STRUCTURE, DYNAMICS AND HYDRATION IN DRUG-DNA RECOGNITION

By Huw E. L. Williams, BSc (Hons)

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy, September 2000

1. Contents

Abstract

The role of deoxyribonucleic acids in the cell has made DNA an attractive target for drug molecules. The anthracycline antitumour antibiotics are potent cytotoxic agents that have found widespread use in cancer chemotherapy. Nogalamycin binds DNA through intercalation, preferentially to 5'-TpG and 5'-CpG sites, by threading through the DNA helix and interacting with both the major and minor grooves simultaneously.

In this thesis, the interaction of nogalamycin with the 5'-TpG site has been investigated using synthetic oligonucleotide duplexes and a combination of high-resolution NMR techniques and NOE-restrained molecular dynamics simulations. The solution structure of the 1:1 complex with $d(ATGCAT)_2$ is described with NOE data unambiguously identifying the position and orientation of the bound drug molecule, allowing conclusions to be drawn regarding the specificity for the TpG site. Binding at one TpG site sterically blocks the interaction at the symmetrically equivalent CpA site. The structural studies are extended to investigate by NMR the role of solvation in drug-DNA recognition and binding. Based on the sign and magnitude of solute-solvent NOEs, it is shown that only a small subset of water molecules visible in the crystal and MD structures are found to be bound in the solution complex, and that a number of these are involved in mediating drug-DNA interactions. The role of the dynamic network of water molecules in stabilising the complex in solution is discussed.

Finally, the binding of nogalamycin at a TpG site carrying a DNA strand break has been investigated using a novel designed single-stranded intermolecular duplex consisting of two hairpins stabilised by GAA loops [d(AC<u>GAA</u>GTGC<u>GAA</u>GC)]. Although stacking of the two hairpins is weak, nogalamycin is shown to bind and stabilise a 1:1 complex by binding at the intercalation site. The complex is discussed in terms of the mechanism by which nogalamycin is able to bind to premelted duplex DNA.

Acknowledgements

I would like to thank a number of people who have helped, encouraged and distracted me over the course of my time in Nottingham.

Firstly, I would like to thank Dr Mark Searle for the studentship, supervision and his guidance throughout my PhD.

Thanks to everyone who I have worked with in B52 and B9a, and especially the Searle group, past and present, for making my time in Nottingham enjoyable; Al, Clare, Gary, Lee, Muriel, Sam, Evripidis, Geoff, and Michelle. I would also like to thank Claire and my parents for support and advice over the years.

Finally, I would like to acknowledge the EPSRC for funding, John Keyte for oligonucleotide synthesis, the technical staff within the department, Dr Charles Laughton for helpful discussions, and all at the university of Nottingham and Four Counties gliding clubs for providing some entertaining times.

Abbreviations and Symbols

σ^{NOE}	-	Relaxation rate for the nuclear Overhauser effect in the laboratory reference frame
σ^{ROE}	-	Relaxation rate for the nuclear Overhauser effect in the rotating reference frame
1D	-	1 dimensional
2D	-	2 dimensional
A	-	Adenine
C	-	Cytosine
COSY	-	Correlated spectroscopy
CPK	-	Corey-Pauling-Koltun model convention
dA	-	Deoxyadenosine
dC	-	Deoxycytidine
dG	-	Deoxyguanosine
DNA	-	Deoxyribonucleic acid
DQF	-	Double quantum filtered
dT	-	Deoxythymidine
EDTA	-	Ethylenediaminetetraacetic acid

FID	-	Free induction decay
FT NMR	-	Fourier transformed nuclear magnetic resonance spectroscopy
G	-	Guanine
HMBC	-	Heteronuclear multibond correlation spectroscopy
HMQC	-	Heteronuclear Multiquantum Coherence spectroscopy
Hoechst 33258	-	2-(4-hydroxyphenyl)-5-[5-(4-methylpiperarazin-1-yl)-benzimidazole-2-yl] benzimidazole
ISPA	-	Isolated spin-pair approximation
m-AMSA	-	Amsacrine
MD	-	Molecular dynamics
NMR	-	Nuclear magnetic resonance
NOE	-	Nuclear Overhauser effect
NOESY	-	Nuclear Overhauser effect spectroscopy
Nogalamycin	-	2,6-Epoxy-2H-naphthaceno[1,2-β] oxocin-14-carboxylic acid, 11-[(6-deoxy-3-C-methyl-2,3,4-tri-O-methyl-α-L-mannopyranosyl)oxy]-4-(dimethylamino)-3,4,5,6,9,11,12,13,14,16-decahydro-3,5,8,10,13-pentahydroxy-6,13-dimethyl-9,16-dioxo-, methyl ester

