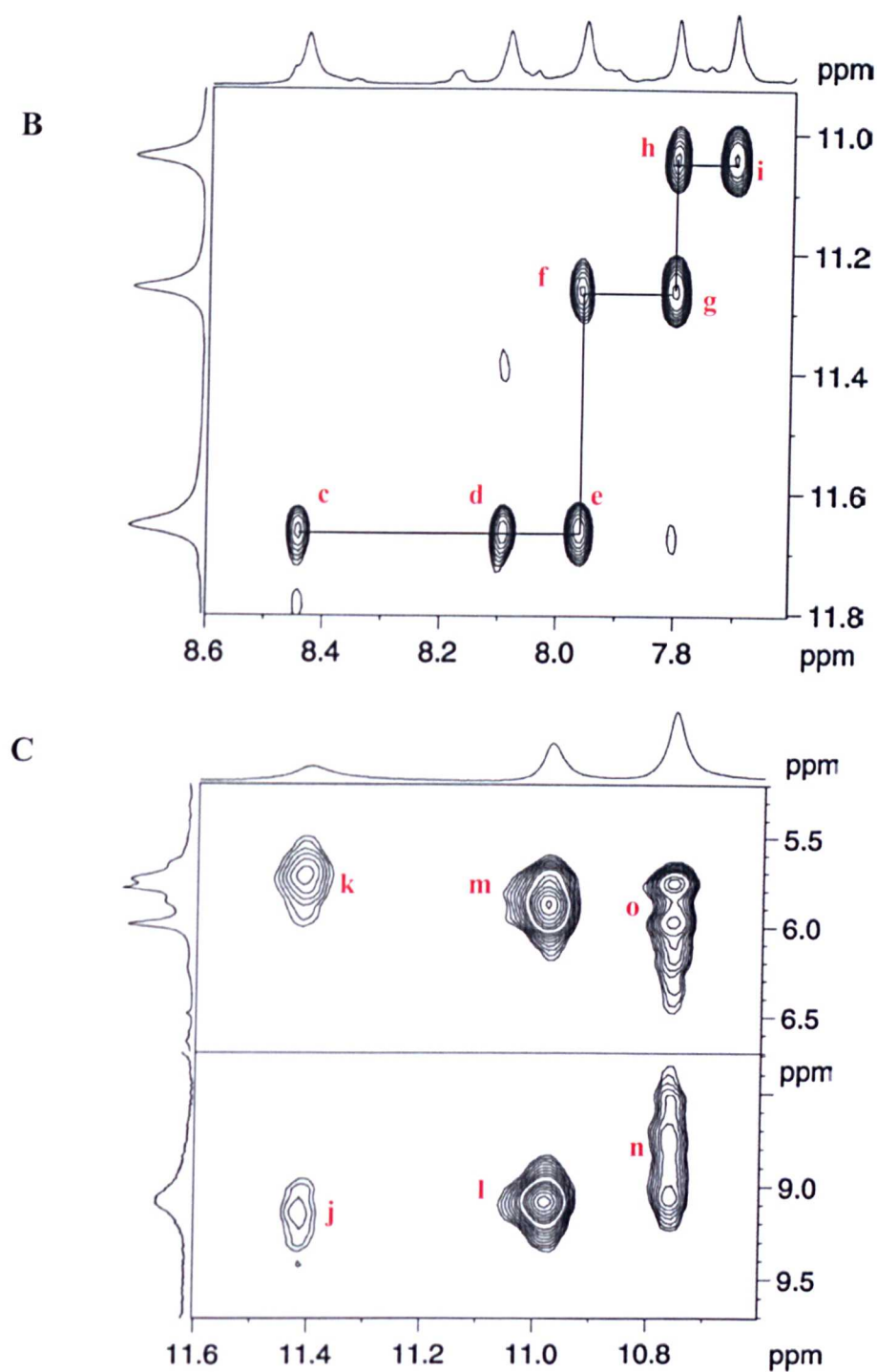


Figure 5.6 Expanded regions A, B and C of the NOESY spectrum of $d(\text{TTAGGGT})_4$ in H_2O solution, recorded at 300 ms mixing time and 293 K. NOEs labelled (a) – (o) are assigned as: (a) G4NH–G5NH (b) G4NH–G5NH (c) G4NH–A3H8 (d) G4NH–A3H2 (e) G4NH–G4H8 (f) G5NH–G4H8 (g) G5NH–G5H8 (h) G6NH–G5H8 (i) G6NH–G6H8 (j) G4NH–G4NH2 (k) G4NH–G4NH2* (l) G5NH–G5NH2 (m) G5NH–G5NH2* (n) G6NH–G6NH2 (o) G6NH–G6NH2*. Hydrogens of the amino groups exposed to the solvent are discriminated as NH2*.

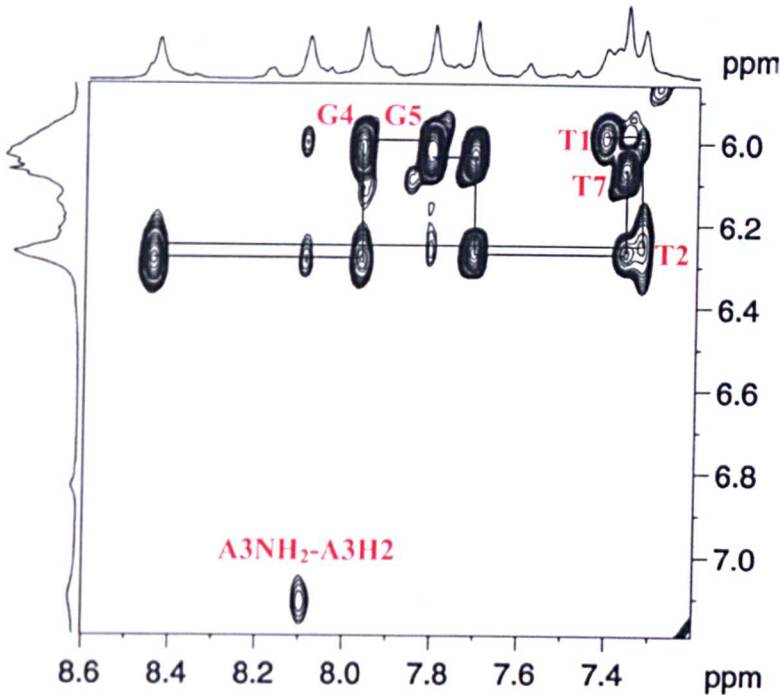


Figure 5.7 Expanded region of the NOESY spectrum of d(TTAGGGT)₄ in H₂O solution, recorded at 300 ms mixing time and 293 K, showing H8/H6–H1' NOE sequential pathway and also the NOE between ANH₂ and A3H2 protons.

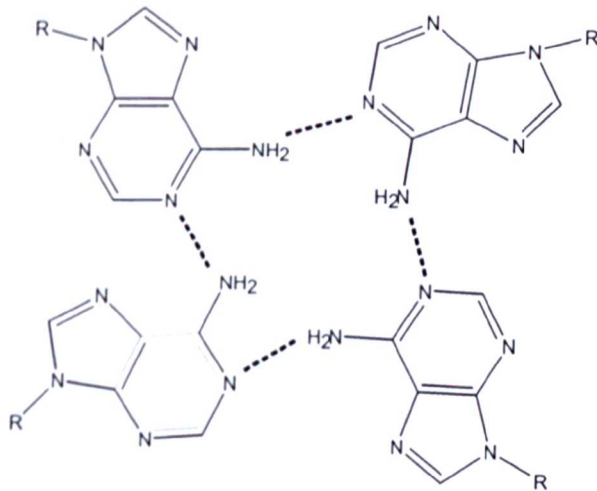


Figure 5.8 Hydrogen-bonding alignment of adenine nucleotides in a A-tetrad.

5.3.3 Non exchangeable protons assignments

Portions of the NOESY spectrum in D₂O solution recorded at 250 ms mixing time and 298 K are plotted in figure 5.7 and 5.9. Expansion of the NOESY spectrum in figure 5.7 shows the sequential pathway followed for the assignment of the sugar H1' protons and the base H8/H6 protons of each nucleotide in the d(T1-T2-A3-G4-G5-G6-T7)₄ quadruplex structure. We can trace all the sequential connectivities of each H8/H6 base proton involving its own and its 5' flanking H1' sugar protons. Expansion A of the NOESY spectrum in figure 5.9 shows NOE connectivities between the base H8/H6 protons and their own and their 5' flanking H2'/H2'' sugar protons, while expansion B in figure 5.9 shows NOE connectivities between the base H8/H6 protons and their own and their 5' flanking H3' sugar protons. The intensities of the NOE cross peaks between base H8/H6 protons to sugar protons H1', H2', H2'', H3' are similar not only for the core G4, G5, G6, but also for the T1, T2, A3 and T7 residues in the d(TTAGGGT)₄ quadruplex structure. This suggests that the nucleotides along the four strands in the symmetrical quadruplex structure are in close distance to their adjacent nucleotides on the same strand. The non-exchangeable NOE assignments show clearly that the T1, T2 and T7 nucleotides are not randomly located, even though it is unlikely that they hydrogen bond with their interstrand neighbouring nucleotides, but they are involved in stacking with their neighbouring nucleotides on the same strand and contribute to the formation and stability of the quadruplex structure. In addition, all the sequential assignment pathways observed in the NOESY spectra follow only one direction from the base to the 5' flanking sugar H1', H2'/H2'' and H3' protons, which indicates that each individual strand has the right handed helical alignment, as observed in the B-DNA helix.

Assignments of all the sugar spin systems H1', H2', H2'', H3', H4', H5'/H5'' were obtained from a combination of NOESY and TOCSY spectra. The H2' and H2'' resonances were assigned stereospecifically on the basis of their NOE cross peak intensities in a short mixing time (70 ms) NOESY spectrum. In any sugar conformation the interproton distance between H1–H2' is always shorter than the H1–H2'' distance. The resonances of the H2'' for all the nucleotides in the d(TTAGGGT)₄ quadruplex appear downfield of the resonance of the H2' protons, as happens predominantly with the B-DNA helix. The coupling constants measured from DQF-COSY spectra for the H2'–H2'' cross peaks were found to be large (>13 Hz) and characteristic of the sugar conformation being predominantly S-type (36).

Interproton distances between H8 and its own H1' protons for the guanine residues G4, G5, G6 and also for the other base residues were quantified in a short mixing time NOESY and found to be weaker than the intensity of the NOE between base H6 and methyl protons CH₃ in any of the T1, T2, T7 thymine residues. All the distance restraints determined for the base H6/H8–H1' sugar protons have a distance > 3.0 Å. The distance between intra-nucleotide H6–CH₃ protons in a thymine residue is fixed to 3.0 Å where *anti* and *syn* glycosidic conformations can be differentiated by distances of 3.7 Å and 2.5 Å, respectively, between the H8 and H1' protons. (37). This indicates that all residues in the d(TAGGGT)₄ quadruplex structure adopt *anti* glycosidic torsion angle conformation. This agrees with what has been found with all the intermolecular G-quadruplex structures except d(AGGGT)₄, where the terminal A has been proposed to adopt a *syn* conformation (26). The data presented here for d(TTAGGGT)₄ is not consistent with A adopting the *syn* conformation.

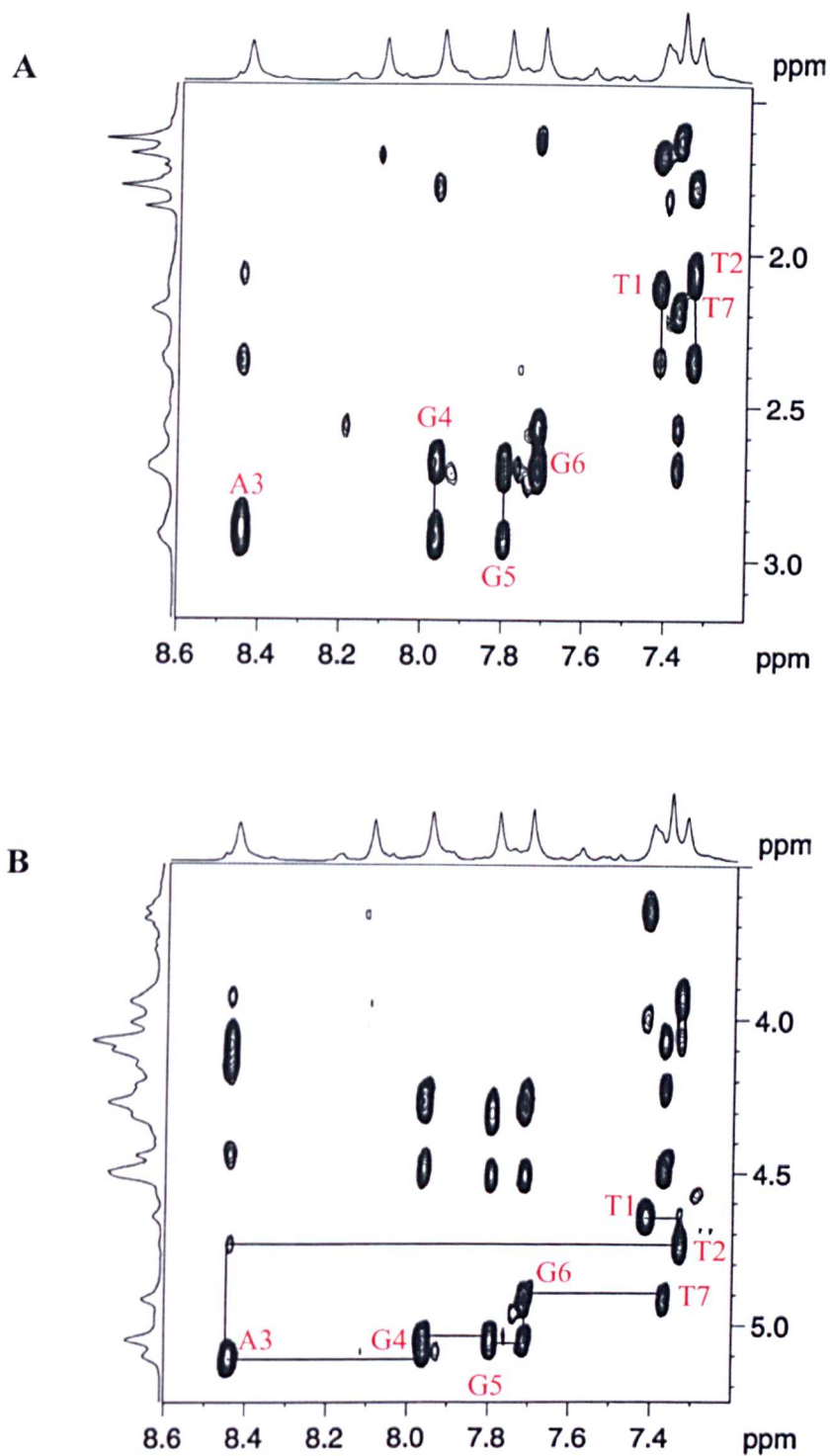


Figure 5.9 Expanded regions A and B of the NOESY spectrum of d(TTAGGGT)₄ in D₂O solution, recorded at 250 ms mixing time and 298 K, showing NOEs between H8/H6 base protons and H1', H2'/H2'' (A) and H3' (B) sugar protons.

5.3.4 NMR restraints

NOE restraints include distances between non-exchangeable protons (152 restraints), exchangeable protons (24 restraints) and in addition hydrogen-bonding restraints (24 restraints) in order to define the hydrogen-bonding alignment of guanines in a G-tetrad, figure 5.1. We did not add any hydrogen-bonding restraints between adenine residues even though NMR data indicate evidence of A-tetrad formation. In addition, there are no hydrogen-bonding restraints for the thymine residues T1, T2 and T7 in the d(TTAGGGT)₄ quadruplex structure. The interstrand NOEs in the parallel quadruplex structure have been assigned based on the initial energy-minimised structure of the quadruplex d(TAGGGT)₄. We observe interstrand NOEs only for the AGGG segment and no NOE between thymine nucleotides, which implies a better definition of the quadruplex structure in this particular region. Figure 5.10 shows the distribution of exchangeable, non-exchangeable and hydrogen-bonding restraints for each residue within one strand. There is almost the same number of restraints of non-exchangeable protons for all the seven residues while exchangeable restraints are observed only for guanine residues and adenine residues. There are 178 NOEs in total per strand, excluding hydrogen-bond restraints, which were quadrupled during structure calculations due to the four-fold symmetry of the quadruplex.

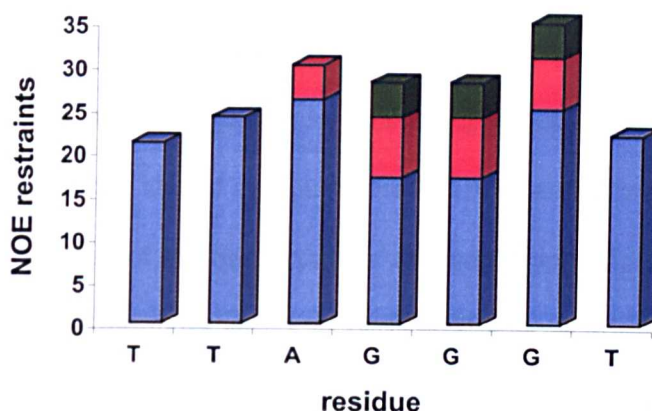


Figure 5.10 Histogram for the NOE restraints associated with each residue in the d(TTAGGGT) sequence. The blue bar represents the non-exchangeable NOE restraints, the red bar the exchangeable NOE restraints and last the green bar the hydrogen-bond restraints.

5.3.5 Restrained molecular dynamics calculations

Our structure calculations were based on a starting model of the d(TTAGGGT)₄ quadruplex, generated from a deposited structure of the d(TTGGGGT)₄ quadruplex with appropriate modifications using the LEAP module of AMBER 6. The starting structure is consistent with the NMR data in a qualitative form. There is a parallel alignment of the four strands in the quadruplex and each strand possesses a right-handed twisted helical alignment similar to B-DNA helix, as indicated from the pattern of the NOE interactions across the whole sequence. The nucleotides in each strand are aligned with the potential for hydrogen-bonding interactions within individual tetrads with the glycosidic torsion angle in an *anti* conformation. Distances calculated for the base H8/H6 protons with the sugar H1' proton for each nucleotide from NOESY spectra at short mixing time suggest that the glycosidic torsion angle of each nucleotide adopts an *anti* conformation.

The vast majority of NMR approaches to quadruplex structure determination have used restrained molecular dynamics *in vacuo* and in the absence of counterions,

except for an NMR-MD simulation study of the hairpin quadruplex structure $d[(G_3T_4G_3)]_2$ that involved explicit water molecules and counterions external and internal to the structure (37). A detailed approach to the determination of a quadruplex structure should not ignore these factors as it is well established that the local conformation of the quadruplex is dependent on the presence of counterions and hydration of the molecule. Our NMR-MD approach included explicit solvent molecules and potassium ions both external to the solute and bound internally within the quadruplex structure. We used the particle mesh ewald method for accurate modelling of long-range electrostatics and for obtaining stable trajectories under these conditions. Although the exact location of the potassium ions cannot be determined from our experimental data, we positioned two ions within the three stacked G-tetrads as defined by the X-ray study of the $d(TGGGGT)$ (22).

Our protocol for the NMR-MD calculations initially included a careful equilibration of the initial structure and then a slow heating to target temperature 300 K, while gradually allowing the DNA co-ordinates to move free and unrestrained. Then NMR restraints are introduced to the equilibrated structure and the system is allowed to undergo 1 ns restrained molecular dynamics to satisfy the NMR restraints and search adequate conformational space under the influence of water molecules and cations. The NMR restraints are satisfied well within the 1 ns simulation and there are only small violations, associated mainly with the terminal residues. The NMR-MD calculations were judged to equilibrate and converge well within the 1 ns of the run as shown from RMSD analysis in figure 5.11. The last 500 ps of the simulation were selected for structural analysis. The pairwise RMSD from the mean structure over all heavy atoms over the final 500 ps of the simulation is 1.65 (± 0.22) Å.

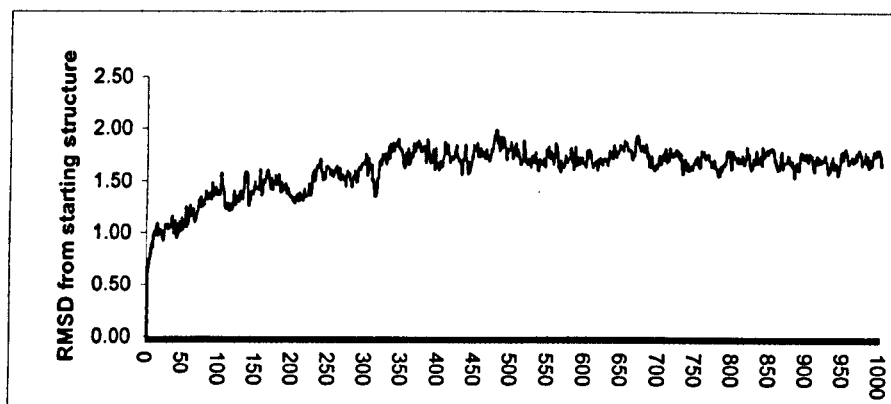


Figure 5.11 RMSD analysis of 1000 snapshots of the d(TTAGGGT)₄ structure taken from the 1 ns restrained MD simulation.

5.3.6 Quadruplex structure

The structure of the quadruplex d(TTAGGGT)₄ in figure 5.12 and 5.13 represents the average minimised structure calculated from the last 500 ps of the restrained MD simulation. The right-handed parallel quadruplex structure shows the guanine residues to have an almost planar conformation and all form a well-defined G-tetrad conformation. Although the NMR refinement did not use any hydrogen-bond restraints between the adenine residues on adjacent strands, the structure shows clearly that there is a formation of a certain planar conformation of the adenine residues that allows them to hydrogen bond with each other. Each of the thymine bases of the structure has a buckled conformation and do not form any specific hydrogen-bonding alignment. These general features of the structure agree with our NMR evidence and emphasize that the purine residues in the structure are able to form tetrads, stack on top of each other and stabilize the quadruplex structure, while the thymine residues have a more flexible conformation. The latter point is also emphasised from figure 5.14, which shows five randomly chosen structures

superimposed. The loose conformation of the thymine residues reflects the lack of interstrand NOEs between thymine nucleotides.

The core quadruplex structure in figure 5.15 from the average minimised structure of d(TTAGGGT)₄, consisting of the purine bases of the AGGG core, has been further analysed in terms of hydrogen bonding and helical parameters. The average distances, with standard deviations between atoms involved in formation of hydrogen bonds from the whole set of 500 structures, are shown in table 5.3. Distances of the Hoogsteen N1–O6 and N2–N7 hydrogen bonds between the guanine bases of the three G-tetrads show that all hydrogen bonds are formed and maintained in all the structures. In addition, there is no significant difference in hydrogen-bond distances between the G-tetrads. The hydrogen bonds formed by the N1–N6 atoms of the adenine bases have a little more flexibility, but they still are well defined in all calculated structures, though no specific H-bond restraints were included. Our calculations for hydrogen-bond distances between O4–N3 and O2–N3 atoms found in thymine bases confirmed that thymines do not form any recognisable hydrogen-bonding alignment in an analogous fashion to the purine residues in the sequence.

It is worth noting that during the course of unrestrained molecular dynamics of the quadruplex d(TTAGGGT)₄ structure hydrogen bonds between the guanine and adenine residues are also stable. Analogous analysis of the unrestrained simulation showed that distances of the Hoogsteen N1–O6 and N2–N7 hydrogen bonds within G-tetrads have an average value of 3.00 ± 0.15 Å, while the N1–N7 hydrogen bonds have an average distance value of 3.12 ± 0.20 Å. These results emphasise that stable G-tetrads and A-tetrad is a characteristic of the quadruplex d(TTAGGGT)₄ structure and not an effect of the hydrogen-bond restraints.

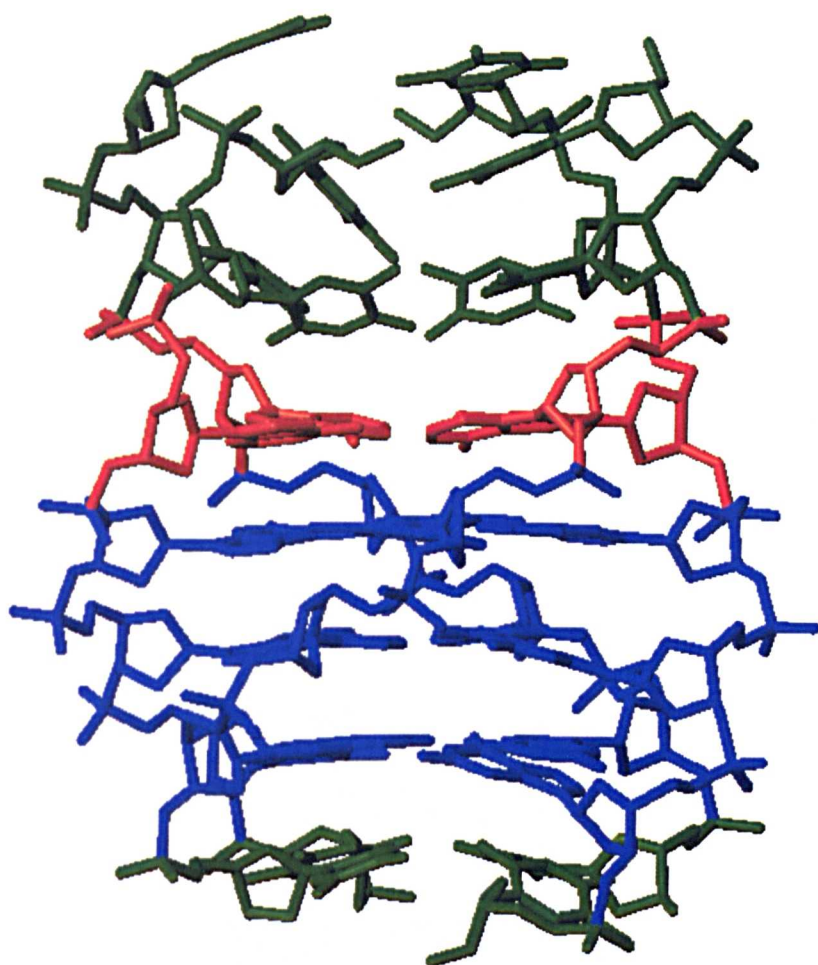


Figure 5.12 Average minimised structure of d(TTAGGGT)₄ from the last 500 snapshots (1 per ps) of the 1 ns restrained simulation. Guanines are coloured in blue, adenines in dark orange and thymines in dark green.

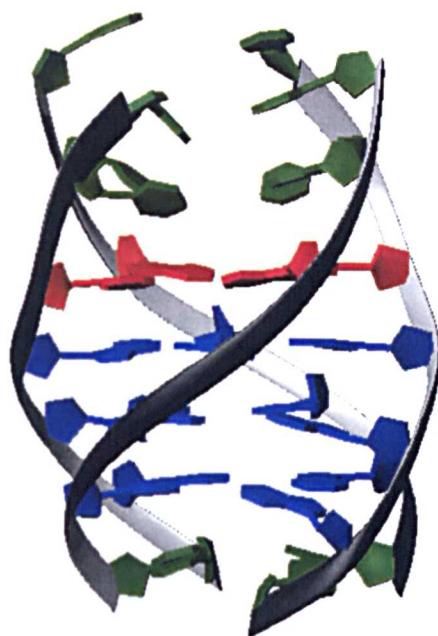


Figure 5.13 Schematic representation of the average minimised structure of d(TTAGGGT)₄ showing the backbone as ribbons and bases as planes.

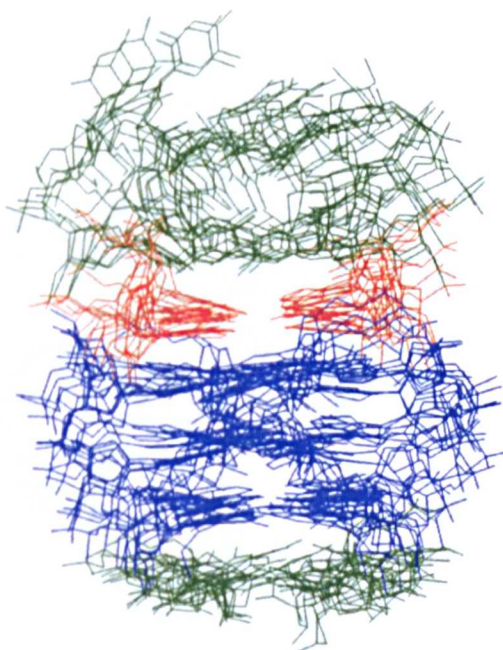


Figure 5.14 Five randomly chosen structures of d(TTAGGGT)₄ from the last 500 snapshots (500 ps) of the 1 ns restrained simulation. Guanines are coloured in blue, adenines in dark orange and thymine in dark green.

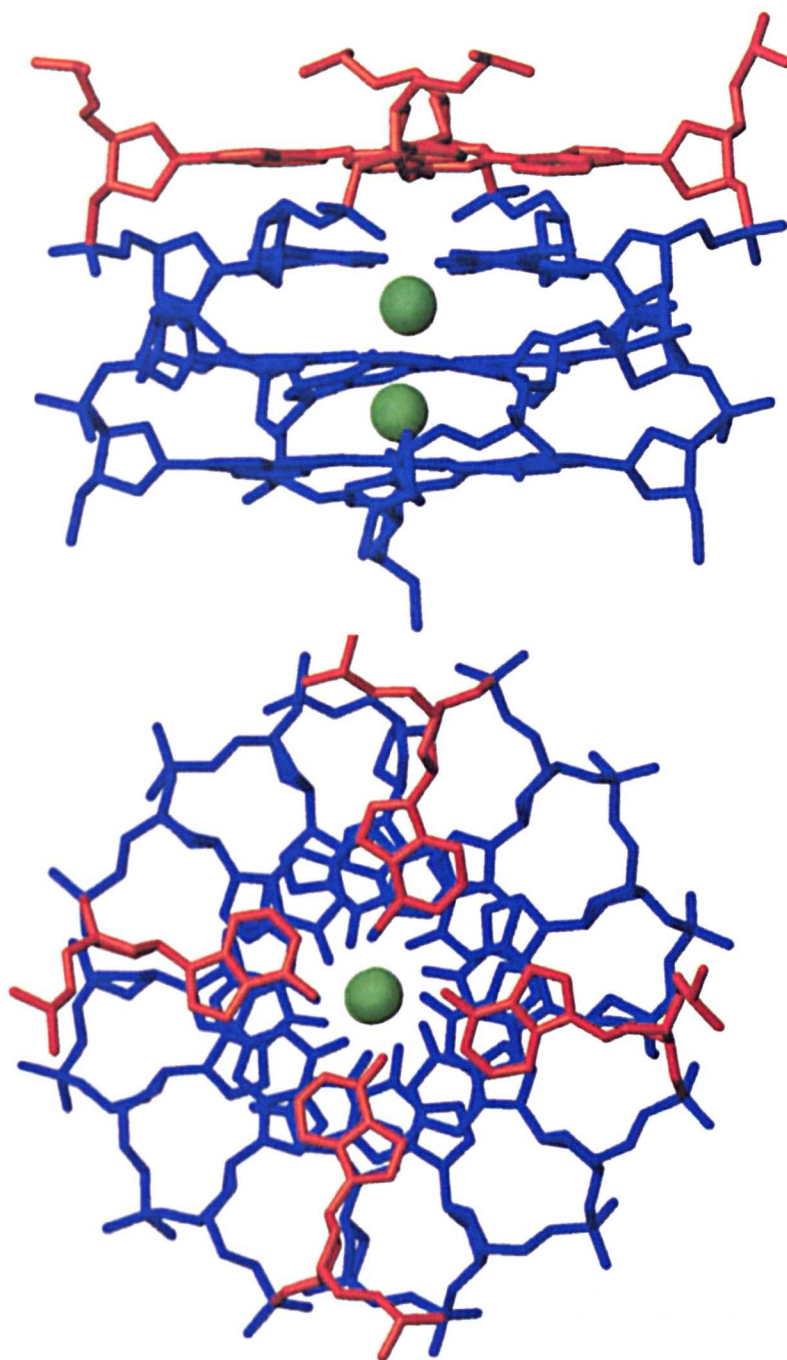


Figure 5.15 Core quadruplex (AGGG)₄ from the average minimised structure of d(TTAGGGT)₄. Guanines are coloured in blue, adenines in dark orange and K⁺ ions lie between of the G-tetrads.

Table 5.3 Average hydrogen-bond distance and standard deviation (parentheses) for core quadruplex (AGGG)₄.

Residue	N7–N2 H-bond	N1–O6 H-bond	N1–N6 H-bond
G4	2.94 (0.07)	2.89 (0.07)	
G5	2.96 (0.06)	2.89 (0.07)	
G6	2.97 (0.06)	2.87 (0.07)	
A3			3.12 (0.16)

It is also evident from the figure 5.15 that the core quadruplex forms a cavity through the centre of its structure, which is maintained in all the calculated structures. Inside this cavity and between two neighbouring G-tetrads the potassium ions are located almost intact from where they were initially positioned. Although the exact number of potassium ions in the quadruplex structure is unanswered, the present findings show that the potassium ions play a role in the stabilisation of the local conformation of the guanine residues in quadruplex structures. Titration studies by others have established the K⁺ binding stoichiometry, consistent with this model (41).

5.3.7 Distinctive features of d(TTAGGGT)₄ structure

Variation of the helicoidal parameters, table 5.4, for the quadruplex d(TTAGGGT)₄ has been calculated over the last 500 structures (500 ps) of the restrained MD simulation using the CURVES program (33). The helicoidal parameters have been analysed for each individual strand and each strand found to have very similar values. These include the glycosidic torsion angle χ and the pseudorotation angle P of all the nucleotides in the sequence and also helical twist and axial rise at the A3–G4, G4–G5 and G5–G6 steps. The glycosidic torsion angles

χ , as was suggested from the NMR data, are clearly in the *anti* conformation for all the nucleotides including A3 (centered on -121°). This value agrees very well with the previous reported value of -115° of a similar quadruplex structure d(TTGGGGT)₄ (20). However, this contrasts with the data of Hosur *et al.* (26) for d(AGGGT)₄, where the adenine was shown to be in the *syn* conformation on the basis of quantitative NOE calculations. Thus, additional thymines on the 5' end appear to have a significant effect in stabilizing and modulating nucleotide conformation.

Table 5.4 Average value and standard deviation (parentheses) over all four strands of helical twist, axial rise for the (AGGG)₄ segment and glycosidic torsion angle χ , pseudorotation phase angle P of d(TTAGGGT)₄.

	Twist ($^\circ$)		Rise (\AA)	
A3-G4	25	(3)	3.0	(0.3)
G4-G5	29	(2)	3.5	(0.2)
G5-G6	26	(2)	3.7	(0.2)
	χ angle ($^\circ$)		Phase P ($^\circ$)	
T1	-146	(17)	140	(19)
T2	-136	(14)	100	(14)
A3	-119	(10)	133	(21)
G4	-117	(9)	134	(10)
G5	-111	(9)	143	(12)
G6	-115	(10)	145	(9)
T7	-109	(15)	127	(24)

The pseudorotation angle P is in the range of the S type sugar conformation for all the residues with a mean value of 132° , although some variations amongst residues are observed. This value is rather smaller than mean values observed for the canonical B-DNA conformation. The purine bases of the AGGG segment contribute

to axial rise values of 3.0 Å, 3.5 Å, 3.7 Å and helical twist of 25°, 29°, 26° for the A3–G4, G4–G5, G5–G6 base steps, respectively. The helical twist for the purine steps is smaller than the average value of 36° for canonical B–DNA conformation, but is in the range of helical twist 25–30° of guanine residues forming G-tetrads (20,22,25). The twist angle is largest at the central G4–G5 step but decreases at the neighbouring A3–G4 and G5–G6 steps. This sequence-dependent variation is likely to be an indication of the differences in base stacking patterns between purine residues on adjacent tetrads formed by the AGGG segment.

Figure 5.16 shows the base stacking of the tetrads formed by (a) A3–G4, (b) G4–G5, (c) G5–G6 steps and the conformation of the A3-tetrad (d) in the average minimised structure. Generally base stacking is observed between rings of purine residues on the same strand, as in B–DNA structure. This stacking pattern is usually observed in parallel G-quadruplex structures (20,22,26). The purine nucleotides in the structure adopt a hydrogen-bond alignment, which involves hydrogen-bond donors available to acceptors of the adjacent nucleotide in the same tetrad. The O6 atoms of the carboxyl groups of guanine nucleotides point towards the cavity of the structure and are assumed to coordinate with the potassium ions. The base stacking for the guanine residues in G4–G5 and G5–G6 is similar, with the imino and amino protons of the G4 and G5 nucleotides stacked on the six-membered ring of the G5 and G6 nucleotides, respectively. The guanine nucleotides of the G5-tetrad show the best overlap with the G4- and G6-tetrads and this may explain why the G5NH resonance exchanges more slowly than the imino protons of the G4 and G6 nucleotides when dissolved in D₂O solution.

Our NMR data highlight a number of additional NOEs, particularly from G4NH to A3H8 and A3H2, that indicate good stacking interactions between the two sets of purine bases. In addition the amino proton of A3 shows similar stability to the guanine imino protons and exhibits NOEs with G4NH and A3H2. The A3NH₂-A3H2 NOE is not expected to arise as an intranucleotide NOE but is consistent with an interaction between adenine bases within an A-tetrad conformation. The formation of an A-tetrad has been reported only for d(AGGGT)₄ (26), but with different conformational properties. Hosur *et al* (26) suggested that adenine nucleotides have two distinct patterns of A-tetrad alignment distinguished by the NOEs between ANH₂-AH2 and ANH₂-AH8, respectively. Observation of these NOEs simultaneously suggests a dynamic behaviour of the A-tetrad exchanging between the two hydrogen-bonding patterns. The dynamism between the two alignments was also evident in their structure calculations, resulting in two different base stacking patterns with the adjacent G-tetrad.

On the contrary, in our quadruplex structure d(TTAGGGT)₄ there is only one strong NOE between the A3NH₂-A3H2 that can force the adenine nucleotides to a single tetrad alignment as shown in figure 5.16d. The adenine nucleotides are stabilised in this A-tetrad alignment by hydrogen bonds between the NH₂-N1 atoms, figure 5.8. The C6-NH₂ of the adenines point into the central cavity of the quadruplex structure in a similar way to that observed for the O6 atoms of the guanines. The base stacking between the A3-G4 step shows partial stacking of the six-membered ring of the adenine with the five-membered ring of the guanine nucleotide of the same strand. The amino protons A3NH₂ stack on top of the imino proton G4NH from the adjacent strand. The base stacking pattern for A3-G4 step in the present structure is also different from that reported for d(AGGGT)₄ (26)

quadruplex. This is justified from the difference in glycosidic torsion angles of the adenine residues in these two structures. We can not explain why adenine nucleotides have a *syn* conformation in d(AGGGT)₄ (26) and *anti* conformation in d(TAGGGT)₄ (26) and d(TTAGGGT)₄, but this is likely an effect from the additional thymine residues on the 5' end.

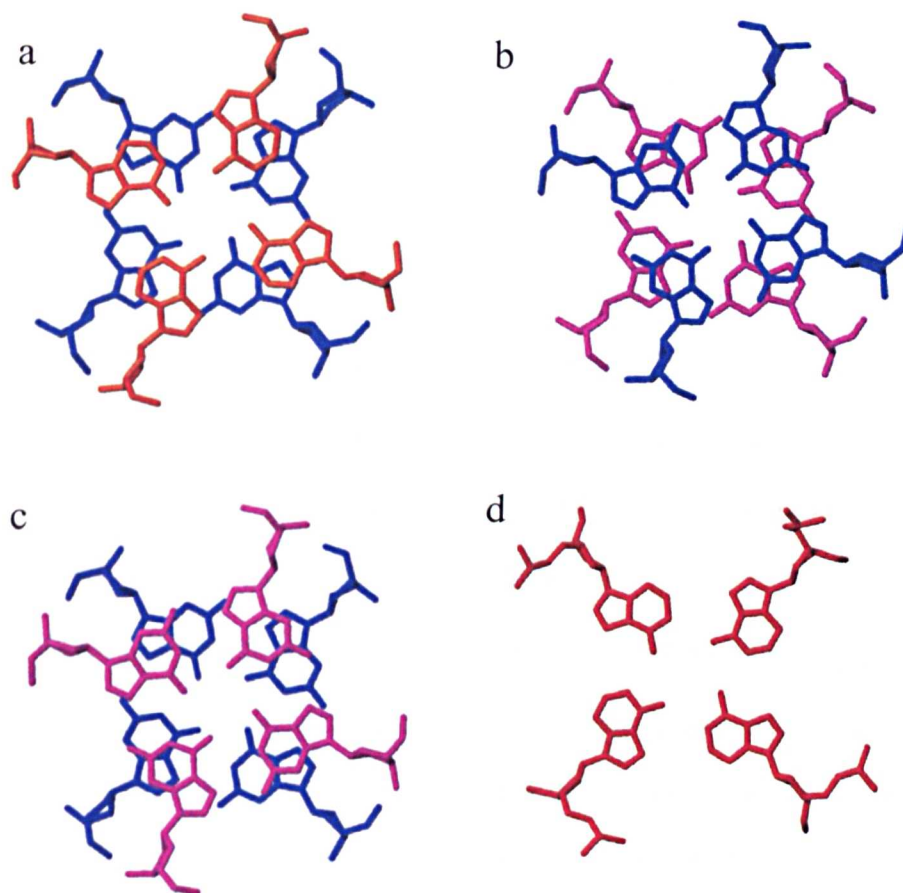


Figure 5.16 Stacking pattern between a) A3–G4, b) G4–G5, c) G6–T7 tetrads and d) A3-tetrad conformation in the average minimised structure of the d(TTAGGGT)₄.

5.4 Conclusions

We have determined the high resolution structure of the quadruplex d(TTAGGGT)₄ by a NMR restrained molecular dynamics approach that includes explicit treatment of solvent and counterions. The d(TTAGGGT)₄ quadruplex forms a stable structure mainly in the purine core due to strong stacking interactions. The

hydrogen bonding of the G-tetrads persists in all structures with minimum deviation from ideal geometries. Our structure determination supported by NMR data showed that adenine nucleotides on separate strands could also form stable hydrogen-bonded tetrads. The A-tetrad observed in our sequence d(TTAGGGT)₄ behaves differently from the one reported in d(AGGGT)₄. We observe a single preferred conformation with all adenine nucleotides adopting *anti* glycosidic torsion angles. However, the A-tetrad in the calculated structures is more flexible than G-tetrads as indicated from the hydrogen-bond distances of the adenine N6-N1 atoms. The thymine nucleotides have a much more flexible conformation and do not appear to form hydrogen-bonded T-tetrads, but do appear to stack with their adjacent nucleotides in B-like DNA conformation. We can observe a broad imino resonance at 5 °C, which could be attributed to the thymine NH though there is no evidence for hydrogen bonding. We can presume that generally pyrimidine nucleotides are unable to form stable tetrads unless they are surrounded and stabilised by purine tetrads.

Previous investigations have provided us with the NMR characteristics of sequences containing the human telomeric repeat, but structure determination of such a quadruplex has not been reported in the literature. Thus, from this point of view the high resolution structure of the d(TTAGGGT)₄ adds valuable information to the family of parallel quadruplex structures determined so far. The quadruplex structure of the d(TTAGGGT)₄ also provides us with an NMR model for ligand recognition studies. Knowledge of the characteristics of such complex structures is essential in order to understand the molecular basis of ligand recognition and examine the effect of ligand binding on such structures. On-going research has highlighted the role of ligand-quadruplex interactions in telomerase inhibition and anticancer activity. Drug-quadruplex interactions are the subject of the next chapter.

5.5 References

1. Neidle, S. 1998 *Oxford Handbook of Nucleic Acid Structure*, Oxford University Press, New York.
2. Blackburn, E. H. 1991 Structure and Function of Telomeres. *Nature* **350**, 569-573.
3. Zakian, V. A. 1995 Telomeres – Beginning to Understand the End. *Science* **270**, 1601-1607.
4. Greider, C. W. & Blackburn, E. H. 1996 Telomeres, Telomerase and Cancer. *Sci. Am.* **274**, 92-97.
5. Shippen, D. E. 1993 Telomeres and Telomerases. *Curr. Opin. Gen. Dev.* **3**, 759-763.
6. Simonsson, T., Pecinka, P. & Kubista, M. 1998 DNA Tetraplex Formation in the Control Region of c-myc. *Nucleic Acids Res.* **26**, 1167-1172.
7. Woodford, K. J., Howell, R. M. & Usdin, K. 1994 A Novel K⁺-Dependent DNA-Synthesis Arrest Site in a Commonly Occurring Sequence Motif in Eukaryotes. *J. Biol. Chem.* **269**, 27029-27035.
8. Shimizu, A. & Honjo, T. 1984 Immunoglobulin Class Switching. *Cell* **36**, 801-803.
9. Feigon, J., Koshlap, K. M. & Smith, F. W. 1995 H-1 NMR Spectroscopy of DNA Triplexes and Quadruplexes. *Nuclear Magn. Reson. Nucleic Acids* **261**, 225-255.
10. Shafer, R. H. 1998 Stability and Structure of Model DNA Triplexes and Quadruplexes and their Interactions with Small Ligands. *Prog. Nucleic Acid Res. Mol. Biol.* **59**, 55-94.
11. Patel, D. J., Bouaziz, S., Kettani, A. & Wang, Y. 1998 *Oxford Handbook of Nucleic Acid Structure*, p. 389-453, Oxford University Press, New York.
12. Walsh, K. & Gualberto, A. 1992 Myod Binds to the Guanine Tetrad Nucleic-Acid Structure. *J. Biol. Chem.* **267**, 13714-13718.
13. Erlitzki, R. & Fry, M. 1997 Sequence Specific Binding Protein of Single-Stranded and Unimolecular Quadruplex Telomeric DNA from Rat Hepatocytes. *J. Biol. Chem.* **272**, 15881-15890.

14. Sundquist, W. I. & Heaphy, S. 1993 Evidence for Interstrand Quadruplex Formation in the Dimerization of Human Immunodeficiency Virus-1 Genomic RNA. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 3393-3397.
15. Raymond, E., Soria, J. C., Izbicka, E., Boussin, F., Hurley, L. & Von Hoff, D. D. 2000 DNA G-quadruplexes, Telomere-Specific Proteins and Telomere-Associated Enzymes as Potential Targets for New Anticancer Drugs. *Invest. New Drugs* **18**, 123-137.
16. Han, H., Bennett, R. J. & Hurley, L. H. 1999 PIPER, A G4-DNA Interactive Compound, Prevents the Unwinding of G-quadruplex DNA by Yeast Sgs1 Helicase. *Clin. Cancer Res.* **5**, 607.
17. Williamson, J. R. 1993 Guanine Quartets. *Curr. Opin. Struct. Biol.* **3**, 357-362.
18. Phillips, K., Dauter, Z., Murchie, A. I. H., Lilley, D. M. J. & Luisi, B. 1997 The Crystal Structure of a Parallel-Stranded Guanine Tetraplex at 0.95 Angstrom Resolution. *J. Mol. Biol.* **273**, 171-182.
19. Hud, N. V., Sklenar, V. & Feigon, J. 1999 Localization of Ammonium Ione in the Minor Groove of DNA Duplexes in Solution and the Origin of DNA A-tract Bending. *J. Mol. Biol.* **286**, 651-660.
20. Wang, Y. & Patel, D. J. 1993 Solution Structure of a Parallel-Stranded G-Quadruplex DNA. *J. Mol. Biol.* **234**, 1171-1183.
21. Kang, C., Zhang, X. H., Ratliff, R., Moyzis, R. & Rich, A. 1992 Crystal-Structure of Four-Stranded Oxytricha Telomeric DNA. *Nature* **356**, 126-131.
22. Wang, Y. & Patel, D. J. 1993 Solution Structure of the Human Telomeric Repeat d[AG₃(T₂AG₃)₃] G-Tetraplex. *Structure* **1**, 263-282.
23. Gupta, G., Garcia, A. E., Guo, Q., Lu, M. & Kallenbach, N. R. 1993 Structure of a Parallel-Stranded Tetramer of the Oxytricha Telomeric DNA-Sequence d(T₄G₄). *Biochemistry* **32**, 7098-7103.
24. Aboulela, F., Murchie, A. I. H., Norman, D. G. & Lilley, D. M. J. 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.
25. Cheong, C. J. & Moore, P. B. 1992 Solution Structure of an Unusually Stable Rna Tetraplex Containing G-Quartet and U-Quartet Structures. *Biochemistry* **31**, 8406-8414.

26. Venczel, E. A. & Sen, D. 1993 Parallel and Antiparallel G-DNA Structures from a Complex Telomeric Sequence. *Biochemistry* **32**, 6220-6228.
27. Patel, P. K. & Hosur, R. V. 1999 NMR Observation of T-tetrads in a Parallel Stranded DNA Quadruplex Formed by *Saccharomyces Cerevisiae* Telomere Repeats. *Nucleic Acids Res.* **27**, 2457-2464.
28. Patel, P. K., Koti, A. S. R. & Hosur, R. V. 1999 NMR Studies on Truncated Sequences of Human Telomeric DNA: Observation of a Novel A-tetrad. *Nucleic Acids Res.* **27**, 3836-3843.
29. Cantor, C.R., Warshaw, M.M. & Shapiro, H. 1970 Oligonucleotide Interactions III. Circular Dichroism Studies of the Conformation of Deoxyoligonucleotides. *Biopolymers* **9**, 1059-1077.
30. Baleja, J. D., Moulton, J. & Sykes, B. D. 1990 Distance Measurement and Structure Refinement with NOE Data. *J. Magn. Res.* **87**, 375-384.
31. Koradi, R., Billeter, M. & Wuthrich, K. 1996 MOLMOL: A Program for Display and Analysis of Macromolecular Structures. *J. Mol. Graph.* **14**, 51-55.
32. Case, D.A., Pearlman, D.A., Caldwell, J.W., Cheatham III, T.E., Ross, W.S., Simmerling, C.L., Darden, T.L., Marz, K.M., Stanton, R.V., Cheng, A.L., Vincent, J.J., Crowley, M., Tsui, V., Radmer, R.J., Duan, Y., Pitera, J., Massova, I., Seibel, G.L., Dingh, U.C., Weiner, P.K. & Kollman, P.A. (1999) AMBER 6, University of California, San Francisco.
33. Cheatham, T. E., Cieplak, P. & Kollman, P. A. 1999 A Modified Version of the Cornell et al. Force field with Improved Sugar Pucker Phases and Helical Repeat. *J. Biomol. Struct. Dyn.* **16**, 845-862.
34. Cheatham, T. E., Miller, J. L., Fox, T., Darden, T. A. & Kollman, P. A. 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.
35. Lavery, R. & Sklenar, H. 1989 Defining the Structure of Irregular Nucleic-Acids – Conventions and Principles. *J. Biomol. Struct. Dyn.* **6**, 655-667.
36. Wang, Y. & Patel, D. J. 1992 Guanine Residues in d(T₂AG₃) and d(T₂G₄) Form Parallel-Stranded Potassium Cation Stabilized G-Quadruplexes with Antiglycosidic Torsion Angles in Solution. *Biochemistry* **31**, 8112-8119.
37. Wang, Y., Delossantos, C., Gao, X. L., Greene, K., Live, D. & Patel, D. J. 1991 Multinuclear Nuclear-Magnetic-Resonance Studies of Na⁺ Cation – Stabilized

Complex Formed by d(GGTTTTCGG) in Solution - Implications for G-Tetrad Structures. *J. Mol. Biol.* **222**, 819-832.

38. Wijk, V. J., Huckriede, B.D., Ippel, J.H., Altona, C. 1992 Furanose Sugar Conformations in DNA from NMR Coupling Constants *Methods Enzymol.* **211**, 286-306.
39. Strahan, G. D., Keniry, M. A. & Shafer, R. H. 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.
40. Jin, R. Z., Gaffney, B. L., Wang, C., Jones, R. A. & Breslauer, K. J. 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. U. S. A.* **89**, 8832-8836.
41. Marathias, V. M. & Bolton, P. H. 2000 Structures of the Potassium-Saturated, 2:1, and Intermediate, 1:1, Forms of a Quadruplex DNA. *Nucleic Acids Res.* **28**, 1967-1977.