PME	-	Particle mesh Ewald
ppm	-	Parts per-million
rf	-	Radio frequency
rMD	-	Restrained molecular dynamics
RMS	-	Root mean square
RNA	-	Ribonucleic acid
ROE	-	Rotating Overhauser effect
ROESY	-	Rotating Overhauser effect spectroscopy
T	-	Thymine
t1	-	First time domain
T1	-	Longitudinal relaxation (spin-lattice)
t2	-	Second time domain
T2	-	Transverse relaxation (spin-spin)
TOCSY	-	Total correlation spectroscopy
TOP1	-	Topoisomerase 1
TOP2	-	Topoisomerase 2
TSP	-	Trimethylisilylpropanoate
UV	-	Ultraviolet
vdW	-	van der Waals

WATERGATE - Water gradient attenuation by tailored excitation

Chapter 1. DNA Targeting Drugs and Biological Function

1.1. Introduction

Nucleic acids were discovered from cell extracts in 1868 by Friedrich Miescher. Although it was recognised that deoxyribonucleic acid (DNA) formed a large part of the cell's nuclear material, its role as the genetic information carrier was not recognised. This stemmed from the misconception that DNA consisted of repeating units of the four nucleotide bases, adenine (A), thymine (T), guanine (G) and cytosine (C) in equal proportions. As purification and analysis procedures improved this premise was disproved, although equimolar proportions of A and T and G and C are found the proportion of AT does not have to equal GC. This discovery suggests that there is variability in the sequence, a prerequisite for an information-encoding molecule. The link between DNA and the genetic information gradually became more compelling. In 1953, the structure was elucidated by a combination of fibre diffraction and modelling (Watson and Crick, 1953). The now famous structure of the double helix was determined and this structure immediately gave rise to mechanisms as to how DNA could replicate, with one strand acting as a template for the other Figure 1-1. Since this time, the knowledge of cellular processes and understanding has been revolutionised.