6. Recognition and stabilisation of quadruplex DNA by a potent new telomerase inhibitor

6.1 Introduction

Telomeres are repetitive DNA sequences at the ends of linear chromosomes that protect the chromosome from recombination, end-to-end fusion and nuclease degradation (1,2). In human cells the telomeric DNA is typically composed of 5-15 kb of double stranded pairs of tandem repeats of guanine rich sequences TTAGGG with a single stranded 3'-end overhang necessary to ensure complete chromosomal DNA replication. With each cell division, telomeres shorten by 50-200 base pairs because synthesis of the lagging strand of DNA is unable to replicate the 3'-end overhang (3,4). Thus, when the telomeres of normal cells shorten to a critical length the cells stop growing and enter a state of senescence (5). At the stage of senescence where telomeres are critically short, end-to-end fusion and chromosomal instability leads to cell death. A cell can escape from this normal cycle and become immortal by stabilising the length of its telomeres. This happens almost always under activation of the telomerase enzyme (6).

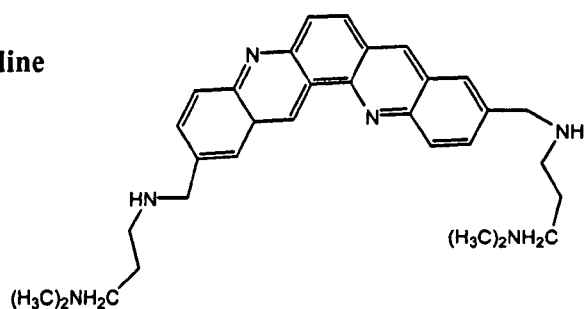
Telomerase is a ribonucleoprotein that is responsible for adding telomeric repeats TTAGGG and maintaining the length of the 3'-ends of chromosomes. The major components of telomerase are an endogenous RNA template (hRN) that contains an 11bp sequence and acts as the template for addition of telomeric repeats to the chromosome and also a reverse transcriptase component (hTERT) that catalyses the addition of telomeric repeats (7-9). All human somatic cells contain

the hRNA template but lack the hTERT that activates telomerase (8). Telomerase is active in 85-90% of all human tumours but not in adjacent normal cells (6). It has been suggested that tumour cells have unlimited proliferative potential beyond the limits of the senescence cell state by activating the telomerase mechanism. Consequently, telomerase has become a high profile target for the development of novel anti-cancer agents (10-12).

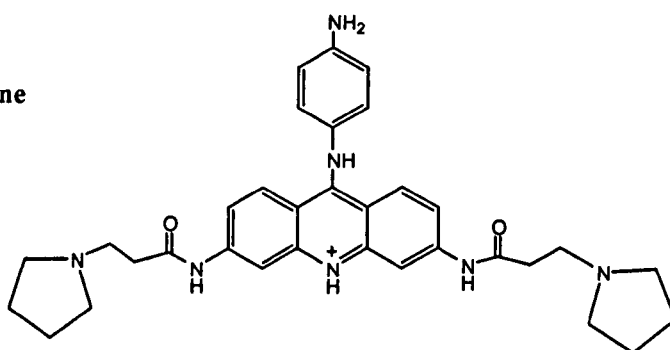
It is well established that the endogenous RNA template region of the telomerase requires a single-stranded, non-folded telomeric DNA primer for the effective addition of the telomeric repeats (7-9). However, G-rich telomeric repeats are able to assemble into novel four-stranded quadruplex structures consisting of guanine tetrads and stabilised by monovalent cations (Na^+ and K^+) (13-16). Initially Zahler and co-workers showed that inhibition of telomerase activity was possible through G-quadruplex stabilisation by K^+ (17). Later, it was suggested that formation of the G-quadruplex DNA is involved in the dissociation of the primer from the RNA template. Experimental data indicated that incorporation of a 7-deaza-dGTP nucleoside analogue instead of dGTP into the telomeric DNA primer d(TTAGGG) was involved in telomerase inhibition, as the nucleoside analogue lacks the N7 atom that is essential for the formation of G-quadruplex structure (18). Thus, small molecules that bind and stabilise the folded quadruplex DNA structure selectively over duplex DNA are potent telomerase inhibitors, with low levels of general cytotoxicity. The distinct geometrical features of G-quadruplex DNA such as four grooves and a channel of negative electrostatic potential can allow specific recognition by small ligands, which either intercalate at GpG steps or at the ends at XpG, GpX steps, or even bind to the grooves (10-12).

Considering the high anti-tumour potential of ligands with the capability of G-quadruplex stabilisation, several research groups have focused on structure-based design approaches to the development of compounds that interact with G-quadruplexes. Up to now, a number of compounds have been synthesised with common feature of an extended aromatic ring system that is capable of interacting with G-tetrads and have shown telomerase inhibition through the stabilisation of G-quadruplexes, figure 6.1. Recently, dibenzophenanthroline derivatives (19) and 3, 6, 9 tri-substituted acridine inhibitors (20) were reported to inhibit telomerase action in tumour cell lines with IC_{50} values of up to 28 nM and 60 nM, respectively. Inhibition by these compounds is correlated to selective stabilisation of the human DNA quadruplex structure, as demonstrated by fluorescence resonance energy transfer and surface plasmon resonance experiments, respectively. Other compounds have also been reported to inhibit telomerase activity with smaller IC_{50} values such as tetra-(N-methyl-4-pyridyl) and tetra-(N-methyl-2-pyridyl) porphyrin compounds (21, 22), 2,6-disubstituted anthraquinones (23), 2,7-disubstituted fluorenones (24), and a dicationic perylene tetracarboxylic diimide derivative, abbreviated PIPER (25). By using absorption or circular dichroism and NMR spectroscopy 2,6 diamidoanthraquinone, (N-methyl-pyridyl) porphyrins and the perylene derivative (PIPER) have been shown to bind and stabilise parallel or anti-parallel G-quadruplex structures.

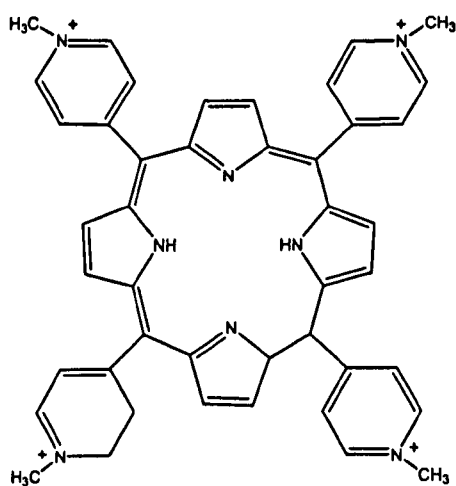
dibenzophenanthroline



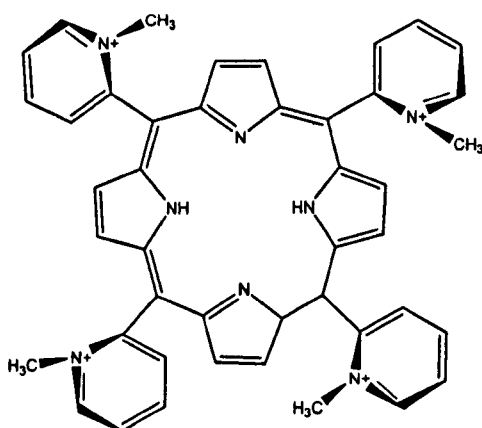
3,6,9 trisubstituted acridine



tetra-(N-methyl-4-pyridyl) porphyrin



tetra-(N-methyl-2-pyridyl) porphyrin



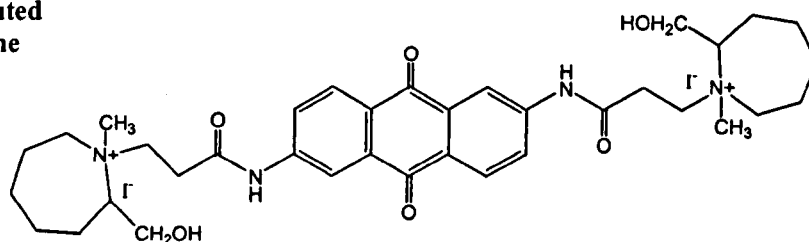
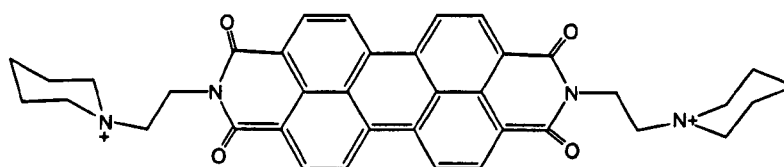
2,6-disubstituted anthraquinone**PIPER**

Figure 6.1 Active telomerase inhibitors by interaction with G-quadruplex DNA.

It is well demonstrated that such ligands bind to quadruplex DNA and that this interaction is associated with telomerase inhibition. The spectroscopic data suggest that such ligands intercalate externally to the surface of G-tetrads rather than intercalate between the stacks of the G-tetrads. However, there is one exceptional study that suggests intercalation of tetra-(N-methyl-2-pyridyl) porphyrin between the G-tetrads based on calorimetric (ITC) and spectroscopic data (26). The same porphyrin derivative was examined by an electrophoretic photocleavage assay (21), and NMR, UV spectroscopy (22), but these data are consistent with the ligands bound externally to the G-tetrads. Molecular modelling studies show that ligands are able to form stable complexes by either intercalating or end-stacking with G-tetrads adding weight to this controversy (26,27). There has been a paucity of detailed structural data available on drug-quadruplex DNA complexes due to intractable NMR spectra arising from extensive drug-induced line broadening. The only structural data to date is of the NMR structure of the PIPER-quadruplex complex, which forms either a sandwich complex bound between the blunt ends of a

quadruplex dimer formed from d(TTAGGG)₄, or intercalates by end-stacking at the GpT step of d(TAGGGTTA)₄ (25). However, more detailed structural studies are required to establish where these G-quadruplex stabilising agents are actually located within a drug–quadruplex DNA complex, and if there is indeed a universal mode of binding with quadruplex DNA or whether intercalation between G-tetrads is also possible.

Here we describe the interaction and stabilisation of quadruplex DNA by a novel fluorinated polycyclic methylacridinium cation RHPS4 (figure 6.2), which shows enhanced binding to higher-ordered DNA structures (triplex/quadruplex) over duplex and single stranded DNA as measured by differential dialysis (28). RHPS4 has been shown to induce telomere shortening with an IC₅₀ value of 0.33 μM, while decreasing tumour cell proliferation of breast 21NT cells at concentrations as low as 0.2 μM (29). RHPS4 is weakly cytotoxic (mean GI₅₀ value in the NCI 60 human tumour cell panel is 13.18 μM), giving a therapeutic index (GI₅₀/IC₅₀) of 40 (28). This activity does not appear to be associated with Taq polymerase and topoisomerase II inhibition, strongly suggesting that RHPS4 is an inhibitor of telomerase function. To investigate structural details of the drug-quadruplex interaction, as part of future rational design processes, we have studied by NMR the RHPS4 complex with the intermolecular parallel-stranded quadruplex d(TTAGGGT)₄, formed from the human telomeric repeat, figure 6.3. In addition, this study represents one of a very limited number of drug–quadruplex DNA structures so far characterized and provides new insights in the understanding of the binding mode of such quadruplex stabilising agents.

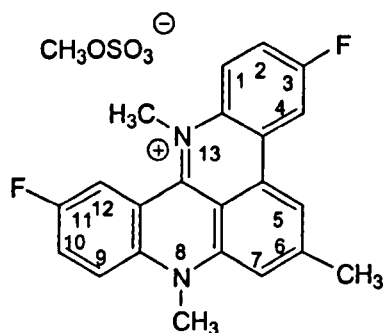


Figure 6.2 Structure of the RHPS4 compound with atom numbering scheme.

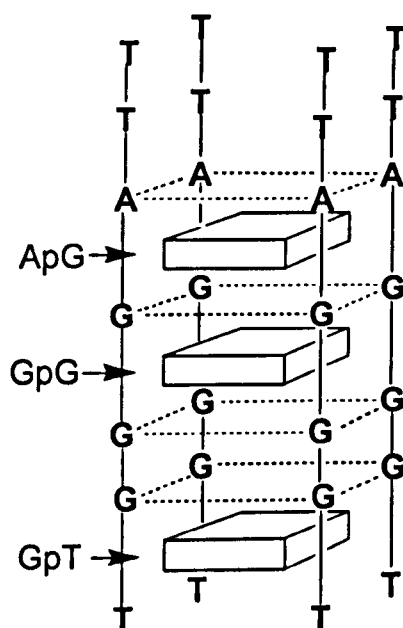


Figure 6.3 Scheme of the intermolecular quadruplex d(TTAGGGT)₄ showing potential sites of intercalation at ApG, GpG and GpT sites.

6.2 Materials and Methods

6.2.1 Sample preparation

Preparation of the d(TTAGGGT)₄ sample was described in section 5.2.1 of chapter 5. RHPS4 (3,11-Difluoro-6,8,13,-trimethyl-8H-quinacridinium methosulfate) was synthesized and purified by Robert Heald of the Cancer Research

Laboratories in Nottingham following the described procedures (28). RHPS4 was used without further purification after checking its purity by ^1H NMR. The RHPS4–quadruplex complex was formed by titrating small aliquots of drug (~5 mM solution) into a solution of the d(TTAGGGT)₄ in 90% H₂O, 10% D₂O, containing concentrations of 100 mM KCl, 10 mM K₂HPO₄ and 1 mM EDTA at pH=7.0. Titration experiment was carried out at 30 °C where the spectrum of the free quadruplex is well resolved from the resonances of the single-stranded d(TTAGGGT).

6.2.2 NMR experiments

All ^1H NMR spectra were recorded at 500 MHz on a Bruker Avance DRX-500 spectrometer and processed on R4600PC and R5000SC Silicon Graphics Indy Workstations using Bruker X-WinNMR v2.6 software. Standard phase sensitive 2D NMR pulse sequences were used to obtain NOESY, TOCSY, DQF-COSY spectra in D₂O and WATERGATE NOESY in 90% H₂O, 10% D₂O solution. One dimensional NMR experiments were recorded over 16000 data points in t_2 for each 400–512 points in t_1 with spectral width of 20000 Hz and a delay time of 1.5 s, at temperatures between 5–85 °C. NOESY spectra at various mixing times 400 ms, 300 ms, 200 ms, 150 ms, 100 ms, were recorded over 2048 data points in t_2 and 400–512 points in t_1 with 64 transients for each, with spectral width of 10000 Hz and a delay time of 1.5 s, at temperature 30–45 °C. TOCSY using 75 ms spin lock and DQF-COSY were recorded over 2048 data points in t_2 and 400–512 points in t_1 with 48 transients for each with spectral width of 10000 Hz and a delay time of 1.5 s, at temperature 45 °C. WATERGATE-NOESY spectra at 300 ms, 200 ms, and 100 ms mixing times were acquired using a sample in 90% H₂O and 10% D₂O solution over 2048 data points in

t_2 and 400–512 points in t_1 with 64 transients for each, and employed a 20,000 Hz spectral width and a delay time of 1.5 s, at temperature 30–45 °C.

6.2.3 NMR restraints

A set of 668 NOE restraints (161 per strand and 24 for both RHPS4 molecules) for non-exchangeable, exchangeable and intermolecular drug–quadruplex protons were determined from NOE cross-peak integration in 500 MHz NOESY spectra collected at 318 K. Volume integrals were normalised to several fixed reference distances within the DNA structure including deoxyribose H2'–H2'' (1.85 Å) and thymine CH₃–H6 (3.00 Å). Distances were estimated from data at a number of mixing times 100–200 ms in D₂O and H₂O solutions using linear regression to extrapolate to 0 ms mixing time (30). As a consequence of the four-fold symmetry of the quadruplex in solution, the set of NOEs assigned were quadrupled for each strand. Distances calculated from the extrapolation method were assigned as strong (1.8–3.0 Å) medium (2.5–4.0 Å) and weak (3.5–5.5 Å). Interproton distances involving exchangeable protons and also intermolecular NOEs between drug molecules and quadruplex were given wider constraints to the range of 2.5–5.0 Å. No restraints corresponding to the ideal hydrogen-bond geometry were included during structure calculations as a way to test the drug binding effect on hydrogen-bonding potential between the purine nucleotides. Distance restraints were checked against energy-minimized structure for large geometrical inconsistencies using MolMol (31). The highly dynamic nature of the RHPS4–quadruplex complex precludes a high-resolution structural characterisation and no further effort was extended to calculate more accurate distance restraints and force the structure to a possible improper determination.

6.2.4 Structure calculations

Energy minimisations and restrained molecular dynamics calculations were performed on an Origin 200 Silicon Graphics Server using the AMBER 6 suite (32) of programs employing the AMBER 95 force field with modifications (33) and Particle Mesh Ewald (PME) (34) method for the treatment of long-range electrostatics. The starting model of the RHPS4–quadruplex complex was generated by using an average energy-minimised d(TTAGGGT)₄ structure. Firstly, the binding sites of the quadruplex were constructed by moving apart the TTA segment and 3'-end thymine nucleotides from the G-tetrads to a separation distance of 6.5 Å, using the LEAP module of the AMBER 6 suite of programs. This manual generation of the binding sites resulted in several violations of the phosphate backbone geometry, which were corrected after a careful minimisation and equilibration of the structure. The drug molecules were manually docked between A3–G4 and G6–T7 steps with orientations according to the set of drug–quadruplex NOE restraints. The molecular structure of the RHPS4 was constructed and minimised using the Macromodel software. The partial charges of the RHPS4 (appendix 3.3) were derived using the HF/6–31G(d)/RESP methodology (35). Force field parameters for the atom types of the drug were adapted from comparable standard parameters within the AMBER parameter set parm98.dat.

The structure of the RHPS4–d(TTAGGGT)₄ complex was solvated in a periodic TIP3 water box of approximate dimensions 60 Å x 60 Å x 60 Å, which extended to a distance of 10 Å from any solute atom and contained 5000 water molecules. Two internal potassium ions K⁺, using standard parameters for the AMBER force field, were manually positioned in the central channel between adjacent G-quartets. The potassium ions were positioned equidistantly to two

adjacent G-quartets to allow octahedral coordination with their four polarized carbonyl oxygen O6 atoms. The quadruplex system was neutralized with 20 potassium ions placed at the most negative locations with the LEAP module. All the potassium ions were treated as part of the solvent.

The drug–quadruplex system was allowed to equilibrate fully before we performed restrained molecular dynamics. Minimisation was performed with 50 steps of steepest descent and 5000 steps of conjugate gradient on the water and counterions first, with the d(TTAGGGT)₄ coordinates frozen, followed by a further 5000 steps on all the components of the system. Next, 10 ps unrestrained molecular dynamics were run at 100 K on the water alone with the DNA and potassium ions constrained, followed by another 10 ps to allow the potassium ions to move. In the following 5 ps of dynamics the temperature of the system was increased from 100 K to 300 K. At this point the set of drug–quadruplex NOE restraints were introduced gradually in the form of square well potentials with a force constant of 30 kcal mol⁻¹ Å⁻² in 10 ps. NOE restraints were assigned manually to specific bases within a given tetrad and according to the orientation of the drug within binding sites. In the next runs, each of them of 10 ps dynamics, the DNA force constant is gradually reduced from 100 to 50, 25, 10, 5, and 2.5 kcal mol⁻¹ Å⁻². Restrained molecular dynamics for 100 ps with the whole set of NOE restraints were followed with the drug and d(TTAGGGT)₄ fully free. Snapshots from the simulation were extracted at each picosecond. In parallel, unrestrained molecular dynamics of the equilibrated drug–quadruplex system were carried out in order to access the impact of the restraints in the structure of the complex. The average structures and the RMSDs from the averages structure were calculated using CARNAL module of AMBER suite.

6.3 Results and Discussion

6.3.1 1D NMR titration and spectral features

Although there have been a number of intermolecular and intramolecular G-quadruplexes structures determined so far by NMR spectroscopy, the nature of drug-quadruplex complexes previously studied made their detailed structural characterisation extremely difficult. We have chosen to study the interaction of RHPS4 with the parallel intermolecular quadruplex d(TTAGGGT)₄ that contains the human telomeric repeat and gives well resolved NMR spectra. We have previously determined the ¹H chemical shifts of the free quadruplex d(TTAGGGT)₄ following standard sequential assignment procedures via NOESY and TOCSY spectra (chapter 5).

RHPS4 was titrated with the quadruplex d(TTAGGGT)₄ at 303 K in 90% H₂O solution and the titration was followed by 1D ¹H-NMR spectra as shown in figure 6.4. The resonances of guanine imino protons between 10–12 ppm are good probes to follow the titration as they are well separated from the crowded regions of the non-exchangeable protons. Titration up to drug:quadruplex ratio of 1:1 causes a shifting upfield and broadening of the guanine imino protons to base line. In other regions of the spectrum such as the aromatic and thymine methyl regions new resonances appear, however there is also a line broadening effect that prevents us from following the new peaks arising. Titrating the drug beyond this point up to a 2:1 ratio produces a new set of resonances that are much sharper and appreciable shifted by up to 0.9 ppm, however the G4 imino protons remain exchange broadened.

Titration of the RHPS4 up to 2:1 titration generally induces a sharper spectrum that is better resolved, however there are still some broadening effects, which does not allow us to attribute the number of resonances to a single species in

the solution. There is also the possibility that there are still some resonances from the single stranded form of the d(TTAGGGT) present in the solution. The methyl protons of the drug are easily distinguished, but most of the aromatic proton resonances are overlapped with the aromatic peaks of the bound quadruplex. A reasonable explanation for the observed data is that the drug binds to the quadruplex d(TTAGGGT)₄ in a number of possible bound conformations in intermediate exchange or binds to a number of sites with similar affinity. The different broadening of the imino proton resonances of G4 versus G6 suggests different drug mobility at different sites. Binding of the asymmetric RHPS4 in slow exchange in either of the bound conformations would lift the symmetry of the quadruplex structure contributing to the line broadening.

1D ¹H-NMR spectra at various temperatures also suggest that the drug is in intermediate exchange, figure 6.5. The number of proton resonances at temperatures below 303 K cannot account for a single symmetrical 2:1 complex. In addition, NOESY spectra of the complex at low temperatures are less informative due to dispersion of the NOE signal among different conformations. At temperatures above 303 K lines sharpen progressively indicating that the drug is in faster exchange on the NMR time scale.

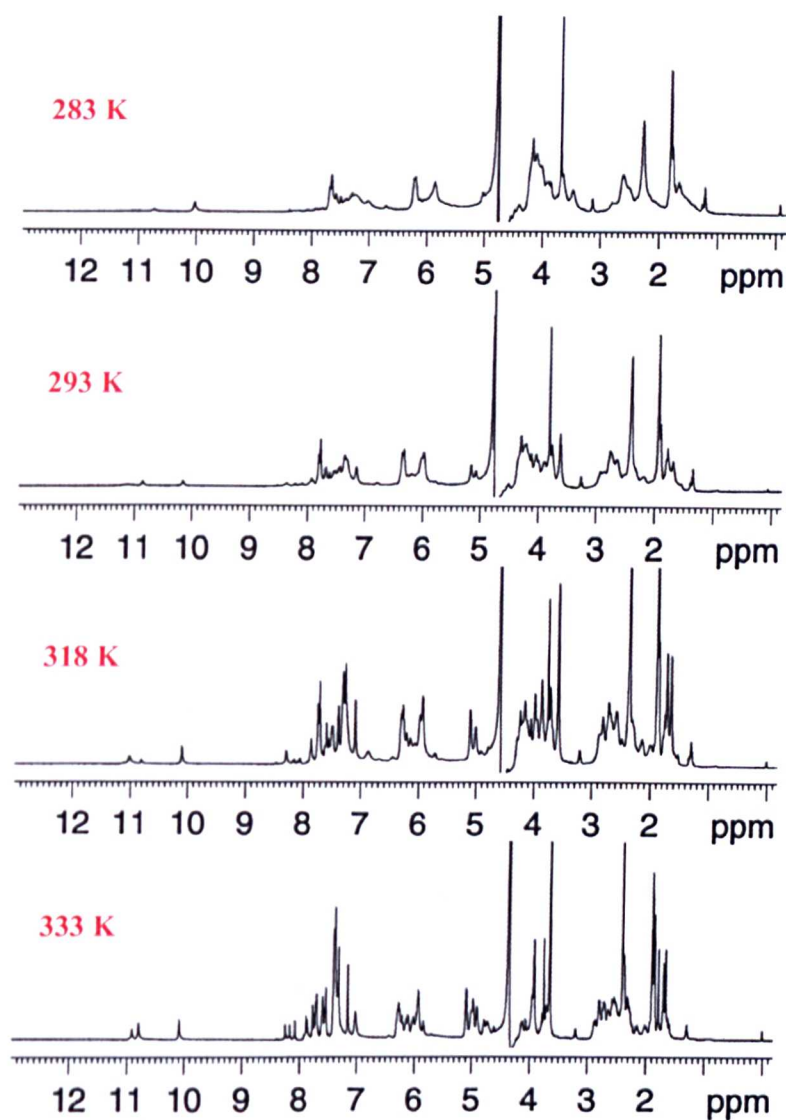


Figure 6.5 1D ¹H-NMR spectra of 2:1 RHPS4–d(TTAGGGT)₄ complex at various temperatures.

The spectrum at 318 K was generally of superior quality and the number of peaks in the aromatic region corresponds to a single set of quadruplex resonances. In addition, at 318 K, we can easily identify the resonances of the aromatic and methyl protons of the bound RHPS4. Thus, 2D NMR data of the 2:1 drug–quadruplex were collected at this temperature. The effect of RHPS4 on quadruplex stability is clearly evident from the intensity of the guanine imino proton resonances. Guanine imino

melted. However, imino but also aromatic and methyl resonances are still visible in the drug complex at temperatures up to 353 K, figure 6.6. This difference in melting temperature of at least 20 °C indicates a strong stabilisation.