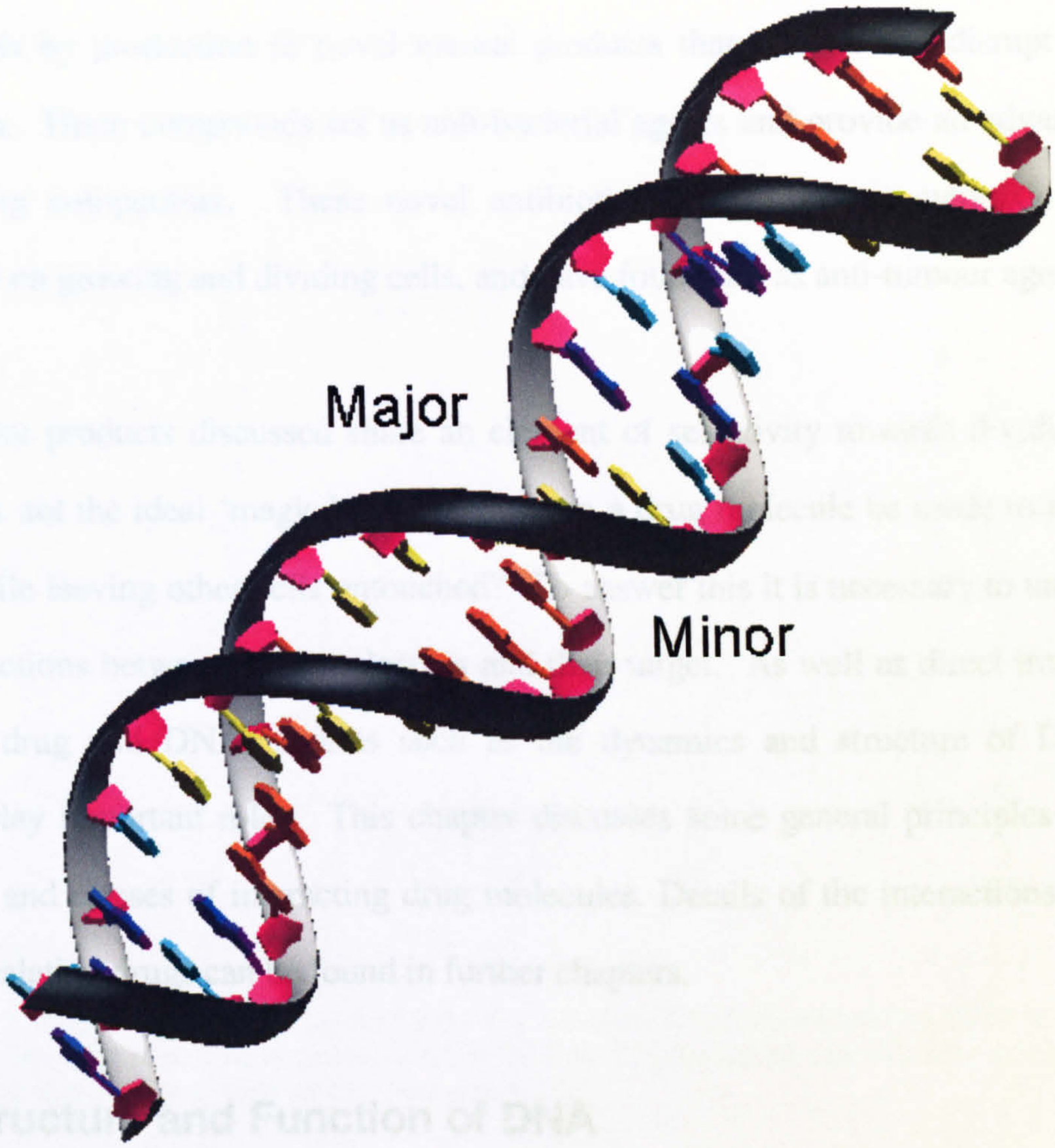


Figure 1-1: The double helix. Two strands of DNA associate together in a complementary fashion forming the major and minor grooves.

DNA molecules serve two essential roles; the controlled production of mRNA transcripts and, ultimately, all the cellular proteins, and the production of hi-fidelity copies to offspring cells. It is not surprising that such a crucial molecule is the target of many cellular proteins that are involved with not only replication and transcription, but also packaging repair and control. The disruption of DNA and DNA-protein

interactions by drug molecules has a profound effect on the cell, and many bacteria exploit this by production of novel natural products that interact and disrupt cellular operations. These compounds act as anti-bacterial agents and provide an advantage by suppressing competitors. These novel antibiotics possess potent cytotoxic effects, especially on growing and dividing cells, and have found use as anti-tumour agents.

The natural products discussed share an element of selectivity towards dividing cells, but this is not the ideal 'magic bullet'. How can a drug molecule be made to affect the target while leaving other cells untouched? To answer this it is necessary to understand the interactions between drug molecules and their target. As well as direct interactions between drug and DNA, aspects such as the dynamics and structure of DNA and solvent play important roles. This chapter discusses some general principles of DNA structure and classes of interacting drug molecules. Details of the interactions between the intercalating drugs can be found in further chapters.

1.2. Structure and Function of DNA

1.2.1 DNA Structure: The Role of Weak Interactions

The double helix suggested by Watson and Crick was conclusively proven by single crystal X-ray crystallography, from which finer structural details began to emerge. Any of the four deoxyribose bases can be connected together via a 5'-3' phosphodiester linkage on the deoxyribose sugar. The two complementary strands can associate together, with adenine (A) always pairing with thymine (T), and guanine (G) with cytosine (C). This is consistent with the data for the relative proportions of each base

pair found experimentally. The model for B-form DNA is close to the conformation adopted *in vivo*. In this conformation two grooves are present, referred to as the major and minor grooves. These grooves are crucial in forming interactions with proteins and small molecules.