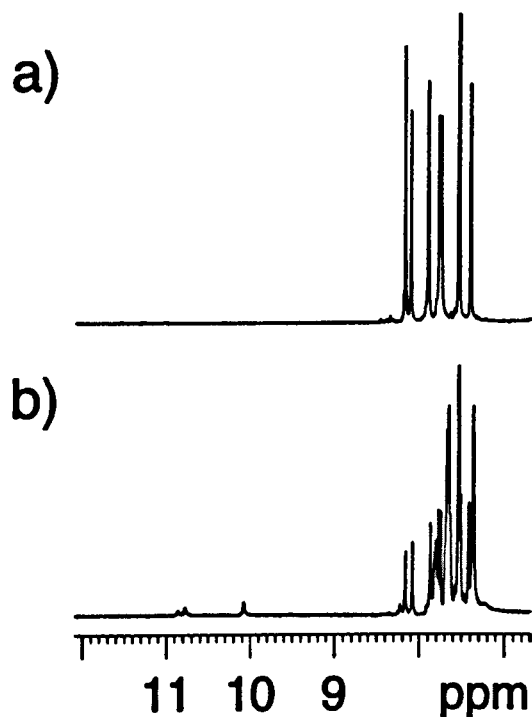


Figure 6.6 1D NMR data recorded in 90% H₂O solution showing the melting of the guanine imino proton resonances for G4, G5 and G6 and aromatic protons of the a) free d(TTAGGGT)₄ at 333 K and b) 2:1 complex RHPS4-d(TTAGGGT)₄ at 353 K.

6.3.2 NMR analysis of the 2:1 complex

The single set of drug resonances recorded at 318 K indicates a fast exchange between different orientations within the binding sites. Chemical shifts of drug protons were readily assigned via a combination of NOESY and TOCSY (figure 6.7) experiments at 318 K. A full set of RHPS4 proton assignments is found in the Appendix 1.10. A set of 24 intermolecular NOEs (table 6.1) of drug-quadruplex interactions were identified and used to pinpoint the drug binding sites (figure 6.8). By far, the largest number of NOEs involves the base and sugar protons of G4 and

G6, while no NOEs are identified to G5 at the centre of the G-tetrads stack. The latter excludes any possibility of drug molecules intercalated between the G5–G6 step. NOEs are detected at the A3–G4 step to both purine nucleotides indicating that the ApG step is a primary intercalation site. A number of NOEs to G6 show that the drug also inserts, but to a lesser extent, at the G6–T7 step, however, no NOEs are observed to T7 despite the perturbations to its chemical shifts. Two very strong NOEs from the drug 8-CH₃ to G4 NH and G6 NH show that end stacking on the G-quadruplex is energetically very favourable. NOEs from drug 8-CH₃ and 13-CH₃ to G4 H8 and G6 H8 and from drug H5, H7, H9, H10, H1 and H12 protons to a number of DNA base and sugar protons position the edges of the acridine ligand in the grooves. A number of NOE cross peaks between RHPS4 and protons of G4, G6, and A3 are shown in figure 6.9.

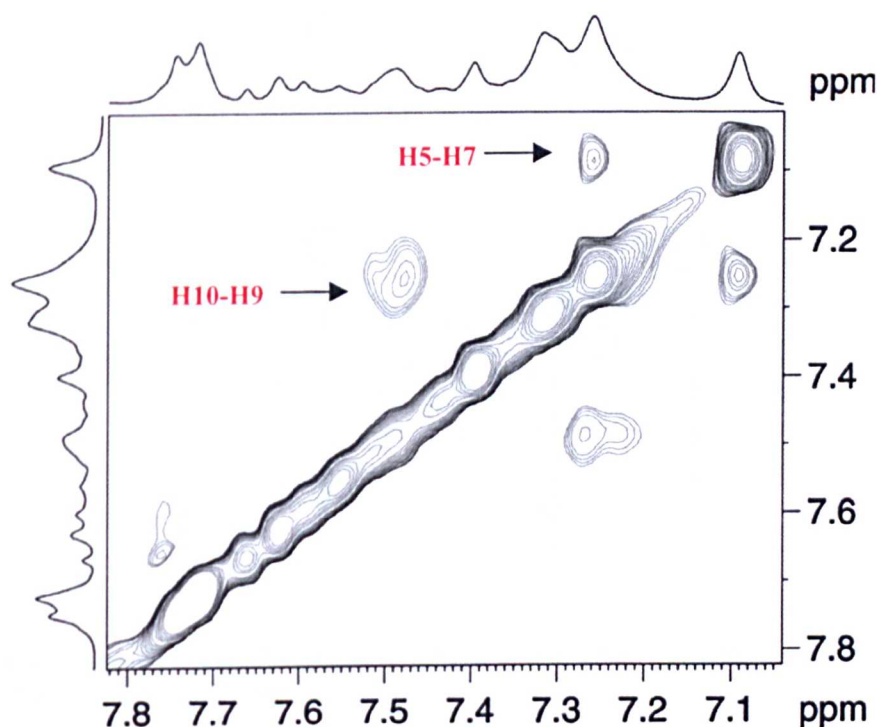


Figure 6.7 Portion of the TOCSY spectrum of the 2:1 RHPS4–d(TTAGGGT)₄ complex in H₂O solution, recorded at 318 K with mixing time 75 ms, showing the assignment of cross coupling interactions within RHPS4.

Table 6.1 The set of 24 intermolecular NOEs between RHPS4 molecules and A3, G4 and G6 nucleotides of the quadruplex structure d(TTAGGGT)₄ used in structure calculations.

RHPS4/nucleotide	A3	G4	G6
6-Me		H8	H3'
8-Me	H1', H8	H1', H8, NH	H1', H8, NH
13-Me		H8, NH	H8, NH
H5		H4'	H2'/H2''
H7	H2'/H2''	H5'	H2'/H2''
H9		H4', H5'	
H10/H12	H2'/H2''	NH	NH

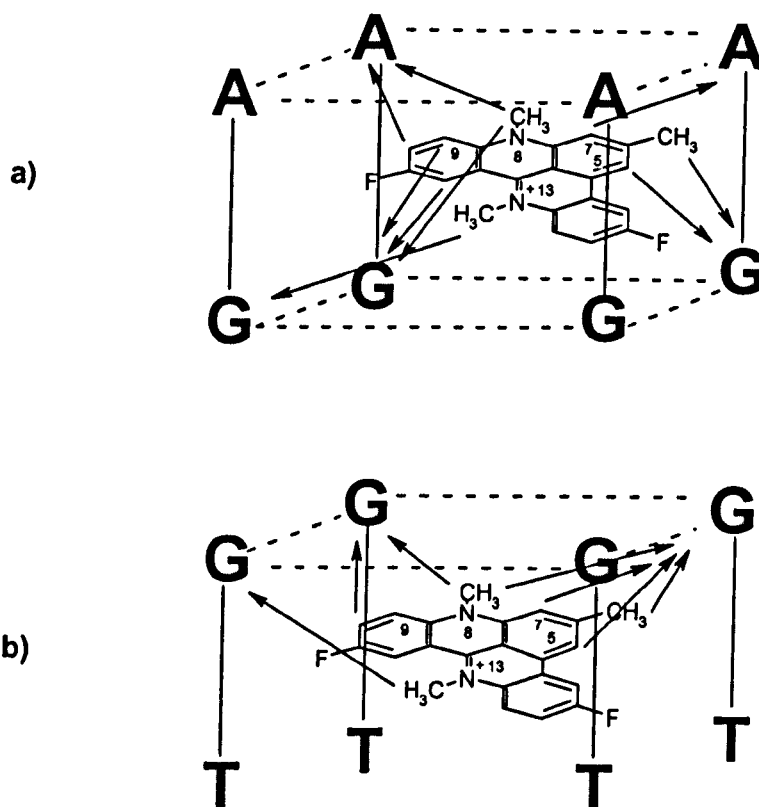


Figure 6.8 Schematic representations of selected intermolecular NOEs between RHPS4 and a) A3-G4 step and b) G6-T7 step.

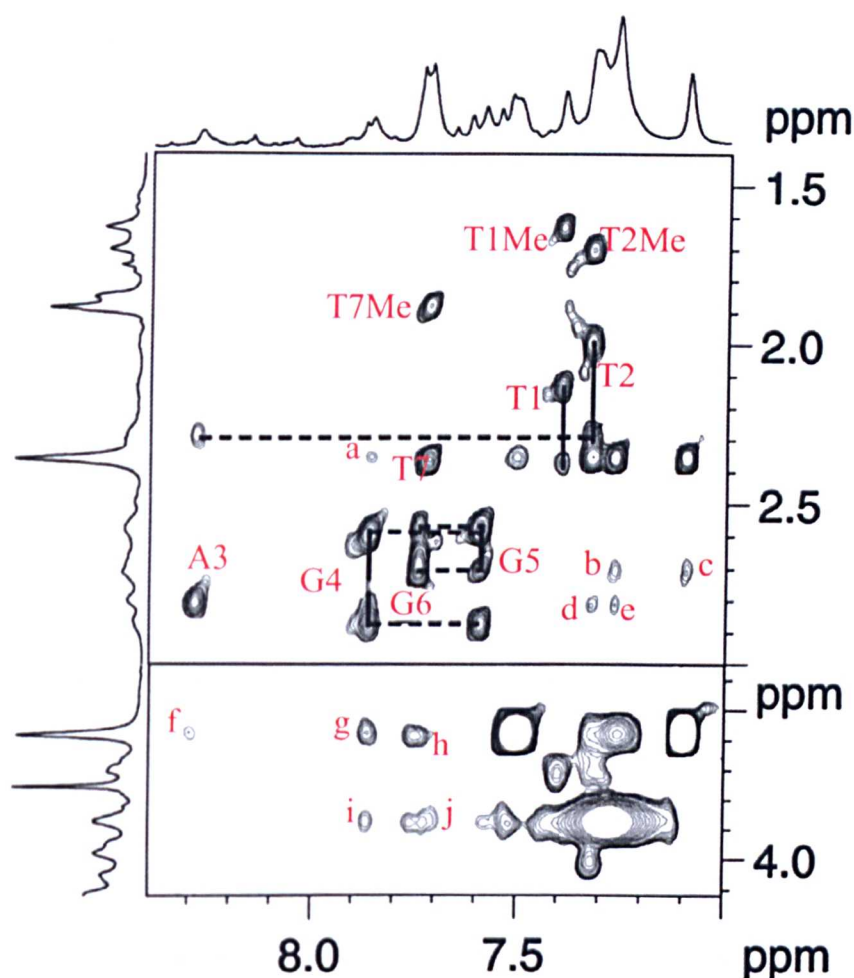


Figure 6.9 Portion of the 250 ms NOESY spectrum of the 2:1 RHPS4- $d(\text{TTAGGGT})_4$ complex recorded at 318 K highlighting a number of key NOEs. In the top panel, NOEs labelled T1 through to T7 represent intranucleotide base H6 or H8 to deoxyribose H2'/H2". Dashed horizontal lines highlight internucleotide NOE connectivities. Thymine methyl to H6 NOEs are also assigned. Key drug-DNA NOEs are labelled a-j as follows: (a) 6-Me-G4H8, (b) H7-G6H2'/H2", (c) H5-G6H2'/H2", (d) H10 or H12-A3H2'/H2", (e) H7-A3H2'/H2", lower panel; (f) 8-Me-A3H8, (g) 8-Me-G4H8, (h) 8-Me-G6H8, (i) 13-Me-G4H8, and (j) 13-Me-G6H8/T7H6. RHPS4 numbering scheme is shown in figure 6.2.

NOESY experiments at several mixing times (400–100 ms) were recorded on the 2:1 RHPS4–d(TTAGGGT)₄ complex at temperatures of 318 K (figure 6.10) and 303 K in H₂O and D₂O solutions. Expanded regions A and B of the NOESY spectrum of the complex in D₂O solution showing H8/H6–H1' and H8/H6–H3' NOE connectivities, recorded at 200 ms mixing time and 318 K are plotted in figure 6.11. At 318 K, the intermolecular bound quadruplex structure retains a four-folded symmetry and NOE signals for each strand are equivalent. Almost all NOE connectivities can be followed between the non-exchangeable protons such as H8/H6–H1' and H8/H6–H2'/H2'' along the sequence, allowing us to determine the chemical shifts of all the protons already assigned for the free quadruplex d(TTAGGGT)₄. We also observe NOEs between base protons H8/H6 and its own and 5' flanking sugar protons H2'/H2'' and H3' as they are observed in free quadruplex NOESY spectrum. While some of the NOE cross peaks concerning thymine nucleotides are weak, the NOE cross peaks between guanines are strong and show that they stack very well with each other. The effect of ligand binding on line widths means that some weak NOEs observed in the free quadruplex are difficult to detect in the bound state. Thus, there are a few examples of NOEs that are not observed in the bound state of the quadruplex d(TTAGGGT)₄ such as the H1'–H6 proton connectivity between T1–T2, T2–A3 and G6–T7 residues and also NOEs between aromatic H8–H8 or H8–H6 protons.

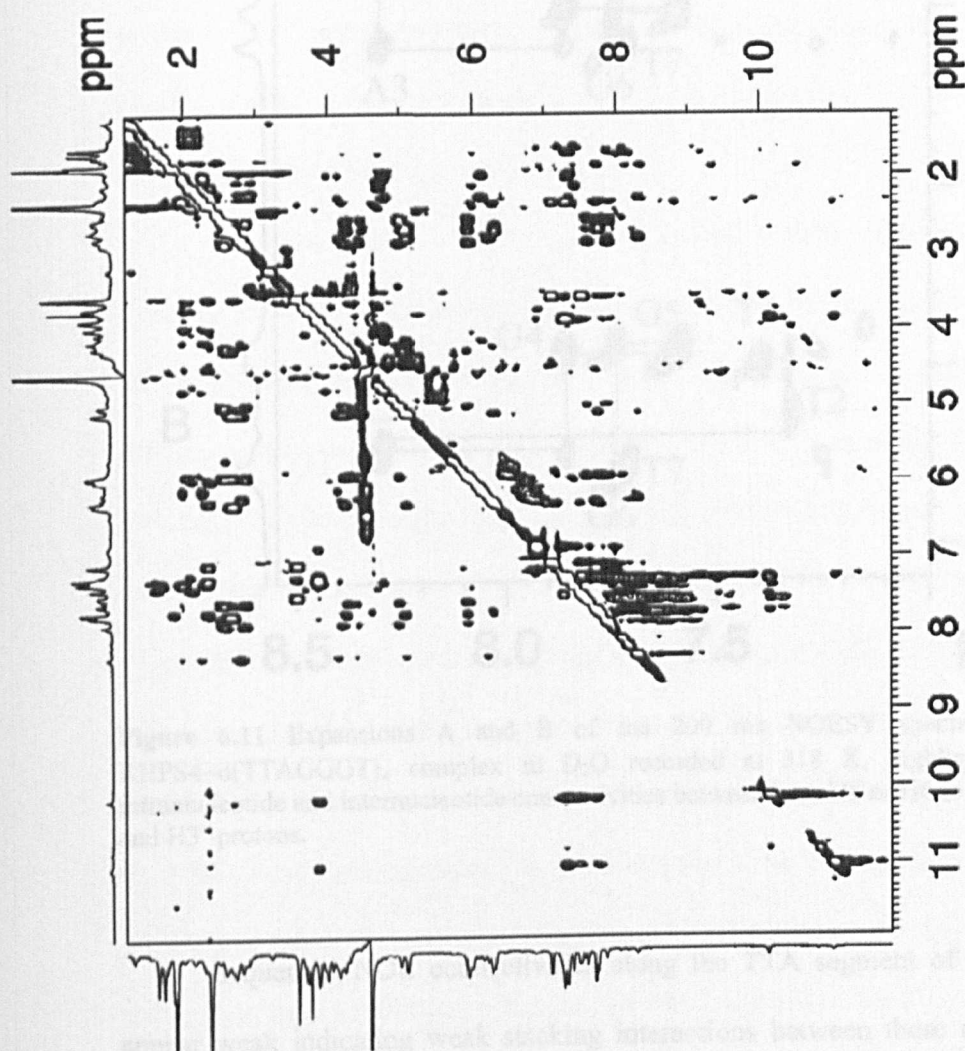


Figure 6.10 NOESY spectrum of the 2:1 RHPS4-d(TTAGGGT)₄ complex in H₂O solution, recorded at 318 K with mixing time 300 ms.

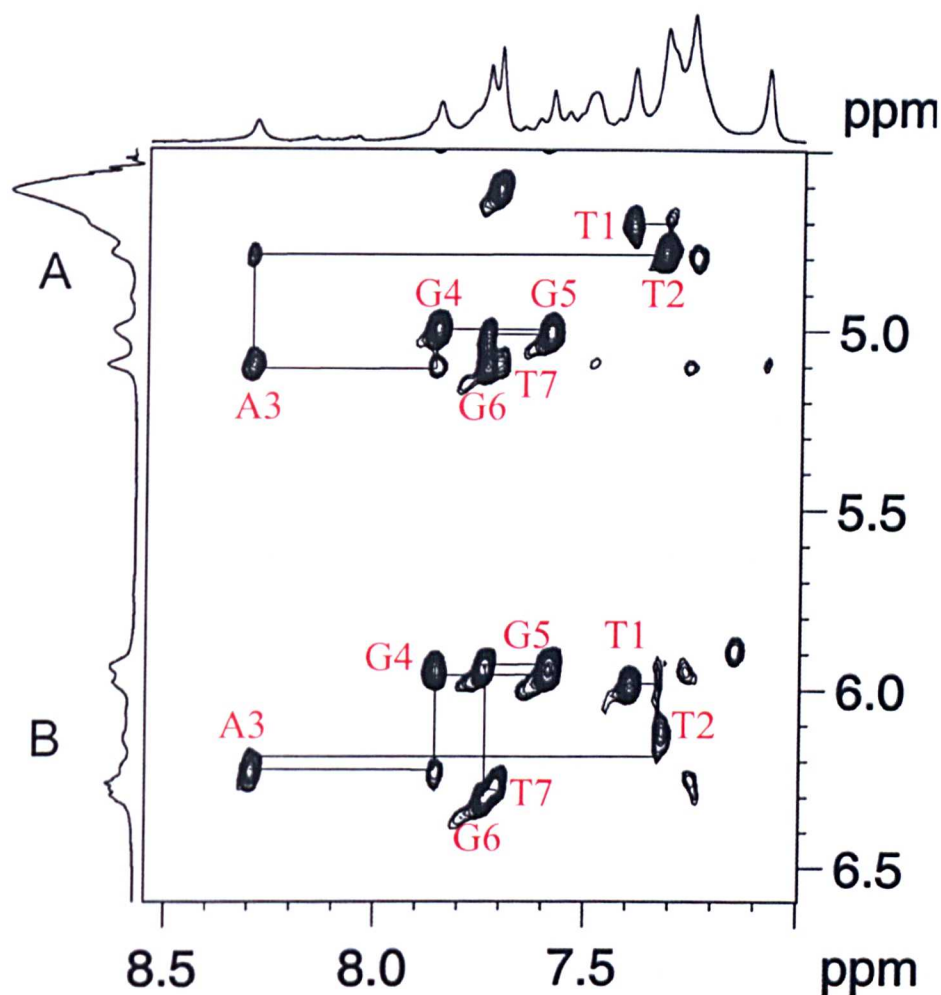


Figure 6.11 Expansions A and B of the 200 ms NOESY spectrum of the 2:1 RHPS4–d(TTAGGGT)₄ complex in D₂O recorded at 318 K, highlighting sequential intranucleotide and internucleotide connectivities between base H8 or H6 to deoxyribose H1' and H3' protons.

Sequential NOE connectivities along the TTA segment of the quadruplex appear weak indicating weak stacking interactions between these nucleotides. In addition, as in the free state, thymine imino protons of the bound quadruplex DNA are not detected at temperatures above 283 K suggesting again that drug binding does not stabilise the relatively loose conformation of the quadruplex at the 5'-end. The differences in chemical shifts of non-exchangeable protons between the bound (2:1 complex) and the free quadruplex d(TTAGGGT)₄, show that the thymines T1 and T2

are unperturbed by ligand binding. However, the base H6 and CH₃ protons of T7 shift by 0.2 and 0.35 ppm, respectively. Proton resonances of the purine residues of the quadruplex show only a small perturbation ~0.1 ppm. Thus, the NOE connectivities suggest that drug binding between the A3–G4 and G6–T7 sites stabilizes mainly the core (GGG)₄ quadruplex. Thymine nucleotides show no evidence of T-tetrad formation or good stacking interactions and should have a similar “loose” conformation as in the free quadruplex. A full set of DNA assignments is found in the appendix 1.7.

Free quadruplex NOESY spectra in H₂O solution showed all the NOEs expected from the guanine imino and adenine amino protons that suggest alignment of the residues in hydrogen-bonded tetrads. Guanine imino and adenine amino protons of the AGGG segment are visible in the bound state and also have a high melting temperature suggesting that the purine nucleotides form hydrogen-bonded tetrads, even at the temperature of 318 K. Specifically, the 4-NH₂ group of A3 is still visible and stabilised against exchange, as well as being upfield shifted by 0.4 ppm, suggesting that the A-tetrad is still intact in the drug complex. However, we could not observe NOEs from the adenine residues between the amino NH₂ protons and base H2 proton that indicated A-tetrad formation in the free quadruplex state, nor do we observe particularly strong NOEs with the drug protons. This is likely to arise partly from the drug binding effect on line widths, but also due to dynamic effects of the adenine residues caused by poorer stacking interactions with the π -system of the drug compared to G4-tetrad.

In NOESY spectra in H₂O solution (figure 6.12) we observe NOEs between G4NH–G5NH, G5NH–G6NH, G6NH–G6H8, G6NH–G5H8, G5NH–G5H8 but do not observe G5NH–G4H8, G4NH–G4H8 and also A3H8–G4NH, A3H2–G4NH.

The last two NOEs should not be visible if drug intercalates between A3–G4 tetrads, however, the sequential NOE pathway between guanine imino protons and base H8 protons in the AGGG segment is disrupted in the bound state because of the extensive line broadening of the G4 imino proton. Localisation of line broadening and unseen NOEs for A3 and G4 amino and imino protons, respectively, is an indication that these nucleotides are most affected by drug binding. This strongly suggests that the A3–G4 step is the primary intercalation site for RHPS4.

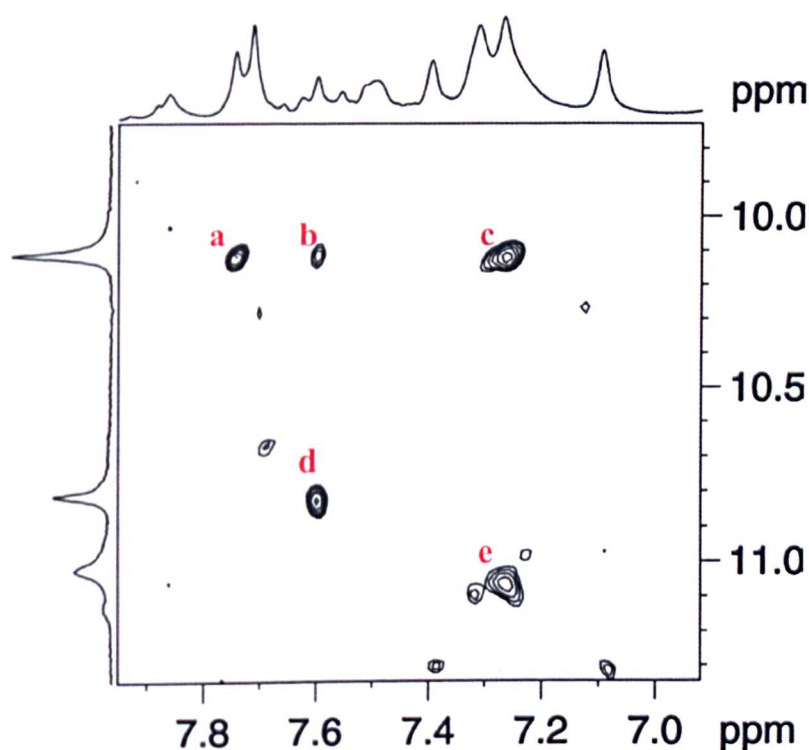


Figure 6.12 Expanded region of the NOESY spectrum of the 2:1 RHPS4–d(TTAGGGT)₄ complex in H₂O solution, recorded at 318 K with mixing time 300 ms. NOEs labelled (a) – (e) are assigned as: (a) G6NH–G6H8, (b) G6NH–G5H8, (c) G6NH–H10/H12, (d) G5NH–G5H8, (e) G4NH–H10/H12.

The NMR titration spectra suggest that the drug complex is highly dynamic with the ligand in intermediate exchange between different bound orientations at room temperature. At higher temperatures (>318 K) the ligand is in faster exchange with all possible orientations contributing to an average NOE intensity. The NOE data clearly show that the ligand interacts with A3–G4 and G6–T7 steps and that the RHPS4 complex is stabilised through extensive π -stacking with the G-tetrads. In addition, NOE cross peak patterns suggest that ligand binding leaves the structure of the G-tetrads largely unperturbed, nor are there significant conformational changes to the rest of the quadruplex structure. In the complex of PIPER with d(TAGGGTTA)₄, formation of the 1:1 complex appears to favour end-stacking at the GpT step rather than the ApG step which the authors attribute to the difficulty of disrupting the stronger base stacking between purine tetrads (25). In contrast, in our complex RHPS4 binding to both the GpT and ApG step appears to be energetically favourable and RHPS4 appears to show relatively little discrimination between these two intercalation sites. Our data suggest that end-stacking on the G-tetrads significantly stabilises the quadruplex structure ($\Delta T_m > 20$ °C) even though the conformation of the terminal thymine and adenine nucleotides appears to be reasonably dynamic. The reported studies of the PIPER complex similarly observe fast exchange of the ligand between different bound conformations while the core (GGG)₄ structure of the quadruplex remains unperturbed.

6.3.3 Molecular modelling of the 2:1 complex

In order to extract structural information regarding the interaction of the RHPS4 ligand with the distinct geometrical features of the d(TTAGGGT)₄ quadruplex, we have carried out an MD approach using NOE restraints. Due to the

asymmetry of the drug, some quadruplex resonances might be expected to be multiplied upon drug binding however the single set of resonances observed indicates a fast reorientation of the bound drug on the NMR time scale. The four-fold degeneracy of the quadruplex structure that is retained in solution at 318 K complicates the determination of a unique solution structure since intermolecular NOEs have to be assigned to specific nucleotides.

Based on the set of 24 drug–quadruplex NOEs, we have docked the drug molecules at the A3–G4 and G6–T7 step with specific orientations in order to satisfy the majority of the NOE restraints. Although we have modelled a single bound drug conformation at the ApG and GpT intercalation sites, the dynamic nature of the interaction, evident from ^1H (figure 6.5) and ^{19}F (figure 6.13) temperature-dependent line widths, does not preclude multiple interconverting conformations all of which will contribute to the average NOE intensity. The analysis of NOEs from a single DNA proton to several drug protons apparently remote from each other on the acridine ring can arise from the drug rotating between the four equivalent stacked conformations, though this is further complicated by the four-fold symmetry of the quadruplex which makes each strand equivalent by NMR.

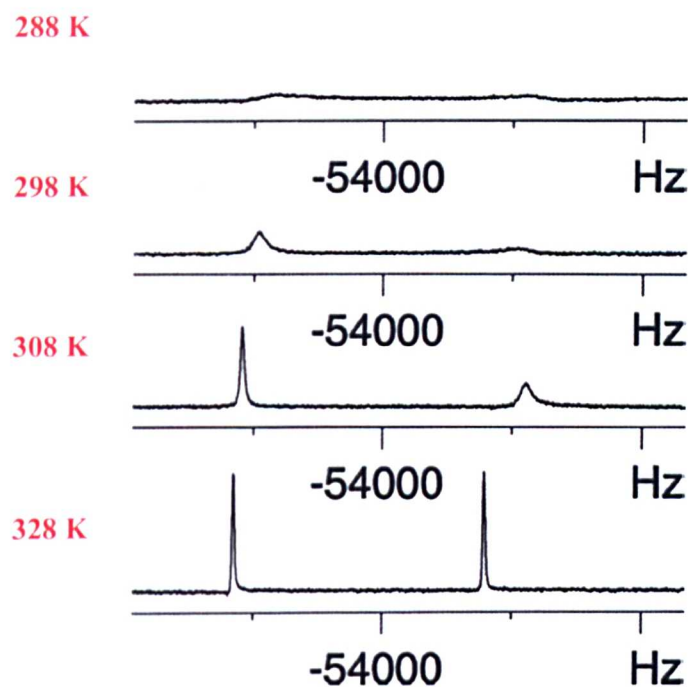


Figure 6.13 ^{19}F 1D NMR spectra at various temperatures illustrating the dependence in line widths of the fluorine resonances for F3 and F11 fluorine atoms of the RHPS4 in the 2:1 complex.

The initial structure of the complex had large geometrical violations arising from the manual generation of the binding sites. A careful minimization and equilibration of the structure was necessary. The equilibration was followed by 100ps of NOE restrained MD simulation and energy minimisation. In the final energy-minimised structure no restraint violations $>0.3\text{\AA}$ were observed for the DNA, however, 3 of the 24 drug–DNA NOEs were found to have distances over 5\AA . The majority of the drug–DNA NOEs is satisfied when the 8- CH_3 and 6- CH_3 groups of the ligand are orientated towards the grooves and the 13- CH_3 lies close to the central ion channel of the quadruplex structure. The 180° reorientation of the ligand having the 13- CH_3 pointing into the groove of the quadruplex structure is in accordance with only a few drug–DNA NOEs, while most of the distances from the 8- CH_3 and 6- CH_3 and the aromatic protons to the quadruplex protons are found over 5\AA . Thus, the set

of drug–DNA NOEs suggest that there are a small number of more favourable orientations for the drug. The low energy structure that satisfies the majority of the distance restraints is shown in figure 6.14, with the partial positive charge on the acridine 13-N acting as a pseudo potassium ion positioned above the centre of the G-tetrad. In both ApG and GpT intercalation sites, the drug is seen to converge to the same orientation while the π -system of the drug is seen to overlap with two bases of each G-tetrad. The overlap between the π -system of the drug and the complementary G-tetrad of G4 and G6 (figure 6.15), suggests several key positions around the acridine ring such as the position of the two fluorine atoms that may provide future points for ligand substitution to enhance quadruplex affinity and specificity through interactions in the grooves.

The structure of the RHPS4 complex shown in figure 6.14, which lacks thymine nucleotides of the 5' end for clarification, has *anti* conformations for all nucleotides and right-handed helical backbone geometry for each strand. The guanine residues appear to have an almost planar conformation and stack on top of each other while Hoogsteen N1–O6 and N2–N7 hydrogen bonds maintain well-defined G-tetrads. The adenine nucleotides are less stable than the guanines because of only partial stacking interactions with the π -system of the drug. Significant flexibility of the thymine nucleotides at the 5' end of the quadruplex structure is observed, giving rise to a disordered conformation throughout the trajectory, as in the unbound state.

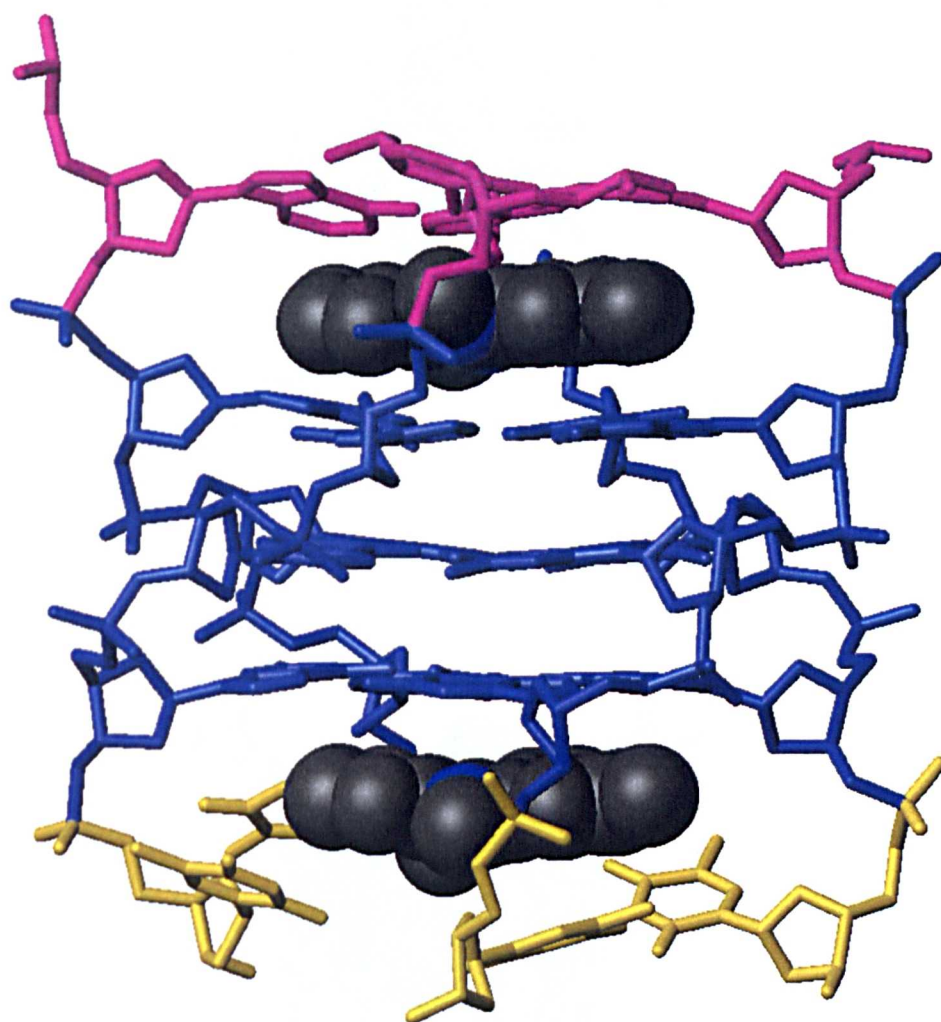


Figure 6.14 Structure of the 2:1 complex with d(TTA*GGG*T)₄ showing the AGGGT core (for clarity) with RHPS4 intercalated at the A3–G4 and G6–T7 sites.

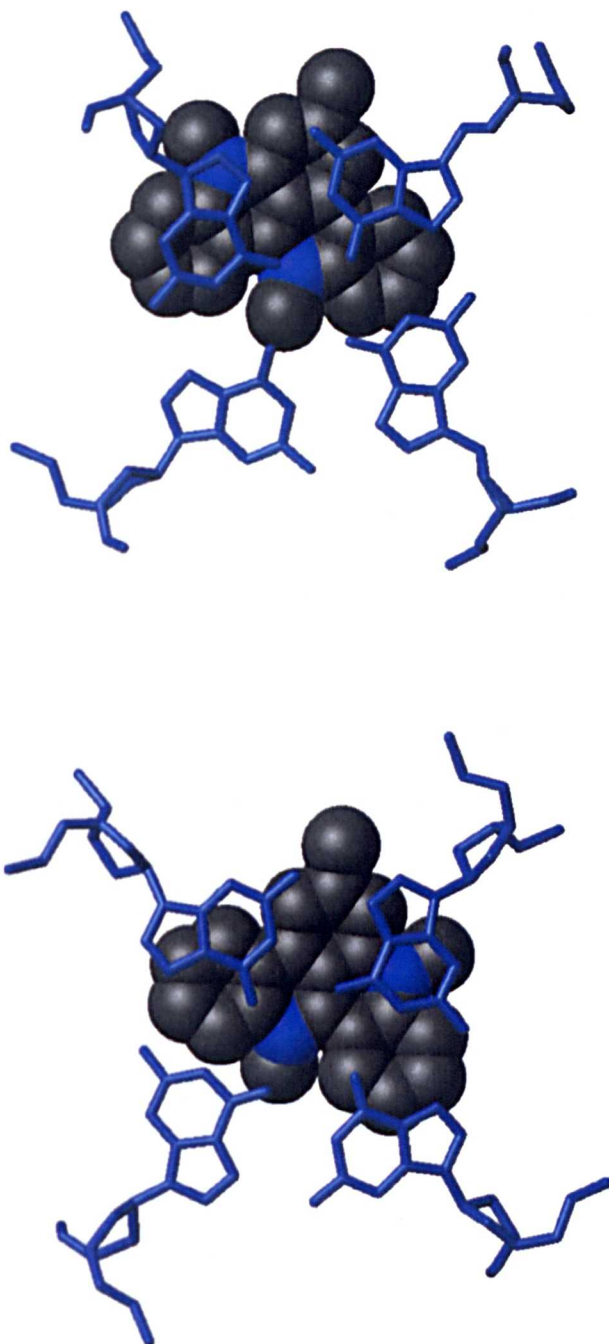


Figure 6.15 The π -stacking of the RHPS4 with the (a) G4 and (b) G6 tetrads showing the orientation with respect to the various grooves.

Analysis of the unrestrained molecular dynamics simulation of the complex over 1000 ps show that the structure (AGGG)₄ remains largely unchanged from the NMR-derived minimised structure. This RMSDs between the starting NMR structure and those calculated from dynamics is 1.73 ± 0.16 Å for (AGGG)₄ and 2.63 ± 0.38 Å for all heavy atoms of (TTAGGGT)₄. Variation of the N1–N6, N1–O6 and N2–N7 Hoogsteen hydrogen-bond distances for A-tetrad and G-tetrads during the 1000 ps of the simulation is plotted in figure 6.16. Variation of the hydrogen bonds for the G-tetrads is not significant and within the limits of acceptable hydrogen-bond distance, while the hydrogen-bond distance of N1–N6 shows more flexibility and suggests a significant breathing of the A-tetrad structure. Average values of Hoogsteen hydrogen bonds of G-tetrads (table 6.2) show no significant difference compared to the free quadruplex structure (table 5.3). In contrast, the average N1–N6 distance of A-tetrad is increased ~ 0.35 Å. Thus, the data suggest that RHPS4 binding stabilises the core (GGG)₄ quadruplex structure and while the A-tetrad is less stable than in the unbound state, the overall effect of binding is favourable.

Table 6.2 Average hydrogen-bond distances and standard deviation (parentheses) for core quadruplex (AGGG)₄.

Residue	N7–N2 H-bond	N1–O6 H-bond	N1–N6 H-bond
G4	3.04 (0.15)	3.03 (0.15)	
G5	2.99 (0.13)	2.99 (0.13)	
G6	2.98 (0.13)	2.98 (0.13)	
A3			3.46 (0.23)

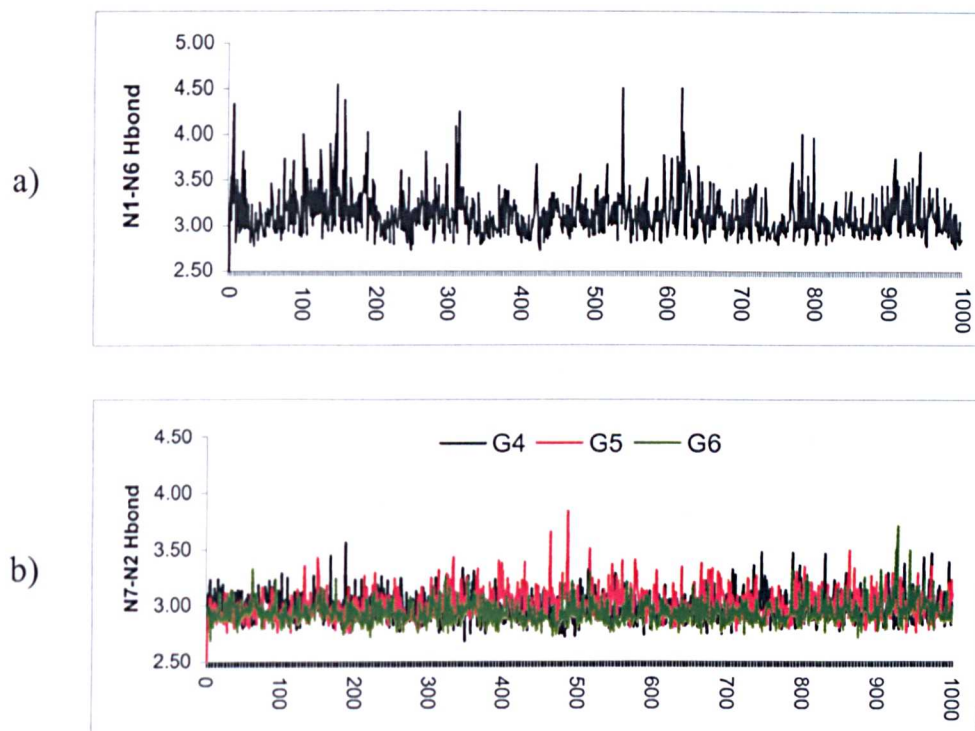


Figure 6.16 Variation of the N1–N6 and N2–N7 Hoogsteen hydrogen-bond distances in Å a) for A-tetrad b) for G-tetrads respectively during the 1000 ps of the simulation.

6.4 Conclusions

Recent detailed studies of cationic porphyrin–DNA quadruplex interactions have suggested both GpG intercalation and violation of the neighbouring site exclusion principle (26). In contrast, the perylenetetracarboxylic diimide intercalator (PIPER) described by Federoff *et al.* (25), and the polycyclic methylacridinium cation reported here, are unable to disrupt the G-tetrad core by insertion at a GpG step with displacement of stabilising K^+ ions. Instead, the drug forms complex-stabilising interactions by stacking on the ends of the G-quadruplex. Such a process is likely to be energetically favourable compared to interior stacking with the G-tetrads requiring disruption of the extremely stable stacking interaction at the GpG steps and displacement of the monovalent cations from the central channel of the

structure. In addition, the kinetics of disruption and opening of the G-tetrads by the ligand would be expected to be slow due to large activation barriers. Similar conclusions have been drawn from modelling studies of a disubstituted anthraquinone (27), and a number of tri-substituted acridine derivatives with the human intramolecular quadruplex where the drug is proposed to bind to the TTA cross-over loop while end-stacking with a G-tetrad (20).

The fluorinated polycyclic methylacridinium cation RHPS4 is a potent telomerase inhibitor as tested in tumour cell lines (29) and interacts selectively with high order DNA structures. Here, it has been shown that RHPS4 forms a stable G-quadruplex complex with d(TTAGGGT)₄ by end-stacking externally to the G-tetrads of the ApG and GpT steps. The NMR data clearly suggest that the RHPS4 complex is dynamic in nature with the drug bound in different orientations that are exchanging on the NMR time scale. However, the set of intermolecular NOEs assigned for the complex are most satisfied with the drug positioned as in figure 6.14, enabling us to suggest key positions where further substitution may enhance G-quadruplex affinity and specificity. Synthetic studies towards addition of side chains to the core structure of the RHPS4 are in progress in collaboration with the CRC laboratories in Nottingham. The RHPS4–d(TTAGGGT)₄ complex represents one of a very limited number of drug–quadruplex structures so far characterised and provides further support for the model that G-quadruplex stabilising ligands bind on the exterior of the G-tetrads stack.

6.5 References

1. Blackburn, E. H. 1994 Telomeres – No End in Sight. *Cell* **77**, 621-623.
2. Zakian, V. A. 1995 Telomeres – Beginning to Understand the End. *Science* **270**, 1601-1607.
3. Delange, T. 1994 Activation of Telomerase in a Human Tumor. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 2882-2885.
4. Bryan, T. M. & Cech, T. R. 1999 Telomerase and the Maintenance of Chromosome Ends. *Curr. Opin. Cell Biol.* **11**, 318-324.
5. Harley, C. B. & Villeponteau, B. 1995 Telomeres and Telomerase in Aging and Cancer. *Curr. Opin. Genetic Dev.* **5**, 249-255.
6. Kim, N. W., Piatyszek, M. A., Prowse, K. R., Harley, C. B., West, M. D., Ho, P. L. C., Coviello, G. M., Wright, W. E., Weinrich, S. L. & Shay, J. W. 1994 Specific Association of Human Telomerase Activity with Immortal Cells and Cancer. *Science* **266**, 2011-2015.
7. Lingner, J., Hughes, T. R., Shevchenko, A., Mann, M., Lundblad, V. & Cech, T. R. 1997 Reverse Transcriptase Motifs in the Catalytic Subunit of Telomerase. *Science* **276**, 561-567.
8. Nakamura, T. M., Morin, G. B., Chapman, K. B., Weinrich, S. L., Andrews, W. H., Lingner, J., Harley, C. B. & Cech, T. R. 1997 Telomerase Catalytic Subunit Homologs from Fission Yeast and Human. *Science* **277**, 955-959.
9. Feng, J. L., Funk, W. D., Wang, S. S., Weinrich, S. L., Avilion, A. A., Chiu, C. P., Adams, R. R., Chang, E., Allsopp, R. C., Yu, J. H., Le, S. Y., West, M. D., Harley, C. B., Andrews, W. H., Greider, C. W. & Villeponteau, B. 1995 The RNA Component of Human Telomerase. *Science* **269**, 1236-1241.
10. Mergny, J. L., Mailliet, P., Lavelle, F., Riou, J. F., Laoui, A. & Helene, C. 1999 The Development of Telomerase Inhibitors: The G-quartet Approach. *Anti-Cancer Drug Des.* **14**, 327-339.
11. Perry, P. J. & Jenkins, T. C. 1999 Recent Advances in the Development of Telomerase Inhibitors for the Treatment of Cancer. *Exp. Opin. Invest. Drugs* **8**, 1981-2008.
12. Raymond, E., Soria, J. C., Izbicka, E., Boussin, F., Hurley, L. & Von Hoff, D. D. 2000 DNA G-quadruplexes, Telomere-Specific Proteins and Telomere-Associated Enzymes as Potential Targets for New Anticancer Drugs. *Invest. New Drugs* **18**, 123-137.
13. Williamson, J. R. 1993 Guanine Quartets. *Curr. Opin. Struct. Biol.* **3**, 357-362.

14. Feigon, J., Koshlap, K. M. & Smith, F. W. 1995 H-1 NMR Spectroscopy of DNA Triplexes and Quadruplexes. *Nuc. Magn. Reson. Nuc. Acids* **261**, 225-255.
15. Shafer, R. H. 1998 Stability and Structure of Model DNA Triplexes and Quadruplexes and Their Interactions with Small Ligands. *Progr. Nucleic Acid Res. Mol. Biol.* **59**, 55-94.
16. Patel, D. J., Bouaziz, S., Kettani, A. & Wang, Y. 1998 *Oxford Handbook of Nucleic Acid Structure*, p. 389-453, Oxford University Press, New York.
17. Zahler, A. M., Williamson, J. R., Cech, T. R. & Prescott, D. M. 1991 Inhibition of Telomerase by G-Quartet DNA Structures. *Nature* **350**, 718-720.
18. Fletcher, T. M., Sun, D. K., Salazar, M. & Hurley, L. H. 1998 Effect of DNA Secondary Structure on Human Telomerase Activity. *Biochemistry* **37**, 5536-5541.
19. Mergny, J. L., Lacroix, L., Teulade-Fichou, M. P., Hounsou, C., Guittat, L., Hoarau, M., Arimondo, P. B., Vigneron, J. P., Lehn, J. M., Riou, J. F., Garestier, T. & Helene, C. 2001 Telomerase Inhibitors Based on Quadruplex Ligands Selected by a Fluorescence Assay. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 3062-3067.
20. Read, M., Harrison, R. J., Romagnoli, B., Tanious, F. A., Gowan, S. H., Reszka, A. P., Wilson, W. D., Kelland, L. R. & Neidle, S. 2001 Structure-Based Design of Selective and Potent G-Quadruplex-Mediated Telomerase Inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 4844-4849.
21. Han, F. X. G., Wheelhouse, R. T. & Hurley, L. H. 1999 Interactions of TMPyP4 and TMPyP2 with Quadruplex DNA. Structural Basis for the Differential Effects on Telomerase Inhibition. *J. Am. Chem. Soc.* **121**, 3561-3570.
22. Wheelhouse, R. T., Sun, D. K., Han, H. Y., Han, F. X. G. & Hurley, L. H. 1998 Cationic Porphyrins as Telomerase Inhibitors: The Interaction of Tetra-(N-methyl-4-pyridyl)porphine with Quadruplex DNA. *J. Am. Chem. Soc.* **120**, 3261-3262.
23. Neidle, S., Harrison, R. J., Reszka, A. P. & Read, M. A. 2000 Structure-Activity Relationships among Guanine-Quadruplex Telomerase Inhibitors. *Pharmacol. Ther.* **85**, 133-139.
24. Perry, P. J., Read, M. A., Davies, R. T., Gowan, S. M., Reszka, A. P., Wood, A. A., Kelland, L. R. & Neidle, S. 1999 2,7-Disubstituted Amidofluorenone Derivatives as Inhibitors of Human Telomerase. *J. Med. Chem.* **42**, 2679-2684.
25. Fedoroff, O. Y., Salazar, M., Han, H. Y., Chemeris, V. V., Kerwin, S. M. & Hurley, L. H. 1998 NMR-based Model of a Telomerase-Inhibiting Compound Bound to G-Quadruplex DNA. *Biochemistry* **37**, 12367-12374.

26. Haq, I., Trent, J. O., Chowdhry, B. Z. & Jenkins, T. C. 1999 Intercalative G-tetraplex Stabilization of Telomeric DNA by a Cationic Porphyrin. *J. Am. Chem. Soc.* **121**, 1768-1779.
27. Read, M. A. & Neidle, S. 2000 Structural Characterization of a Guanine-quadruplex Ligand Complex. *Biochemistry* **39**, 13422-13432.
28. Heald, R. A., Modi, C., Cookson, J. C., Hutchinson, I., Laughton, C. A., Gowan, S. M., Kelland, L. R. & Stevens, M. F. G. 2001 Antitumor Polycyclic Acridines. 8.¹ Synthesis and Telomerase-Inhibitory Activity of Methylated Pentacyclic Acridinium Salts. *J. Med. Chem.* **45**, 590-597.
29. Gowan, S. M., Heald, R. A., Stevens, M. F. G. & Kelland, L. R. 2001 Potent Inhibition of Telomerase by Small-Molecule Pentacyclic Acridines Capable of Interacting with G-Quadruplexes. *Mol. Pharmacol.* **60**, 981-988.
30. Baleja, J. D., Moulton, J. & Sykes, B. D. 1990 Distance Measurement and Structure Refinement with NOE Data. *J. Magn. Reson.* **87**, 375-384.
31. Koradi, R., Billeter, M. & Wuthrich, K. 1996 MOLMOL: A Program for Display and Analysis of Macromolecular Structures. *J. Mol. Graph.* **14**, 51-55.
32. Case, D.A., Pearlman, D.A., Caldwell, J.W., Cheatham III, T.E., Ross, W.S., Simmerling, C.L., Darden, T.L., Marz, K.M., Stanton, R.V., Cheng, A.L., Vincent, J.J., Crowley, M., Tsui, V., Radmer, R.J., Duan, Y., Pitera, J., Massova, I., Seibel, G.L., Dingh, U.C., Weiner, P.K. & Kollman, P.A. (1999) AMBER 6, University of California, San Francisco.
33. Cheatham, T. E., Cieplak, P. & Kollman, P. A. 1999 A Modified Version of the Cornell *et al.* Force Field with Improved Sugar Pucker Phases and Helical Repeat. *J. Biomol. Struct. Dyn.* **16**, 845-862.
34. Yang, D. M., Lee, W. T., Darden, H. & Pedersen, L. G. 1995 Toward the Accurate Modeling of DNA - The Importance of Long-Range Electrostatics. *J. Am. Chem. Soc.* **117**, 5001-5002.
35. Bayly, C. I., Cieplak, P., Cornell, W. D. & Kollman, P. A. 1993 A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges – The RESP Model. *J. Phys. Chem.* **97**, 10269-10280.

Appendices

Appendix 1

Appendix 1.1 ¹H chemical shifts of the d(CTTTTGCAAAAG)₂ duplex at 298K.

	H1'	H2'	H2''	H3'	H4'	H6/H8	H5'/H5''	H5/Me	H2	NH
C1	5.90	2.32	2.65	4.68	4.14	7.93	3.82/ 3.88	5.98		
T2	6.25	2.33	2.68	4.95	4.29	7.70	4.20	1.71		13.70
T3	6.26	2.32	2.70	4.95	4.33	7.56	4.20	1.67		13.70
T4	6.14	2.19	2.66	4.94	4.25	7.51	4.14	1.71		13.85
T5	5.82	2.09	2.47	4.91		7.31	4.14	1.69		14.00
G6	5.78	2.61	2.65	4.97	4.35	7.86	4.09			12.58
C7	5.40	1.82	2.2	4.97	4.76	7.29	4.08	5.36		
A8	5.70	2.68	2.79	5.00	4.32	8.17	4.15		7.14	
A9	5.72	2.59	2.79	5.01	4.39	8.09	4.19		7.05	
A10	5.83	2.55	2.84	5.01	4.4	8.03	4.22		7.11	
A11	5.99	2.51	2.85	4.99	4.42	7.94	4.26		7.61	
G12	5.97	2.28	2.31	4.98	4.59	7.50	4.18			14.29

Appendix 1. ^1H chemical shifts of the d(GAAAAGCTTTTC)₂ duplex at 298K.

	H1'	H2'	H2''	H3'	H4'	H5'/H5''	H6/H8	H5/Me	H2	NH
G1	5.50	2.40	2.60	4.80	4.15	3.65	7.84			
A2	5.78	2.70	2.82	5.03	4.38	4.06/4.15	8.19		7.41	
A3	5.80	2.59	2.82	5.03	4.41	4.19	8.09		7.17	
A4	5.84	2.55	2.82	5.03	4.42	4.22	8.01		7.45	
A5	5.94	2.50	2.82	5.01	4.41	4.22	7.91		7.13	
G6	5.66	2.42	2.57	4.91	4.37	4.20/4.16	7.48			
C7	5.86	2.05	2.49	4.68	4.23	4.11/4.17	7.28	5.09		
T8	6.01	2.21	2.61	4.88	4.23	4.11	7.46	1.53		12.75
T9	6.15	2.19	2.65	4.90	4.23	4.14	7.50	1.65		13.46
T10	6.17	2.18	2.55	4.91	4.26	4.14	7.45	1.73		13.46
T11	6.16	2.24	2.63	4.92	4.31	4.18	7.50	1.70		14.12
C12	6.27	2.28	2.56	4.58	4.21	4.05	7.63	5.80		

Appendix 1.3 ^1H chemical shifts for the DNA duplex of the 2:1 d(GAAAAGCTTTTC)₂-H33258 complex at 298K.

	H1'	H2'	H2''	H3'	H4'	H5'/H5''	H6/H8	/H5/Me	H2	NH
C1	5.98	2.39	2.74	4.72	4.17	3.86/3.94	8.01	6.04		
T2	6.53	2.30	2.71	4.98	4.37	4.13/4.19	7.73	1.80		14.74
T3	5.46	1.90	2.28	4.63	4.02		7.46	1.70		14.74
T4	5.09	1.66	2.13	4.47	3.36		7.14	1.51		13.90
T5	4.57	1.25	1.49	4.45			6.74	1.45		13.65
G6	5.73	2.59	2.56	4.35	4.12		7.80			12.65
C7	5.41	1.43	2.02	4.74	4.04	4.13	7.21			
A8	5.44	2.82	2.82	5.05	4.35	3.94/4.07	8.31		7.22	
A9	6.11	2.78	2.86	5.13	4.48	4.16/4.23	8.21		7.60	
A10	5.45	2.12	2.56	4.67		3.86/4.08	7.91		7.94	
A11	5.09	1.94	2.31	4.54	3.70	2.80/2.92	7.61		8.28	
G12	5.03	1.78	1.89	4.29	3.80	3.33	7.07			15.30

Appendix 1.4 ^1H chemical shifts of the d(GTTTTGGCCAAAAC)₂ duplex at 298K.

	H1'	H2'	H2''	H3'	H4'	H5'/H5''	H6/H8	H5/Me	H2	NH
C1	5.78	2.29	2.59	4.62	4.06	3.78	7.83	5.88		
T2	6.22	2.30	2.66	4.92	4.30	4.07	7.68	1.67		13.74
T3	6.24	2.30	2.67	4.92	4.27	4.18	7.54	1.68		13.86
T4	6.12	2.16	2.63	4.92	4.23	4.10	7.49	1.68		13.97
T5	5.75	1.99	2.39	4.88	4.23	4.10	7.27	1.67		13.97
G6	5.59	2.67	2.72	4.98	4.34	4.14	7.85			12.76
G7	5.83	2.54	2.67	4.94	4.36	4.15	7.72			12.80
C8	5.85	1.98	2.38	4.92	4.12		7.29	5.22		
C9	5.20	1.87	2.18	4.75	3.98		7.38	5.55		
A10	5.70	2.69	2.79	5.00	4.32	4.05	8.17		7.19	
A11	5.71	2.58	2.78	5.01	4.38		8.09		7.03	
A12	5.82	2.54	2.83	5.00	4.39	4.19	8.02		7.10	
A13	5.98	2.49	2.83	4.98	4.40	4.20	7.92		7.58	
G14	5.93	2.24	2.29	4.57	4.23	4.15	7.47			14.27

Appendix 1.5 ^1H chemical shifts of the d(GCAAATTTGC)₂ duplex at 298K.

	H1'	H2'	H2''	H3'	H4'	H5'/H5''	H6/H8	H5/Me	H2
G1	5.95	2.76	2.58	4.83	4.21	3.68	7.93		
C2	5.45	2.31	1.99	4.84	4.41	4.14	7.40	5.43	
A3	5.89	2.95	2.78	5.07	4.42	4.14/4.03	8.24		7.25
A4	5.94	2.90	2.67	5.09	4.51	4.26/4.14	8.13		7.44
A5	6.13	2.76	2.49	4.96	4.23	4.13/4.13	8.14		7.35
T6	5.13	2.08	1.53	4.46			6.87	1.29	
T7	5.04	2.08	1.61	4.47	3.92		7.01	1.42	
T8	4.70	1.97	1.63	3.92	3.61	3.44	6.91	1.51	
G9	5.93	2.72	2.64	5.02	4.42	4.15	7.90		
C10	6.23	2.21	2.17	4.50	4.25	4.11	7.49	5.50	
G11	5.95	2.75	2.57	4.84	4.22	3.69	7.94		
C12	5.40	2.20	1.86	4.83	4.37	4.10	7.35	5.43	
A13	5.88	2.93	2.81	5.08	4.42	4.12/4.02	8.25		
A14	6.00	2.82	2.72	5.10	4.50	4.27/4.13	8.16		7.48
A15	5.54	2.49	2.20	4.74	4.16	3.65/3.50	8.03		7.51
T16	5.18	1.99	1.57	4.36			6.89	1.19	
T17	5.30	2.21	1.70	4.49	4.36	3.96	7.11	1.42	
T18	5.68	2.33	1.86	4.37	4.18	4.03	7.06	1.56	
G19	5.91	2.72	2.64	5.01	4.42	4.17	7.89		
C20	6.22	2.21	2.17	4.51	4.27	4.10	7.50	5.51	

Appendix 1.6 ^1H chemical shifts of the d(TTAGGGT)₄ quadruplex at 298K.

	H1'	H2'	H2''	H3'	H4'	H5'/H5''	H6/H8	/H5/Me	H2	NH
T1	6.00	2.10	2.34	4.64	4.00	3.65	7.41	1.67		
T2	6.25	2.06	2.34	4.74	4.06	3.93	7.33	1.78		
A3	6.28	2.86	2.92	5.10	4.44	4.16/ 4.10	8.43		8.09	
G4	6.01	2.67	2.91	5.05	4.49	4.27	7.95			11.23
G5	6.03	2.66	2.74	5.04	4.51	4.30	7.79			11.45
G6	6.27	2.57	2.70	4.91	4.52	4.27	7.70			11.84
T7	6.07	2.17	2.19	4.49	4.23	4.07	7.36	1.63		

Appendix 1.7 ^1H chemical shifts for the DNA quadruplex of the 2:1 d(TTAGGGT)₄–RHPS4 complex at 318K.

	H1'	H2'	H2''	H3'	H4'	H5'/H5''	H6/H8	H5/Me	NH
T1	5.98	2.14	2.38	4.71	4.05	3.70	7.39	1.63	
T2	6.11	2.00	2.30	4.79	4.42	4.15	7.32	1.71	
A3	6.22	2.79	2.82	5.10	4.43	4.16	8.29		
G4	5.94	2.59	2.86	5.00	4.46	4.21	7.85		10.11
G5	5.93	2.57	2.70	5.01	4.46	4.29	7.58		10.80
G6	6.30	2.68	2.71	5.10	4.47	4.27	7.74		11.05
T7	6.27	2.34	2.36	4.61		4.23	7.70	1.88	

Appendix 1.8 ^1H chemical shifts for H33258 molecule of the 2:1 d(GAAAAGCTTTTC)₂-H33258 complex at 298K.

Pip H2/H6	3.33/4.29	bz2 H4	7.46
Pip H3/H5	3.13/3.73	bz2 H6	7.98
bz1 H6	7.16	bz2 H7	8.22
bz1 H7	7.76	bz2 H3	11.63
bz1 H3	11.42	ph H2/H6	7.42
bz1 H4	8.27	ph H3/H5	8.136
pip NMe	2.94		

Appendix 1.9 ^1H chemical shifts for H10 molecule of the 1:1 d(CGCAAATTTGCG)₂-H10 complex at 298K.

OMe	4.37	bz1 H6	8.00
pip NMe	3.08	bz1 H7	8.13
bz2 H6	7.00	bz1 H4	8.37
bz2 H7	7.77	bz1 H3	11.40
bz2 H4	8.33	pip H2/H6	3.51/4.39
bz2 H3	12.13	pip H3/H5	3.76/4.29
ph H3	8.15	ph H2	7.60

Appendix 1.10 ^1H chemical shifts for RHPS4 of the 2:1 d(TTAGGGT)₄-RHPS4 complex at 318K.

H5	7.09	H2	6.87
H7	7.27	H1	7.24
H9	7.30	Me6	2.36
H10	7.26	Me8	3.58
H12	7.28	Me13	3.88
H4	7.76		

Appendix 2

Appendix 2.1 NOE restraints for the d(CTTTTGCAAAAG)₂ duplex structure, for one strand only.

residue	atom	residue	atom	distance	residue	atom	residue	atom	distance
T2	H7	C1	H6	4.27	C1	H3'	C1	H6	4.73
T2	H7	T2	H6	3.44	C1	H3'	T2	H6	3.79
T3	H7	T3	H6	3.34	C1	H4'	C1	H6	4.45
T4	H7	T4	H6	3.30	C1	H5'2	C1	H6	4.65
T5	H7	T5	H6	3.52	C1	H5'1	C1	H6	4.24
T3	H7	T3	H1'	5.49	A8	H2'1	A8	H8	2.70
T2	H7	C1	H3'	3.67	C7	H2'1	A8	H8	3.88
T2	H2'2	T2	H6	3.73	A8	H2'2	A8	H8	3.48
T3	H2'2	T3	H6	3.27	A9	H2'2	A9	H8	3.67
C1	H2'1	C1	H6	3.25	A9	H2'1	A9	H8	2.75
C1	H2'2	C1	H6	4.09	A8	H2'1	A9	H8	5.46
T4	H2'2	T4	H6	3.36	C7	H2'2	A8	H8	4.30
T2	H2'1	T2	H6	2.87	A10	H2'1	A10	H8	3.06
T3	H2'1	T3	H6	2.57	A11	H2'1	A11	H8	3.16
T4	H2'1	T4	H6	3.10	G6	H2'2	G6	H8	3.90
T5	H2'2	T5	H6	4.44	T5	H2'2	G6	H8	4.07
C7	H2'1	C7	H6	3.14	A11	H2'2	G12	H8	3.57
T5	H2'1	T5	H6	3.12	A11	H2'1	G12	H8	4.32
T4	H2'1	T5	H6	4.65	T5	H2'1	G6	H8	4.51
C7	H2'2	C7	H6	4.58	A10	H2'2	A10	H8	3.71
C1	H4'	T2	H6	4.77	G6	H2'1	G6	H8	3.14
T4	H2'2	T5	H6	3.41	A8	H1'	A8	H8	4.77
T2	H3'	T2	H6	4.73	C7	H1'	A8	H8	5.24
T3	H3'	T3	H6	4.59	A8	H3'	A8	H8	5.63
T5	H3'	T5	H6	4.63	A9	H3'	A9	H8	5.30
C1	H1'	C1	H6	4.51	A10	H3'	A10	H8	4.85
T2	H1'	T2	H6	4.80	A11	H3'	A11	H8	4.69
T2	H1'	T3	H6	4.74	G6	H3'	G6	H8	5.37
T3	H1'	T3	H6	4.74	A10	H1'	A10	H8	5.42
T3	H1'	T4	H6	4.60	A9	H1'	A10	H8	4.54
T4	H1'	T4	H6	4.76	A8	H1'	A9	H8	5.58
T4	H1'	T5	H6	5.33	A9	H1'	A9	H8	5.25
T5	H1'	T5	H6	4.87	A11	H1'	A11	H8	5.53
C1	H1'	T2	H6	6.00	T5	H1'	G6	H8	5.55
T5	H6	G6	H8	4.04	G6	H1'	G6	H8	5.42
G12	H4'	G12	H8	4.70	T3	H4'	T3	H1'	3.10
A11	H4'	A11	H8	5.70	T4	H4'	T4	H1'	2.82
A10	H4'	A10	H8	5.70	T5	H2'1	T5	H3'	2.56
G6	H2'2	G6	H8	3.56	T4	H2'1	T4	H3'	2.61
G12	H2'2	G12	H8	3.50	T5	H2'2	T5	H3'	2.96

residue	atom	residue	atom	distance	residue	atom	residue	atom	distance
A9	H3'	A9	H1'	4.00	A11	H2'1	A11	H3'	2.73
T2	H3'	T2	H1'	4.25	G6	H2'1	G6	H3'	3.18
T3	H3'	T3	H1'	3.89	A9	H2'1	A9	H3'	2.30
A8	H3'	A8	H1'	4.41	A8	H2'1	A8	H3'	2.54
G6	H3'	G6	H1'	4.75	A11	H2'2	A11	H3'	2.94
T5	H3'	T5	H1'	4.62	A10	H2'2	A10	H3'	3.18
T4	H3'	T4	H1'	4.29	G6	H2'2	G6	H3'	2.57
C7	H2'1	C7	H1'	3.53	G12	H4'	G12	H1'	4.82
T5	H2'1	T5	H1'	3.11	A11	H4'	A11	H1'	3.18
T4	H2'1	T4	H1'	3.53	A10	H4'	A10	H1'	3.26
C7	H2'2	C7	H1'	2.54	G6	H4'	G6	H1'	3.34
G12	H2'1	G12	H1'	2.61	A9	H4'	A9	H1'	3.32
G12	H2'2	G12	H1'	2.52	A8	H4'	A8	H1'	3.49
C1	H2'1	C1	H1'	3.31	C1	H4'	C1	H1'	3.39
C1	H2'2	C1	H1'	2.38	T4	H2'1	T4	H4'	4.95
T4	H2'2	T4	H1'	2.42	C7	H2'1	C7	H4'	4.29
A11	H2'2	A11	H1'	2.58	C7	H2'2	C7	H4'	3.66
A11	H2'1	A11	H1'	2.98	C1	H2'1	C1	H3'	2.75
A10	H2'2	A10	H1'	2.51	A8	H2'1	A8	H2'2	2.45
A10	H2'1	A10	H1'	3.03	C1	H2'2	C1	H3'	4.70
A9	H2'2	A9	H1'	2.50	G12	H2'2	G12	H4'	3.62
A9	H2'1	A9	H1'	2.72	G12	H2'1	G12	H4'	2.90
A8	H2'2	A8	H1'	2.16	C1	H5'1	C1	H3'	4.57
A8	H2'1	A8	H1'	2.79	C1	H4'	C1	H3'	3.25
T5	H2'2	T5	H1'	2.44	G12	H2'1	G12	H5'	3.61
A8	H4'	A8	H3'	2.85	C1	H2'1	C1	H4'	4.32
A11	H4'	A11	H3'	2.76	T2	H2'1	T3	H4'	4.25
G6	H4'	G6	H3'	2.75	C1	H2'2	C1	H4'	2.96
T3	H4'	T3	H3'	2.63	C1	H5'2	C1	H4'	5.27
T2	H4'	T2	H1'	3.11	T4	H2'2	T4	H4'	4.95
A8	H2'2	A8	H4'	5.32	T5	H2'1	T5	H2'2	2.22
A11	H2'2	A11	H4'	3.74	C7	H2'1	C7	H2'2	2.23
A9	H2'2	A9	H4'	4.87	G6	H2'1	G6	H1'	3.55
A9	H2'1	A9	H4'	5.47	G6	H2'2	G6	H1'	2.47
A10	H2'1	A10	H4'	4.62	T2	H3	A23	H2	4.30
A11	H2'1	A11	H4'	4.49	T3	H3	A23	H2	4.26
G6	H2'1	G6	H4'	5.50	T5	H3	A20	H2	4.10
A11	H2'1	A11	H2'2	2.05	T3	H3	A22	H2	4.88
T2	H2'1	T2	H2'2	2.16	T4	H3	A21	H2	6.25
C1	H2'1	C1	H2'2	2.04	T4	H3	A22	H2	6.25
T4	H2'1	T4	H2'2	2.17	T5	H3	A21	H2	6.25
G6	H1	A20	H2	6.25					

Appendix 2.2 NOE restraints for the d(CAAAAGCTTTTG)₂ duplex structure, for one strand only.

residue	atom	residue	atom	distance	residue	atom	residue	atom	distance
G1	H8	G1	H1'	4.29	A2	H3'	A2	H2'1	2.64
A2	H8	G1	H1'	4.13	A2	H4'	A2	H2'1	4.19
G1	H8	G1	H2'1	2.94	A2	H3'	A2	H2'2	2.85
A2	H8	G1	H2'1	3.83	A2	H4'	A2	H2'2	4.84
G1	H8	G1	H2'2	3.82	A2	H1'	A2	H3'	4.43
A2	H8	G1	H2'2	3.71	A2	H1'	A2	H4'	2.83
G1	H8	G1	H3'	4.70	A2	H3'	A2	H4'	2.79
A2	H8	G1	H3'	5.34	G1	H1'	A2	H5'1	4.20
G1	H5'1	G1	H8	4.34	A2	H1'	A2	H5'1	5.08
G1	H1'	G1	H2'1	3.38	A2	H3'	A2	H5'1	3.19
G1	H2'2	G1	H2'1	2.89	A2	H3'	A2	H5'2	3.43
G1	H3'	G1	H2'1	2.99	A2	H4'	A2	H5'2	3.84
G1	H5'1	G1	H2'1	3.06	A3	H8	A2	H1'	4.84
G1	H1'	G1	H2'2	3.63	A3	H8	A2	H2'1	5.24
G1	H3'	G1	H2'2	2.74	A3	H8	A3	H1'	5.29
G1	H5'1	G1	H2'2	4.16	A4	H8	A3	H1'	3.39
G1	H1'	G1	H3'	4.88	A3	H8	A3	H2'1	2.78
G1	H1'	G1	H5'1	4.37	A3	H8	A3	H2'2	3.22
G1	H4'	G1	H5'1	3.16	A3	H8	A3	H3'	4.21
G1	H1'	A2	H5'1	3.30	A3	H8	A3	H4'	5.41
A2	H8	G1	H1'	4.14	A3	H8	A3	H5'1	5.61
A2	H8	G1	H2'1	3.83	A3	H1'	A3	H2'1	3.82
A2	H8	G1	H2'2	3.71	A3	H2'2	A3	H2'1	1.95
A2	H8	G1	H3'	5.34	A3	H3'	A3	H2'1	2.78
A2	H8	A2	H1'	4.02	A3	H1'	A3	H2'2	2.60
A3	H8	A2	H1'	4.84	A3	H1'	A3	H3'	3.75
A2	H8	A2	H2'1	2.61	A3	H3'	A3	H4'	3.95
A3	H8	A2	H2'1	3.85	A3	H3'	A3	H5'1	3.31
A2	H8	A2	H2'2	3.43	A3	H4'	A3	H5'1	2.22
A2	H8	A2	H3'	4.34	A4	H8	A3	H1'	3.39
A2	H8	A2	H4'	4.53	A4	H8	A4	H1'	4.13
A2	H8	A2	H5'1	5.76	A5	H8	A4	H1'	3.93
A2	H8	A2	H5'2	5.33	A4	H8	A4	H2'1	2.73
A2	H1'	A2	H2'1	2.95	A4	H8	A4	H2'2	3.17
A2	H2'2	A2	H2'1	2.79	A4	H8	A4	H3'	4.08
A4	H8	A4	H4'	6.03	G6	H8	A5	H2'2	3.18
A4	H5'1	A4	H8	4.78	G6	H8	A5	H3'	4.52
A4	H2'2	A4	H2'1	1.95	C7	H6	G6	H1'	3.38
A4	H3'	A4	H2'1	2.95	G6	H8	G6	H2'1	2.86
A4	H4'	A4	H2'1	3.54	C7	H5	G6	H2'1	4.52
A4	H1'	A4	H2'2	2.56	C7	H6	G6	H2'1	3.58
A4	H4'	A4	H2'2	5.22	C7	H5	G6	H2'2	3.82
A4	H1'	A4	H3'	3.78	C7	H6	G6	H2'2	3.17

residue	atom	residue	atom	distance	residue	atom	residue	atom	distance
A4	H1'	A4	H4'	2.40	C7	H6	G6	H3'	5.33
A4	H3'	A4	H4'	3.25	G6	H8	G6	H4'	5.87
A4	H3'	A4	H5'1	4.13	A5	H8	G6	H8	5.60
A4	H4'	A4	H5'1	2.45	G6	H1'	C7	H5	5.73
A5	H8	A4	H1'	3.93	G6	H8	C7	H5	4.24
A5	H8	A5	H1'	4.46	G6	H1'	G6	H8	4.24
G6	H8	A5	H1'	3.83	G6	H1'	G6	H2'1	2.77
G6	H8	A5	H2	6.50	G6	H2'2	G6	H2'1	2.29
A5	H8	A5	H2'1	2.70	G6	H3'	G6	H2'1	3.52
A5	H8	A5	H2'2	2.97	G6	H4'	G6	H2'1	4.36
G6	H8	A5	H2'2	3.18	G6	H1'	G6	H2'2	2.47
A5	H8	A5	H3'	4.07	G6	H3'	G6	H2'2	3.53
G6	H8	A5	H3'	4.52	G6	H1'	G6	H4'	3.13
A5	H8	A5	H4'	5.37	G6	H3'	G6	H4'	3.22
A5	H8	A5	H5'1	6.41	C7	H6	G6	H1'	3.38
A5	H8	G6	H8	5.60	C7	H5	G6	H2'1	4.52
A5	H1'	A5	H2	5.53	C7	H6	G6	H2'1	3.58
A5	H1'	A5	H2'1	3.89	C7	H5	G6	H2'2	3.82
A5	H2'2	A5	H2'1	2.73	C7	H6	G6	H2'2	3.17
A5	H3'	A5	H2'1	3.02	C7	H6	G6	H3'	5.33
A5	H1'	A5	H2'2	2.68	C7	H6	C7	H1'	3.85
A5	H4'	A5	H2'2	3.77	T8	H6	C7	H1'	4.43
A5	H1'	A5	H3'	4.45	C7	H6	C7	H2'1	3.27
A5	H1'	A5	H4'	3.13	T8	H6	C7	H2'1	3.46
A5	H3'	A5	H5'1	3.13	C7	H6	C7	H2'2	4.40
G6	H8	A5	H1'	3.83	C7	H6	C7	H3'	4.30
G6	H8	A5	H2	6.50	T8	H6	C7	H3'	4.12
C7	H6	C7	H4'	4.11	T10	H6	T9	H1'	5.02
G6	H1'	C7	H5	5.73	T10	H6	T10	H1'	5.10
G6	H8	C7	H5	4.24	T11	H6	T10	H1'	5.00
C7	H6	C7	H5	2.94	T10	H6	T10	H2'1	2.81
C7	H6	C7	H5'2	4.35	T10	H6	T10	H4'	3.87
C7	H2'1	C7	H6	5.08	T10	H3'	T10	H2'1	2.50
C7	H1'	C7	H2'1	3.14	T10	H3'	T10	H4'	3.35
C7	H2'2	C7	H2'1	2.22	T10	H4'	T10	H5'1	3.32
C7	H3'	C7	H2'1	2.75	T10	H1'	T10	H2'1	3.43
C7	H4'	C7	H2'1	3.56	T10	H1'	T10	H2'2	2.46
T8	H5'1	C7	H2'1	3.74	T10	H1'	T10	H3'	4.72
C7	H1'	C7	H2'2	2.65	T11	H6	T10	H1'	5.00
C7	H3'	C7	H2'2	3.33	C12	H6	T11	H3'	4.92
T8	H5'1	C7	H2'2	4.50	T11	H6	T11	H1'	5.44
C7	H1'	C7	H3'	3.36	C12	H6	T11	H1'	4.13
C7	H3'	C7	H4'	3.45	T11	H6	T11	H2'1	2.53
C7	H3'	C7	H5'2	5.57	C12	H6	T11	H2'1	3.98

residue	atom	residue	atom	distance	residue	atom	residue	atom	distance
C7	H3'	C7	H4'	3.45	T11	H6	T11	H2'1	2.53
C7	H3'	C7	H5'2	5.57	C12	H6	T11	H2'1	3.98
C7	H5	C7	H2'1	6.44	T11	H3'	T11	H6	3.98
T8	H6	C7	H1'	4.43	T11	H1'	T11	H4'	4.88
T8	H6	C7	H2'1	3.46	T11	H1'	T11	H5'1	6.20
T8	H6	C7	H3'	4.12	T11	H5'1	T11	H3'	3.91
T8	H2'2	T8	H6	2.79	C12	H6	T11	H3'	4.92
T8	H5'1	C7	H2'1	3.74	C12	H6	T11	H1'	4.13
T8	H5'1	C7	H2'2	3.78	C12	H6	T11	H2'1	3.98
T8	H1'	T8	H2'1	3.25	C12	H6	C12	H1'	5.05
T8	H2'2	T8	H2'1	2.30	C12	H6	C12	H2'1	3.61
T8	H1'	T8	H2'2	2.56	C12	H6	C12	H2'2	4.83
T9	H6	T9	H1'	5.45	C12	H6	C12	H5	4.64
T10	H6	T9	H1'	5.02	C12	H6	C12	H5'1	5.41
T9	H6	T9	H2'1	2.79	C12	H1'	C12	H2'1	5.22
T9	H3'	T9	H6	2.79	C12	H3'	C12	H2'1	2.90
T9	H2'2	T9	H2'1	2.23	C12	H5	C12	H2'1	4.51
T9	H1'	T9	H2'2	2.80	C12	H5'1	C12	H2'1	5.45
T9	H1'	T9	H3'	4.72	C12	H1'	C12	H3'	4.30
T9	H5'1	T9	H4'	3.05	C12	H3'	C12	H4'	4.41

Appendix 2.3 NOE restraints for the 2:1 d(CTTTTGCAAAAG)₂-H33258 complex structure, for one strand only.

residue	atom	residue	atom	distance	residue	atom	residue	atom	distance
C1	H1'	C1	H6	4.92	A10	H2'2	A11	H8	3.79
T2	H1'	T2	H6	5.02	T2	H2'2	T3	H6	3.43
T2	H1'	T3	H6	5.24	C7	H2'2	C7	H6	3.41
T3	H1'	T3	H6	4.22	C7	H2'1	C7	H6	2.75
T3	H1'	T4	H6	6.26	T3	H2'1	T4	H6	4.31
T4	H1'	T4	H6	6.13	T4	H2'2	T4	H6	3.93
T4	H1'	T5	H6	4.60	T3	H2'2	T4	H6	3.61
T5	H1'	T5	H6	4.44	G6	H2'2	C7	H6	4.74
A10	H1'	A10	H8	6.24	A11	H2'2	G12	H8	3.42
C7	H5	C7	H6	3.52	G12	H2'1	G12	H8	3.65
A9	H1'	A9	H8	5.36	T4	H2'2	T5	H6	3.47
A8	H1'	A9	H8	4.79	T4	H2'1	T5	H6	3.92
A9	H1'	A10	H8	4.48	T5	H2'1	T5	H6	2.68
A10	H1'	A11	H8	5.20	T5	H2'2	T5	H6	3.43
A11	H1'	A11	H8	6.24	A9	H3'	A9	H8	5.03
A11	H1'	G12	H8	4.84	A8	H3'	A8	H8	5.10
G12	H1'	G12	H8	5.88	T2	H3'	T2	H6	5.01
A9	H2'2	A9	H8	3.46	T2	H3'	T3	H6	4.08
A9	H2'2	A10	H8	3.39	T2	H4'	T2	H6	4.52
A9	H2'1	A10	H8	4.00	C7	H3'	C7	H6	4.29
C1	H2'2	C1	H6	4.93	C1	H3'	T2	H6	4.21
C1	H2'1	C1	H6	3.69	C1	H3'	C1	H6	3.93
C7	H2'2	A8	H8	4.44	C1	H4'	C1	H6	5.31
A10	H2'1	A10	H8	3.60	C1	H5'1	C1	H6	4.40
C7	H2'1	A8	H8	4.61	C1	H5'2	C1	H6	4.34
A10	H2'2	A10	H8	4.50	A10	H3'	A10	H8	4.78
C1	H2'1	T2	H6	3.80	T3	H3'	T3	H6	4.51
T2	H2'2	T2	H6	3.50	T3	H4'	T3	H6	4.80
T5	H2'2	G6	H8	4.90	A11	H3'	G12	H8	6.00
T5	H2'1	G6	H8	5.50	A11	H3'	A11	H8	5.05
T3	H2'1	T3	H6	3.33	T5	H3'	T5	H6	4.02
A11	H2'1	A11	H8	3.78	T4	H3'	T4	H6	4.50
A10	H2'1	A11	H8	4.04	G12	H3'	G12	H8	4.91
A11	H2'2	A11	H8	4.80	G12	H5'2	G12	H8	5.01
T3	H2'2	T3	H6	3.69	T5	H6	T4	H6	6.26
A9	H8	A8	H8	4.90	A11	H2'1	A11	H3'	3.01
G12	H8	G12	H2'2	4.07	A11	H2'2	A11	H3'	3.20
G12	H8	A11	H2'1	4.48	T5	H2'1	T5	H1'	3.65
T4	H6	T3	H3'	6.30	T5	H2'2	T5	H3'	3.45
C7	H6	G6	H1'	5.56	T5	H2'1	T5	H3'	3.25
C1	H6	C1	H5	3.07	T4	H2'1	T4	H3'	2.90

residue	atom	residue	atom	distance	residue	atom	residue	atom	distance
A8	H8	A8	H1'	4.82	G12	H2'1	G12	H3'	3.32
C1	H2'2	C1	H1'	2.75	G12	H2'2	G12	H3'	3.24
A9	H2'1	A9	H1'	3.60	T4	H2'2	T4	H3'	3.29
A9	H2'2	A9	H1'	2.98	A9	H4'	A9	H3'	2.77
A10	H2'1	A10	H1'	3.32	A9	H3'	A9	H1'	4.78
A10	H2'2	A10	H1'	3.06	A9	H4'	A9	H1'	3.69
T3	H2'2	T3	H1'	4.00	A9	H5'2	A9	H4'	2.48
C1	H2'1	C1	H2'2	2.24	A9	H5'1	A9	H4'	3.37
T3	H2'1	T3	H1'	3.47	A8	H4'	A8	H3'	2.90
C7	H2'2	C7	H1'	3.11	A8	H5'2	A8	H3'	2.87
C7	H2'1	C7	H1'	4.03	A8	H5'1	A8	H3'	3.40
T4	H2'1	T4	H1'	3.83	A8	H5'1	A8	H5'2	2.49
A11	H2'1	A11	H1'	3.63	A8	H3'	A8	H1'	5.01
T4	H2'2	T4	H1'	3.05	A8	H4'	A8	H1'	4.58
A11	H2'2	A11	H1'	2.96	A8	H5'1	A8	H4'	2.36
A9	H2'1	A9	H3'	3.00	T3	H4'	T3	H3'	2.72
A9	H2'2	A9	H3'	3.21	T2	H4'	T2	H3'	2.77
G12	H2'2	G12	H1'	3.33	T2	H3'	T2	H1'	4.82
G12	H2'1	G12	H1'	2.85	T2	H4'	T2	H1'	3.33
T2	H2'2	T2	H3'	2.54	T2	H5'2	T2	H4'	3.60
C1	H2'1	C1	H3'	3.12	T2	H2'2	T2	H4'	3.24
C1	H2'2	C1	H3'	3.69	T2	H2'1	T2	H4'	4.96
A10	H2'2	A10	H3'	3.28	C7	H4'	C7	H1'	3.65
T3	H2'2	T3	H3'	3.47	C7	H4'	C7	H3'	3.00
A10	H2'1	A10	H3'	2.93	C7	H2'2	C7	H4'	4.66
C7	H2'2	C7	H3'	4.19	C7	H2'1	C7	H4'	5.14
T3	H2'1	T3	H3'	3.17	C1	H5'1	C1	H3'	3.24
C7	H2'1	C7	H3'	3.50	C1	H5'2	C1	H3'	2.96
T5	H2'2	T5	H1'	3.10	C1	H5'1	C1	H4'	2.76
C1	H5'2	C1	H4'	2.75	T2	H2'2	T2	H1'	2.85
C1	H4'	C1	H3'	3.18	T2	H1'	T2	H2'1	3.11
C1	H3'	C1	H1'	3.96	C1	H2'1	C1	H1'	3.22
C1	H5'1	C1	H5'2	2.40	G6	H3'	G6	H2'2	3.64
C1	H4'	C1	H1'	3.37	C7	H1'	C7	H3'	4.20
T2	H2'1	T2	H2'2	2.28	A11	H5'1	A11	H5'2	2.18
A10	H5'2	A10	H3'	3.60	T4	H7	T3	H6	3.71
A10	H2'1	A10	H2'2	2.22	T3	H7	T3	H6	3.03
T3	H3'	T3	H1'	3.45	T3	H7	T2	H1'	4.76
T3	H4'	T2	H1'	5.25	T2	H7	C1	H5	3.99
A11	H4'	A11	H3'	3.18	T2	H7	C1	H1'	2.93
A11	H3'	A11	H1'	4.70	T4	H7	T3	H1'	6.00
T4	H4'	T4	H3'	3.15	T3	H7	T2	H3'	5.16
G12	H4'	G12	H3'	2.94	T2	H7	C1	H3'	4.12
G6	H3'	G6	H1'	4.44	T4	H7	T3	H3'	3.81
G6	H4'	G6	H1'	4.57	T2	H7	C1	H2'1	4.26
A11	H2'1	A11	H2'2	2.34	T2	H7	C1	H2'2	4.28
C7	H2'1	C7	H2'2	2.55	T3	H7	T2	H2'2	4.41

residue	atom	residue	atom	distance	residue	atom	residue	atom	distance
G12	H2'1	G12	H2'2	2.24	T3	H7	T2	H2'1	3.33
T4	H2'1	T4	H2'2	2.13	T4	H7	T3	H2'2	3.61
T3	H2'1	T3	H2'2	2.13	T5	H7	T4	H2'2	4.38
T5	H2'1	T5	H2'2	3.22	T4	H7	T3	H2'1	3.24
A8	H2	A9	H2	3.75	T5	H7	T4	H2'1	3.14
A9	H2	A10	H2	3.75	T2	H7	T2	H2'1	3.76
T4	H1'	A22	H2	4.28	T2	H7	C1	H6	3.60
A11	H5'1	A11	H3'	3.25	T3	H7	T2	H6	3.41
A11	H5'2	A11	H3'	4.69	T2	H7	T2	H6	3.84
A11	H5'1	A11	H5'2	2.52	T4	H6	T4	H7	3.49
G12	H5'2	G12	H1'	3.53	T4	H6	T5	H7	4.35
T17	H3	A9	H2	4.79	T15	H3	A10	H2	3.95
T17	H3	A8	H2	3.93	T15	H3	A9	H2	5.00
T16	H3	A10	H2	4.45	T5	H3	T4	H3	4.52
T16	H3	A9	H2	3.92	G6	H1	C7	H4'	3.88
T14	H3	A11	H2	3.90	G6	H1	C7	H6	4.77

HOE26	pip H6	A9	H2	4.43	HOE25	pip H6	A21	H2	4.43
T14	H1'	HOE26	ph H6	3.74	T2	H1	HOE25	ph H6	3.74
T15	H1'	HOE26	ph H6	4.89	T3	H1	HOE25	ph H6	4.89
G12	H1'	HOE26	ph H6	4.64	G24	H1	HOE25	ph H6	4.64
HOE26	pip NMe	A8	H2	4.45	HOE25	pip NMe	A20	H2	4.45
HOE26	pip H6	A8	H2	3.70	HOE25	pip H6	A20	H2	3.70
HOE26	ph H5	A11	H2	3.22	HOE25	ph H5	A23	H2	3.12
A10	H2	HOE26	bz1 H4	2.91	A22	H2	HOE25	bz1 H4	2.91
A10	H2	HOE26	ph H5	5.1	A22	H2	HOE25	ph H5	5.09
HOE26	ph H6	A11	H2	4.67	HOE25	ph H6	A23	H2	4.67
HOE26	bz2 H4	A9	H2	4.07	HOE25	bz2 H4	A21	H2	4.07
A9	H1'	HOE26	bz2 H4	5.96	A21	H1	HOE25	bz2 H4	5.96
T15	H1'	HOE26	ph H5	3.81	T3	H1	HOE25	ph H5	3.81
TT16	H1'	HOE26	bz1 H4	3.73	T4	H1	HOE25	bz1 H4	3.73
G12	H1'	HOE26	ph H5	4.63	G24	H1	HOE25	ph H5	4.63
HOE26	pip H6	A9	H1'	3.47	HOE25	pip H6	A21	H1'	3.47
HOE26	pip NMe	A9	H1'	4.65	HOE25	pip NMe	A21	H1'	4.65
HOE26	pip H5	A8	H2	3.91	HOE25	pip H5	A20	H2	3.91
HOE26	pip H6	A9	H2	4.18	HOE25	pip H6	A21	H2	4.18
HOE26	pip H6	A9	H1'	4.13	HOE25	pip H6	A21	H1'	4.13
HOE26	bz2 H3	A10	H2	4.41	HOE25	bz2 H2	A22	H2	4.41
HOE26	bz2 H3	A9	H2	4.96	HOE25	bz2 H3	A21	H2	4.96
HOE26	bz1 H3	A11	H2	4.19	HOE25	bz1 H3	A23	H2	4.19
HOE26	bz1 H3	A10	H2	4.62	HOE25	bz1 H3	A22	H2	4.62
T16	H3	HOE26	bz2 H4	5.77	T4	H3	HOE25	bz2 H4	5.77
T16	H3	HOE26	bz1 H4	5.16	T4	H3	HOE25	bz1 H4	5.16
T15	H3	HOE26	ph H5	4.90	T3	H3	HOE25	ph H5	4.90
T14	H3	HOE26	ph H6	4.92	T2	H3	HOE25	ph H6	4.92

Appendix 2.4 NOE restraints for the 1:1 d(GCAAATTTGC)₂-H10 complex structure.

residue	atom	residue	atom	distance	residue	atom	residue	atom	distance
G1	H1'	C2	H6	4.40	G11	H2'2	C12	H6	4.05
G1	H2'2	C2	H6	3.22	C12	H1'	C12	H6	3.35
C2	H1'	C2	H6	3.98	C12	H2'1	C12	H6	2.81
C2	H2'1	C2	H6	2.65	C12	H2'2	C12	H6	3.68
C2	H2'2	C2	H6	3.25	C12	H3'	C12	H6	4.98
C2	H3'	C2	H6	4.45	C12	H4'	C12	H6	4.10
C2	H5	C2	H6	2.78	C12	H5	C12	H6	2.84
C2	H5'1	C2	H6	4.94	C12	H5'1	C12	H6	5.88
A5	H1'	T6	H6	4.01	A15	H1'	T16	H6	5.41
A5	H2'1	T6	H6	3.70	A15	H2'1	T16	H6	3.97
A5	H2'2	T6	H6	3.82	T16	H1'	T16	H6	5.15
T6	H1'	T6	H6	4.8	T16	H1'	T17	H6	4.51
T6	H1'	T7	H6	5.86	T16	H2'1	T17	H6	3.00
T6	H2'2	T6	H6	3.77	T16	H2'2	T17	H6	2.86
T6	H3'	T6	H6	5.25	T16	H3'	T16	H6	4.52
T6	H6	A5	H8	5.48	T16	H6	A15	H8	5.59
T6	H6	T7	H6	5.84	T16	H6	T17	H6	5.12
T7	H1'	T7	H6	5.06	T17	H1'	T17	H6	4.90
T7	H1'	T8	H6	5.16	T17	H1'	T18	H6	4.50
T7	H2'1	T7	H6	2.72	T17	H2'1	T17	H6	2.59
T7	H2'2	T7	H6	3.70	T17	H2'1	T18	H6	2.97
T7	H2'2	T8	H6	3.00	T17	H2'2	T17	H6	3.60
T8	H1'	T8	H6	4.25	T17	H2'2	T18	H6	2.65
T8	H2'1	T8	H6	2.85	T17	H3'	T17	H6	4.99
T8	H3'	T8	H6	5.84	T17	H4'	T17	H6	4.28
T8	H6	T7	H6	5.18	T18	H1'	T18	H6	5.14
T8	H6	G9	H8	3.30	T18	H2'1	T18	H6	2.70
G9	H1'	C10	H6	4.41	T18	H2'2	T18	H6	3.61
G9	H2'2	C10	H6	4.63	T18	H6	G19	H8	5.50
G9	H2'1	C10	H6	3.45	G19	H1'	C20	H6	4.89
C10	H1'	C10	H6	3.55	G19	H2'1	C20	H6	2.80
C10	H3'	C10	H6	3.65	G19	H2'2	C20	H6	3.46
C10	H5	C10	H6	2.82	C20	H1'	C20	H6	4.03
G11	H1'	C12	H6	5.69	C20	H3'	C20	H6	3.75
G11	H2'1	C12	H6	4.25	C20	H5	C20	H6	2.82
G1	H1'	G1	H8	4.65	A13	H3'	A13	H8	3.96
G1	H2'1	G1	H8	2.63	A13	H4'	A13	H8	5.52
G1	H2'2	G1	H8	3.25	A14	H1'	A14	H8	4.77
G1	H3'	G1	H8	4.56	A14	H1'	A15	H8	3.64
C2	H1'	A3	H8	4.58	A14	H2'1	A15	H8	3.31
C2	H2'1	A3	H8	4.61	A14	H2'2	A15	H8	2.74
C2	H2'2	A3	H8	3.37	A14	H3'	A14	H8	4.36
C2	H5	G1	H8	2.85	A14	H4'	A14	H8	4.84

residue	atom	residue	atom	distance	residue	atom	residue	atom	distance
A3	H1'	A3	H8	4.58	A15	H1'	A15	H8	5.11
A3	H1'	A4	H8	4.22	A15	H2'1	A15	H8	2.90
A3	H2'2	A3	H8	4.29	A15	H2'2	A15	H8	3.74
A3	H2'1	A3	H8	3.39	A15	H3'	A15	H8	4.91
A3	H4'	A3	H8	5.52	A15	H4'	A15	H8	5.39
A4	H2'2	A4	H8	3.20	T16	H6	A15	H8	5.18
A4	H3'	A4	H8	5.06	T18	H1'	G19	H8	4.96
A4	H4'	A4	H8	4.67	T18	H2'1	G19	H8	3.66
A5	H1'	A5	H8	4.85	T18	H2'2	G19	H8	3.30
A5	H2'1	A5	H8	2.59	T18	H6	G19	H8	4.19
A5	H3'	A5	H8	4.60	G19	H3'	G19	H8	5.18
T6	H6	A5	H8	4.57	T8	H2'2	T8	H6	4.42
T8	H1'	G9	H8	5.06	T16	H2'2	T16	H6	3.64
T8	H2'1	G9	H8	3.47	A4	H1'	A4	H8	4.51
T8	H2'2	G9	H8	3.66	G1	H2'1	G1	H3'	3.45
T8	H6	G9	H8	4.65	G1	H2'1	G1	H4'	4.03
G9	H3'	G9	H8	4.03	G1	H2'2	G1	H3'	3.54
G11	H2'1	G11	H8	2.79	G1	H2'2	C2	H5	4.65
G11	H2'2	G11	H8	3.25	G1	H3'	G1	H1'	4.01
C12	H1'	A13	H8	4.57	G1	H4'	G1	H1'	3.64
C12	H2'1	A13	H8	4.30	G1	H4'	G1	H3'	2.67
C12	H2'2	A13	H8	4.99	G1	H5'1	G1	H1'	5.50
A13	H1'	A13	H8	4.75	G1	H5'1	G1	H3'	2.83
A13	H1'	A14	H8	4.59	C2	H2'1	C2	H1'	4.28
A13	H2'1	A13	H8	3.02	C2	H2'1	C2	H2'2	2.25
A13	H2'2	A13	H8	3.35	C2	H2'1	C2	H3'	3.85
A13	H3'	A13	H8	3.96	C2	H2'1	C2	H5	6.07
C2	H2'1	C2	H5'1	5.13	T7	H2'1	T7	H2'2	2.37
C2	H2'2	C2	H1'	2.48	T7	H2'2	T7	H1'	2.66
C2	H2'2	C2	H3'	3.58	T8	H2'1	T8	H1'	3.12
C2	H2'2	C2	H4'	5.20	T8	H2'2	T8	H1'	2.56
C2	H2'2	C2	H5'1	6.01	T8	H3'	T8	H1'	3.54
C2	H5	G1	H1'	3.39	T8	H4'	T8	H1'	3.73
C2	H5'1	C2	H1'	4.60	T8	H4'	T8	H3'	3.76
A3	H2'1	A3	H1'	3.13	T8	H5'1	T8	H3'	3.00
A3	H2'1	A3	H2'2	2.33	G9	H4'	G9	H1'	3.05
A3	H3'	A3	H1'	3.61	C10	H2'1	C10	H1'	4.05
A3	H4'	A3	H3'	2.89	C10	H2'2	C10	H1'	3.18
A3	H5'2	A3	H3'	3.18	C10	H3'	C10	H1'	3.40
A4	H1'	A3	H2	4.39	G11	H2'1	G11	H3'	2.70
A4	H2'1	A4	H2'2	2.55	G11	H2'1	G11	H4'	3.87
A4	H2'2	A4	H1'	2.73	G11	H2'2	G11	H1'	3.67
A4	H2'2	A4	H4'	4.70	G11	H2'2	G11	H3'	3.42

residue	atom	residue	atom	distance	residue	atom	residue	atom	distance
A4	H2'2	A4	H5'1	5.00	G11	H2'2	C12	H5	5.33
A4	H3'	A4	H1'	4.82	G11	H3'	G11	H1'	4.01
A4	H4'	A4	H1'	4.64	G11	H4'	G11	H1'	3.64
A4	H4'	A4	H3'	2.92	G11	H4'	G11	H3'	2.67
A5	H2'1	A5	H1'	3.79	G11	H5'1	G11	H1'	5.50
A5	H2'1	A5	H2'2	2.40	C12	H2'1	C12	H1'	4.28
A5	H2'1	A5	H3'	2.81	C12	H2'1	C12	H2'2	2.47
A5	H2'2	A5	H1'	2.66	C12	H2'1	C12	H3'	3.67
A5	H2'2	A5	H3'	3.74	C12	H2'1	C12	H5'1	5.79
A5	H3'	A5	H1'	4.86	C12	H2'2	C12	H1'	2.97
A5	H4'	A5	H1'	5.18	C12	H2'2	C12	H3'	4.21
A5	H4'	A5	H3'	3.56	C12	H5'1	C12	H1'	4.81
A5	H5'1	A5	H3'	4.85	A13	H2'1	A13	H1'	3.13
T6	H2'1	T6	H1'	3.89	A13	H2'1	A13	H2'2	2.34
T6	H2'1	T6	H2'2	2.43	A13	H2'2	A13	H5'1	5.99
T6	H2'1	T6	H3'	2.96	A13	H3'	A13	H1'	3.61
T6	H2'2	T6	H1'	2.89	A13	H4'	A13	H3'	2.89
T6	H3'	T6	H1'	3.68	A14	H2'1	A14	H1'	3.60
T7	H2'1	T7	H1'	3.61	A14	H2'2	A14	H1'	2.70
A14	H2'2	A14	H4'	4.11	T7	H4'	T7	H3'	3.82
A14	H2'2	A14	H5'1	5.50	C10	H4'	C10	H3'	3.58
A14	H3'	A14	H1'	4.41	A14	H2'1	A14	H2'2	2.67
A14	H4'	A14	H1'	4.24	A14	H2'1	A14	H4'	5.24
A14	H4'	A14	H3'	2.92	T16	H2'1	T16	H1'	3.73
A14	H5'1	A14	H1'	5.31	T16	H2'1	T16	H2'2	2.19
A15	H2'1	A15	H1'	3.35	C20	H4'	C20	H3'	2.75
A15	H2'1	A15	H2'2	2.43	C2	H2'2	C2	H1'	2.79
A15	H2'1	A15	H3'	3.36	A13	H2'2	A13	H1'	2.96
A15	H2'2	A15	H1'	3.14	T6	H7	A5	H2'1	4.43
A15	H2'2	A15	H3'	3.50	T6	H7	A5	H2'2	4.72
A15	H4'	A15	H3'	3.54	T6	H7	A5	H3'	5.50
A15	H5'1	A15	H3'	3.25	T6	H7	A5	H8	4.47
A15	H5'1	A15	H4'	3.48	T6	H7	T6	H6	3.60
T16	H2'1	T16	H3'	3.02	T7	H7	T6	H1'	5.78
T16	H2'2	T16	H1'	2.67	T7	H7	T6	H2'2	4.85
T16	H2'2	T16	H3'	3.73	T7	H7	T6	H3'	3.48
A14	H2'2	A14	H4'	4.11	T7	H4'	T7	H3'	3.82
A14	H2'2	A14	H5'1	5.50	C10	H4'	C10	H3'	3.58
A14	H3'	A14	H1'	4.41	A14	H2'1	A14	H2'2	2.67
A14	H4'	A14	H1'	4.24	A14	H2'1	A14	H4'	5.24
A14	H4'	A14	H3'	2.92	T16	H2'1	T16	H1'	3.73
A14	H5'1	A14	H1'	5.31	T16	H2'1	T16	H2'2	2.19
A15	H2'1	A15	H1'	3.35	C20	H4'	C20	H3'	2.75
A15	H2'1	A15	H2'2	2.43	C2	H2'2	C2	H1'	2.79
A15	H2'1	A15	H3'	3.36	A13	H2'2	A13	H1'	2.96
A15	H2'2	A15	H1'	3.14	T6	H7	A5	H2'1	4.43

residue	atom	residue	atom	distance	residue	atom	residue	atom	distance
A15	H2'2	A15	H3'	3.50	T6	H7	A5	H2'2	4.72
A15	H4'	A15	H3'	3.54	T6	H7	A5	H3'	5.50
A15	H5'1	A15	H3'	3.25	T6	H7	A5	H8	4.47
A15	H5'1	A15	H4'	3.48	T6	H7	T6	H6	3.60
T16	H2'1	T16	H3'	3.02	T7	H7	T6	H1'	5.78
T16	H2'2	T16	H1'	2.67	T7	H7	T6	H2'2	4.85
T16	H2'2	T16	H3'	3.73	T7	H7	T6	H3'	3.48
T16	H3'	T16	H1'	3.64	T7	H7	T6	H6	3.67
T17	H2'1	T17	H1'	3.68	T7	H7	T7	H6	3.50
T17	H2'1	T17	H2'2	2.42	T8	H7	T7	H1'	6.14
T17	H2'1	T17	H3'	2.67	T8	H7	T8	H6	3.58
T17	H2'2	T17	H1'	3.06	T16	H7	A15	H2'1	4.04
T17	H3'	T17	H1'	3.88	T16	H7	A15	H2'2	4.69
T18	H2'1	T18	H1'	2.53	T16	H7	A15	H3'	5.27
T18	H2'1	T18	H2'2	2.35	T16	H7	A15	H8	4.28
T18	H2'2	T18	H1'	3.55	T16	H7	T16	H6	3.60
T18	H2'2	T18	H3'	3.95	T17	H7	T16	H1'	5.65
T18	H4'	T18	H1'	4.65	T17	H7	T16	H2'2	5.45
T18	H5'1	T18	H1'	5.57	T17	H7	T16	H3'	5.69
G19	H4'	G19	H3'	2.88	T17	H7	T17	H6	3.55
C20	H3'	C20	H1'	3.60	T18	H7	T17	H1'	5.78
G1	H2'1	G1	H2'2	2.24	T18	H7	T17	H2'1	4.67
A3	H2'2	A3	H1'	3.17	T18	H7	T17	H2'2	5.70
A4	H2'1	A4	H4'	4.64	T18	H7	T18	H6	3.70
T7	H2'1	T7	H3'	2.74	G11	H1'	G11	H8	4.72
C20	H6	G19	H8	4.98	C20	H6	G19	H8	5.18
A13	H1'	A14	H8	4.51					

A4	H2	H10	bz1 H4	5.50	H10	pip H2	A14	H1'	4.50
T17	H1'	H10	bz1 H4	4.47	H10	pip H2	A14	H2	4.50
T16	H1'	H10	bz1H4	4.11	H10	ph H2	A15	H2	3.32
A5	H2	H10	bz1 H4	5.00	A14	H1'	H10	ph H6	5.13
T17	H1'	H10	bz1 H7	3.99	A15	H1'	H10	ph H6	3.87
A5	H2	H10	bz1 H7	5.20	T16	H1'	H10	ph H6	5.04
A5	H1'	H10	bz1 H7	5.20	A14	H2	H10	ph H5	4.08
T6	H1'	H10	bz1 H7	3.99	A15	H1'	H10	ph H5	3.87
A4	H2	H10	bz1 H6	4.50	A15	H2	H10	ph H5	3.45
H10	OMe	A3	H2	5.50	T16	H1'	H10	ph H5	5.20
H10	OMe	A4	H2	4.45	T17	H1'	H10	bz2 H4	5.50
H10	OMe	A5	H1'	4.23	T16	H1'	H10	bz2 H4	4.45
H10	OMe	T17	H1'	5.50	A15	H2	H10	bz2 H4	5.00
H10	OMe	A4	H1'	4.20	A15	H2	H10	bz2 H3	3.40
H10	OMe	A5	H2	4.10	A14	H2	H10	bz2 H3	4.23
H10	pip H3	A14	H1'	4.50	A5	H2	H10	bz1 H3	4.36
H10	pip H3	A14	H2	4.50					

Appendix 2.5 NOE restraints for the d(TTAGGGT)₄ quadruplex structure, for one strand only.

residue	atom	residue	atom	distance	residue	atom	residue	atom	distance
T1	H6	T1	H1'	4.50	G4	H8	G4	H2'2	3.45
T2	H6	T1	H1'	6.33	G5	H8	G4	H2'2	3.37
T1	H6	T1	H2'1	2.74	G4	H8	G4	H3'	4.51
T1	H2'2	T1	H6	3.46	G4	H8	G4	H4'	5.68
T1	H6	T1	H3'	4.84	A3	H8	G4	H8	6.33
T1	H6	T1	H4'	6.33	G5	H8	G5	H1'	5.43
T2	H6	T2	H1'	5.29	G5	H8	G5	H2'1	3.24
T2	H6	T2	H2'1	3.32	G5	H8	G5	H2'2	3.42
T2	H6	T2	H2'2	4.16	G5	H8	G5	H3'	4.58
T2	H6	T2	H3'	5.37	G6	H8	G5	H3'	5.09
T2	H6	T2	H4'	6.33	G5	H8	G5	H4'	5.88
A3	H8	T2	H6	6.33	G4	H8	G5	H8	6.33
T7	H6	G6	H1'	6.33	G6	H8	G6	H1'	4.68
T7	H6	G6	H2'1	4.48	G6	H8	G6	H2'1	2.55
T7	H6	G6	H2'2	4.04	G6	H8	G6	H2'2	3.24
T7	H6	G6	H3'	5.28	G6	H8	G6	H3'	4.53
T7	H6	T7	H1'	4.19	G6	H8	G6	H4'	6.08
T7	H6	T7	H2'1	3.62	G6	H8	T7	H6	6.33
T7	H6	T7	H2'2	4.44	G6	H8	T7	H7	4.66
T7	H6	T7	H3'	4.23	T1	H7	T1	H1'	5.91
T7	H6	T7	H4'	6.13	T1	H6	T1	H7	3.77
G6	H8	T7	H6	6.33	T1	H1'	T2	H7	6.26
A3	H8	T2	H1'	5.72	T1	H3'	T2	H7	5.51
A3	H8	T2	H2'1	5.72	T2	H7	T2	H2'1	4.60
A3	H8	T2	H2'2	4.15	T2	H2'2	T2	H7	5.68
A3	H8	T2	H3'	6.33	T2	H3'	T2	H7	4.58
A3	H8	A3	H1'	5.32	T2	H6	T2	H7	3.53
G4	H8	A3	H1'	5.69	G6	H1'	T7	H7	5.84
A3	H8	A3	H2'1	3.41	G6	H2'1	T7	H7	5.23
A3	H8	A3	H2'2	3.81	G6	H2'2	T7	H7	5.27
A3	H8	A3	H3'	5.73	G6	H3'	T7	H7	5.20
A3	H8	A3	H4'	6.33	G6	H4'	T7	H7	6.32
G4	H8	G4	H1'	4.87	G6	H8	T7	H7	4.66
G5	H8	G4	H1'	6.22	T7	H1'	T7	H7	6.05
G4	H8	G4	H2'1	3.02	T7	H2'2	T7	H7	5.19
T7	H6	T7	H7	3.57	A3	H3'	A3	H4'	3.08
T1	H1'	T1	H2'1	3.63	A3	H3'	A3	H5'1	3.08
T1	H2'2	T1	H2'1	2.15	A3	H4'	A3	H5'1	2.88
T1	H3'	T1	H2'1	2.85	A3	H2	G4	H1'	4.90
T1	H2'1	T1	H3'	2.88	G4	H1'	G4	H2'1	3.30
T1	H4'	T1	H2'1	4.26	G4	H2'2	G4	H2'1	2.12
T1	H2'1	T1	H4'	4.60	G4	H3'	G4	H2'1	2.63
T1	H1'	T1	H2'2	2.90	G4	H1'	G4	H2'2	2.58
T1	H2'2	T1	H4'	4.75	G4	H3'	G4	H2'2	2.66

residue	atom	residue	atom	distance	residue	atom	residue	atom	distance
T1	H1'	T1	H3'	4.93	G4	H4'	G4	H2'2	3.91
T1	H1'	T1	H4'	3.84	G4	H1'	G4	H3'	4.26
T1	H3'	T1	H4'	3.40	G4	H1'	G4	H4'	3.86
T1	H3'	T1	H5'1	3.38	G4	H3'	G4	H4'	2.88
T1	H4'	T1	H5'1	3.21	G4	H3'	G4	H5'1	2.82
T2	H1'	T2	H2'1	3.59	G5	H1'	G5	H2'1	3.03
T2	H2'2	T2	H2'1	2.17	G5	H2'2	G5	H2'1	2.09
T2	H3'	T2	H2'1	2.80	G5	H3'	G5	H2'1	2.44
T2	H4'	T2	H2'1	3.94	G5	H4'	G5	H2'1	3.48
T2	H3'	T2	H2'2	3.41	G5	H1'	G5	H2'2	2.57
T2	H1'	T2	H2'2	2.92	G5	H3'	G5	H2'2	2.58
T2	H3'	T2	H2'2	3.19	G5	H4'	G5	H2'2	3.88
T2	H4'	T2	H2'2	3.87	G5	H1'	G5	H3'	3.83
T2	H1'	T2	H4'	3.37	G5	H1'	G5	H4'	4.38
T2	H3'	T2	H4'	3.74	G5	H3'	G5	H4'	2.83
T2	H3'	T2	H5'1	3.25	G5	H3'	G5	H5'1	2.70
T2	H4'	T2	H5'1	2.75	G6	H1'	G6	H2'1	2.98
A3	H2	A3	H1'	4.10	G6	H2'2	G6	H2'1	2.12
A3	H1'	A3	H2'1	3.45	G6	H3'	G6	H2'1	2.51
A3	H3'	A3	H2'1	2.59	G6	H4'	G6	H2'1	4.00
A3	H4'	A3	H2'1	5.00	G6	H1'	G6	H2'2	2.46
A3	H1'	A3	H2'2	2.74	G6	H3'	G6	H2'2	2.88
A3	H3'	A3	H2'2	3.18	G6	H4'	G6	H2'2	3.77
A3	H4'	A3	H2'2	4.27	G6	H1'	G6	H3'	4.11
A3	H1'	A3	H3'	4.16	G6	H1'	G6	H4'	3.36
A3	H1'	A3	H4'	4.53	G6	H3'	G6	H4'	2.88
G6	H3'	G6	H5'1	2.65	T7	H1'	T7	H3'	4.93
T7	H1'	T7	H2'1	3.27	T7	H1'	T7	H4'	3.87
T7	H3'	T7	H2'1	2.88	T7	H3'	T7	H4'	3.4
T7	H1'	T7	H2'2	2.45	T7	H5'1	T7	H3'	2.74
T7	H3'	T7	H2'2	3.34	T7	H4'	T7	H5'1	1.94
T7	H4'	T7	H2'2	4.00	A3	H2'2	A3	H2'1	2.14
A3	H62	A10	H2	5.06	G18	N7	G25	N2	2.95
A24	H61	G4	NH	4.71	G18	O6	G25	N1	2.95
G4	NH	A3	H2	5.65	G4	N2	G25	N7	2.95
G4	NH	A24	H8	5.64	G4	N1	G25	O6	2.95
G4	NH	G4	H21	4.58	G11	N2	G4	N7	2.95
G4	NH	G25	H8	5.00	G11	N1	G4	O6	2.95
G12	NH	G5	H8	5.13	G19	N2	G12	N7	2.95
G4	NH	G5	NH	4.19	G19	N1	G12	O6	2.95
G5	NH	G25	H8	4.54	G19	N7	G26	N2	2.95
G5	NH	G5	H21	4.85	G19	O6	G26	N1	2.95
G5	NH	G26	H8	5.00	G5	N2	G26	N7	2.95
G6	NH	G26	H8	5.88	G5	N1	G26	O6	2.95
G6	NH	G5	NH	4.36	G12	N2	G5	N7	2.95
G6	NH	G27	H8	4.67	G12	N1	G5	O6	2.95

residue	atom	residue	atom	distance	residue	atom	residue	atom	distance
G6	NH	G6	H21	4.63	G20	N2	G13	N7	2.95
G6	H22	G6	NH	4.25	G20	N1	G13	O6	2.95
G5	H22	G26	H8	4.25	G20	N7	G27	N2	2.95
G5	H22	G5	NH	4.25	G20	O6	G27	N1	2.95
G4	H22	G25	H8	4.25	G6	N2	G27	N7	2.95
G4	H22	G4	NH	4.25	G6	N1	G27	O6	2.95
G6	H22	G27	H8	4.25	G13	N2	G6	N7	2.95
G18	N2	G11	N7	2.95	G13	N1	G6	O6	2.95

Appendix 2.6 NOE restraints for the 1:1 d(TTAGGGT)₄–RHPS4 complex structure, for one strand only.

residue	atom	residue	atom	distance	distance
T1	H1'	T1	H6	2.50	4.00
T1	H2"	T1	H1'	1.80	3.00
T1	H2"	T1	H3'	2.50	4.00
T1	H2"	T1	H4'	3.50	5.50
T1	H2"	T1	H6	2.50	4.00
T1	H2"	T1	H5'	3.50	5.50
T1	H2'	T1	H1'	2.50	4.00
T1	H2'	T1	H2"	1.80	3.00
T1	H2'	T1	H3'	2.50	4.00
T1	H2'	T1	H4'	3.50	5.50
T1	H2'	T1	H5'	3.50	5.50
T1	H2'	T1	H6	1.80	3.00
T1	H3'	T1	H1'	3.50	5.50
T1	H3'	T1	H6	2.50	4.00
T1	H4'	T1	H1'	3.50	5.50
T1	H4'	T1	H3'	2.50	4.00
T1	H4'	T1	H6	3.50	5.50
T1	H5'	T1	H3'	2.50	4.00
T1	H5'	T1	H4'	2.50	4.00
T1	H5'	T1	H6	2.50	4.00
T1	Me	T1	H6	2.50	4.00
T2	H1'	T2	H6	2.50	4.00
T2	H2"	T2	H1'	1.80	3.00
T2	H2"	T2	H3'	1.80	3.00
T2	H2"	T2	H5"	3.50	5.50
T2	H2"	T2	H5'	3.50	5.50
T2	H2"	T2	H6	2.50	4.00
T2	H2"	A3	H8	2.50	4.00
T2	H2'	T2	H1'	2.50	4.00
T2	H2'	T2	H2"	1.80	3.00
T2	H2'	T2	H3'	1.80	3.00
T2	H2'	T2	H5"	3.50	5.50
T2	H2'	T2	H5'	3.50	5.50
T2	H2'	T2	H6	1.80	3.00
T2	H2'	A3	H8	3.50	5.50
T2	H3'	T2	H1'	3.50	5.50
T2	H3'	T2	H6	2.50	4.00
T2	H3'	A3	H8	3.50	5.50
T2	H4'	T2	H1'	3.50	5.50
T2	H4'	T2	H3'	2.50	4.00
T2	H5"	T2	H3'	2.50	4.00
T2	H5"	T2	H5'	1.80	3.00
T2	H5"	T2	H6	3.50	5.50
T2	H5'	T2	H1'	3.50	5.50
T2	H5'	T2	H4'	1.80	3.00

T2	H5'	T2	H6	3.50	5.50
T2	Me	T1	H1'	3.50	5.50
T2	Me	T1	H4'	3.50	5.50
T2	Me	T2	H6	2.50	4.00
T2	Me	A3	H4'	3.50	5.50
A3	H1'	A3	H8	2.50	4.00
A3	H1'	G4	H8	3.50	5.50
A3	H2'	A3	H2"	1.80	3.00
A3	H2"	A3	H1'	1.80	3.00
A3	H2"	A3	H3'	1.80	3.00
A3	H2"	A3	H4'	2.50	4.00
A3	H2"	A3	H5'	3.50	5.50
A3	H2"	A3	H8	1.80	3.00
A3	H2"	G4	H8	2.50	4.00
A3	H2'	A3	H1'	1.80	3.00
A3	H2'	A3	H4'	3.50	5.50
A3	H2'	A3	H5'	3.50	5.50
A3	H2'	A3	H8	1.80	3.00
A3	H3'	A3	H8	2.50	4.00
A3	H3'	G4	H8	3.50	5.50
A3	H4'	A3	H1'	1.80	3.00
A3	H4'	A3	H3'	1.80	3.00
A3	H5'	A3	H1'	3.50	5.50
A3	H5'	A3	H3'	2.50	4.00
A3	H5'	A3	H4'	2.50	4.00
A3	H5'	A3	H8	3.50	5.50
G4	H1'	G4	H8	2.50	4.00
G4	H1'	G5	H8	2.50	4.00
G4	H2"	G4	H1'	1.80	3.00
G4	H2"	G4	H3'	1.80	3.00
G4	H2"	G4	H4'	2.50	4.00
G4	H2"	G4	H5'	3.50	5.50
G4	H2"	G4	H8	1.80	3.00
G4	H2"	G5	H8	2.50	4.00
G4	H2'	G4	H1'	1.80	3.00
G4	H2'	G4	H2"	1.80	3.00
G4	H2'	G4	H3'	1.80	3.00
G4	H2'	G4	H4'	3.50	5.50
G4	H2'	G4	H5'	3.50	5.50
G4	H2'	G4	H8	1.80	3.00
G4	H3'	G4	H1'	2.50	4.00
G4	H3'	G4	H8	2.50	4.00
G4	H3'	G5	H8	3.50	5.50
G4	H4'	G4	H1'	2.50	4.00
G4	H4'	G4	H3'	1.80	3.00
G4	H4'	G4	H8	3.50	5.50
G4	H5'	G4	H1'	3.50	5.50
G4	H5'	G4	H3'	2.50	4.00
G4	H5'	G4	H4'	2.50	4.00

G4	H5'	G4	H8	3.50	5.50
G5	H1'	G5	H8	2.50	4.00
G5	H1'	G6	H8	2.50	4.00
G5	H2"	G5	H1'	1.80	3.00
G5	H2"	G5	H3'	1.80	3.00
G5	H2"	G5	H4'	3.50	5.50
G5	H2"	G5	H5'	3.50	5.50
G5	H2"	G5	H8	2.50	4.00
G5	H2'	G5	H1'	2.50	4.00
G5	H2'	G5	H2"	1.80	3.00
G5	H2'	G5	H3'	1.80	3.00
G5	H2'	G5	H5'	2.50	4.00
G5	H2'	G5	H8	2.50	4.00
G5	H2'	G6	H8	1.80	3.00
G5	H3'	G5	H1'	2.50	4.00
G5	H3'	G5	H8	2.50	4.00
G5	H3'	G6	H8	3.50	5.50
G5	H4'	G5	H1'	2.50	4.00
G5	H4'	G5	H3'	1.80	3.00
G5	H4'	G5	H8	3.50	5.50
G5	H5'	G5	H1'	3.50	5.50
G5	H5'	G5	H3'	2.50	4.00
G5	H5'	G5	H4'	2.50	4.00
G5	H5'	G5	H8	2.50	4.00
G6	H1'	G6	H8	2.50	4.00
G6	H2"	G6	H1'	1.80	3.00
G6	H2"	G6	H3'	1.80	3.00
G6	H2"	G6	H8	2.50	4.00
G6	H2'	G6	H1'	2.50	4.00
G6	H2'	G6	H3'	1.80	3.00
G6	H2'	G6	H2"	1.80	3.00
G6	H2'	G6	H8	1.80	3.00
G6	H3'	G6	H1'	2.50	4.00
G6	H3'	G6	H8	2.50	4.00
G6	H4'	G6	H8	3.50	5.50
G6	H5'	G6	H1'	3.50	5.50
G6	H5'	G6	H3'	2.50	4.00
G6	H5'	G6	H4'	2.50	4.00
G6	H5'	G6	H8	3.50	5.50
T7	H1'	T7	H6	2.50	4.00
T7	H2"	T7	H5'	3.50	5.50
T7	H2"	T7	H1'	1.80	3.00
T7	H2"	T7	H3'	1.80	3.00
T7	H2"	T7	H6	2.50	4.00
T7	H2'	T7	H1'	2.50	4.00
T7	H2'	T7	H3'	1.80	3.00
T7	H2'	T7	H5'	3.50	5.50
T7	H2'	T7	H6	1.80	3.00
T7	H3'	T7	H6	2.50	4.00

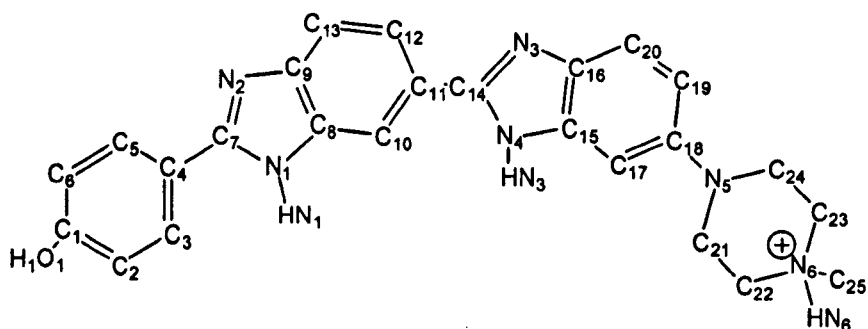
T7	H5'	T7	H1'	3.50	5.50
T7	Me	G6	H1'	3.00	5.50
T7	Me	G6	H3'	3.50	5.50
T7	Me	T7	H6	2.50	4.00
T1	H2'	T1	H2"	1.80	3.00
T7	H2'	T7	H3'	2.50	4.00
T7	H2'	T7	H4'	3.50	5.50
T7	H4'	T7	H1'	2.50	4.00
T7	H4'	T7	H3'	2.50	4.00
G6	NH	G27	H8	2.50	5.50
G13	NH	G5	H8	2.50	5.50
G6	NH	G5	NH	2.50	5.50
G4	NH	G25	H8	2.50	5.50
G4	NH	G5	NH	2.50	5.50
G5	NH	G26	H8	2.50	5.50
G6	H22	G6	NH	2.50	5.50
G5	H22	G5	NH	2.50	5.50
G6	H22	G27	H8	2.50	5.50
G5	H22	G26	H8	2.50	5.50

residue	atom	residue	atom	distance	distance
ACR	Me6	G4	H8	3.50	5.50
ACR	Me6	G6	H3'	3.50	5.50
ACR	Me8	A3	H1'	3.50	5.50
ACR	Me8	G4	H1'	3.50	5.50
ACR	Me8	G4	H8	3.50	5.50
ACR	Me8	G6	H1'	3.50	5.50
ACR	Me8	G6	H8	3.50	5.50
ACR	Me13	G4	H8	3.50	5.50
ACR	Me13	G6	H8	3.50	5.50
ACR	H5	G4	H4'	3.50	5.50
ACR	H5	G6	H2"	3.50	5.50
ACR	H7	A3	H2'	3.50	5.50
ACR	H7	G4	H5'	3.50	5.50
ACR	H9	G4	H4'	3.50	5.50
ACR	H9	G4	H5'	3.50	5.50
ACR	H7	G6	H2"	3.50	5.50
ACR	Me8	G6	NH	3.50	5.50
ACR	Me8	G4	NH	3.50	5.50
ACR	Me8	A3	H8	3.50	5.50
ACR	Me13	G4	NH	3.50	5.50
ACR	Me13	G6	NH	3.50	5.50
ACR	H10/H12	G6	NH	3.50	5.50
ACR	H10/H12	G4	NH	3.50	5.50
ACR	H10/H12	A3	H2"	3.50	5.50

Appendix 3

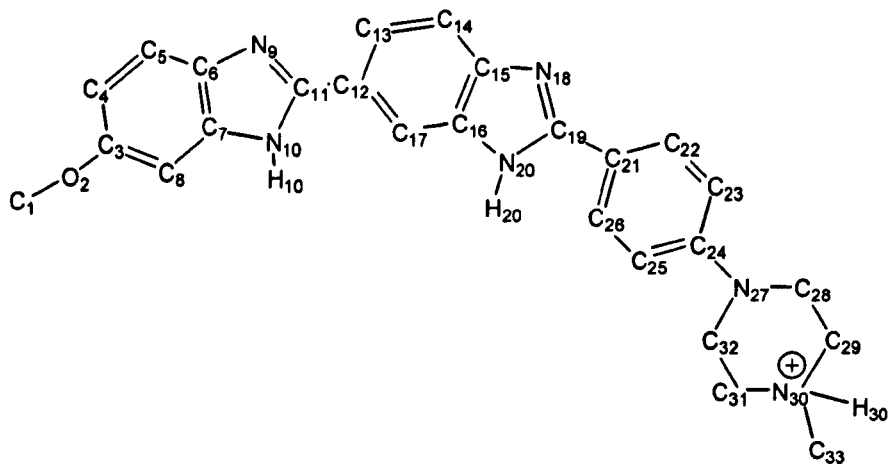
Appendix 3.1 Atom charges for H33258 molecule.

atom	charge	atom	charge	atom	charge
C1	0.3258	C10	-0.4226	H13	0.1558
C2	-0.2492	C14	0.4803	H12	0.1391
C3	-0.2016	N3	-0.3821	H10	0.1787
C4	-0.0398	C15	0.0754	HN3	0.3318
C5	-0.0987	C16	0.3346	H17	0.1592
C6	-0.1812	N4	-0.5909	H19	0.0962
O1	-0.5863	C17	-0.2607	H20	0.1966
H2	0.1553	C18	0.0405	H21	0.1072
H3	0.1478	C19	-0.1378	H21	0.1072
H5	0.1273	C20	-0.2703	H22	0.1746
H6	0.1728	N5	-0.1783	H22	0.1746
H1	0.4434	C21	-0.0726	H23	0.1597
N1	-0.5031	C22	-0.0976	H23	0.1597
C8	0.3216	N6	-0.2781	H24	0.1199
C9	0.1746	C23	-0.1018	H24	0.1199
N2	-0.5354	C24	-0.0867	H25	0.1624
C13	-0.1310	C7	0.4658	H25	0.1624
C12	-0.1830	C25	-0.2543	H25	0.1624
C11	-0.0207	HN1	0.3604	HN6	0.3997



Appendix 3.2 Atom charges for H10 molecule.

atom	charge	atom	charge	atom	charge
C11	0.4423	N10	-0.3481	H26	0.1765
C12	-0.0296	C5	-0.2297	H21	0.3316
C33	-0.0901	C4	-0.2997	H5	0.1905
C14	-0.305	C3	0.3136	H4	0.1764
C15	0.3917	C8	-0.3500	H8	0.1738
C16	0.0792	O2	-0.3417	H1	0.0660
C17	-0.3485	C1	0.1249	H1	0.0660
N20	-0.3742	C22	-0.0617	H1	0.0660
C19	0.4255	C23	-0.0273	H28	0.1174
N18	-0.5984	N30	-0.1506	H28	0.1174
C21	0.008	C31	-0.0039	H29	0.1155
C22	-0.162	C32	-0.0746	H29	0.1155
C23	-0.1179	C33	-0.3012	H30	0.3598
C24	0.2275	H20	0.3490	H31	0.1155
C25	-0.116	H13	0.1380	H31	0.1155
C26	-0.2604	H14	0.1931	H32	0.1174
N27	-0.4317	H17	0.1843	H32	0.1174
N9	-0.603	H22	0.1462	H33	0.1667
C6	0.3394	H23	0.1443	H33	0.1667
C7	0.0479	H25	0.1443	H33	0.1667



Appendix 3.3 Atom charges for RHPS4 molecule.

atom	charge	atom	charge	atom	charge
C1	0.1404	C16	0.0274	H15	0.1760
C2	0.0117	C17	-0.2376	H12	0.1755
C3	-0.1987	C18	0.2567	H17	0.1949
C4	0.2421	C19	-0.1117	H19	0.1849
C5	-0.0831	C20	-0.2763	H20	0.1984
C6	-0.2689	C21	0.0892	H24	0.1247
F7	0.1877	N23	-0.0456	H24	0.1130
C8	0.1184	C26	-0.2426	H24	0.1050
C9	0.1632	F22	-0.1872	H26	0.0998
C10	-0.0063	C25	-0.1417	H26	0.1025
N11	-0.0482	C24	-0.1629	H26	0.1063
C13	-0.0011	H3	0.1733	H25	0.1150
C13	-0.2426	H5	0.1924	H25	0.1015
C14	0.1750	H6	0.1916	H25	0.1196
C15	-0.2565				