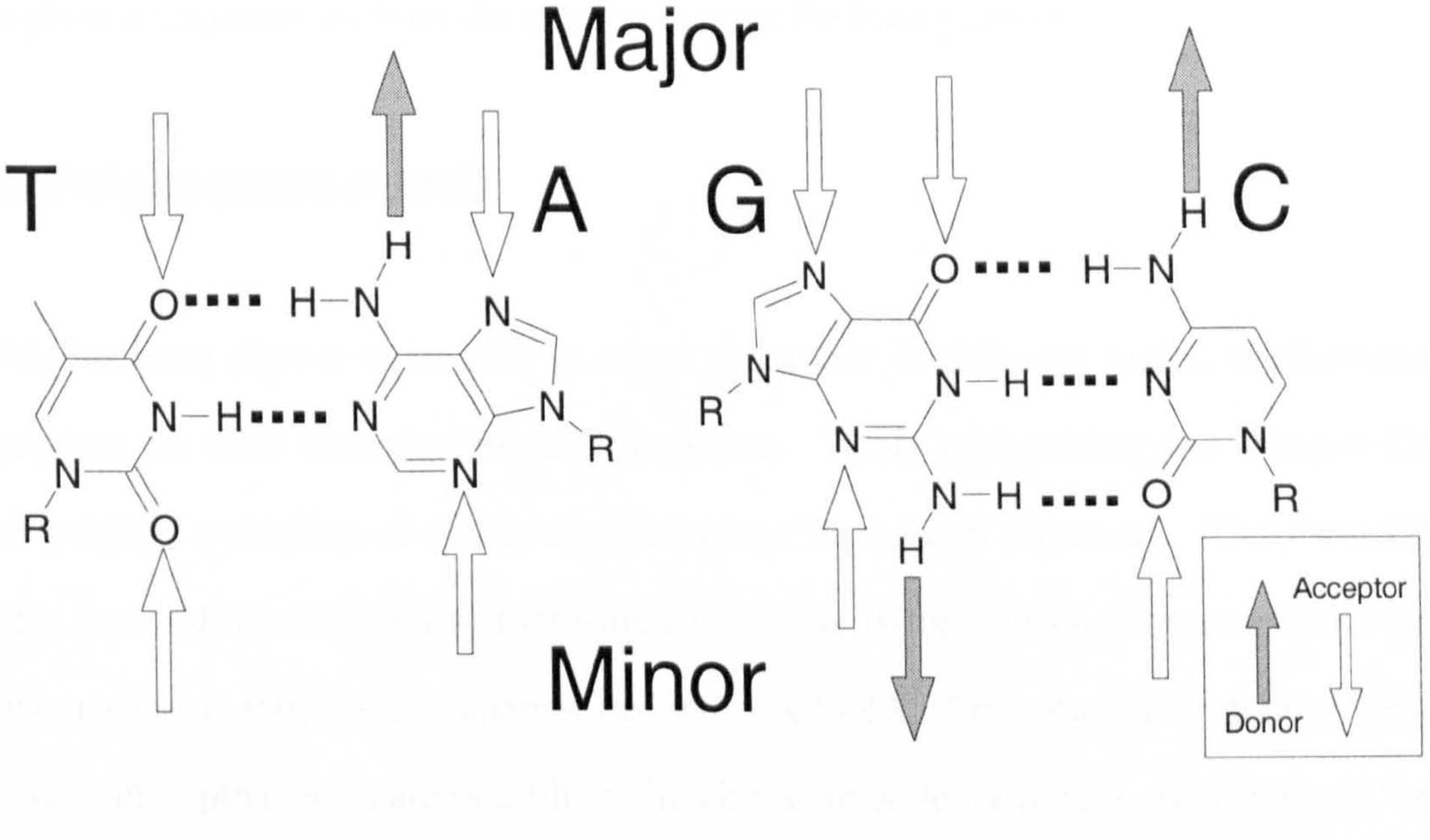


Figure 1-2 Hydrogen bonding dictates the Watson-Crick basepairing shown as dotted lines between basepairs. Additional hydrogen bond donors and acceptors are available in the floors of the two grooves.

The structure of DNA makes it well suited to the role of information carrier. The four bases can be assembled in any combination to build the genetic code, and polymers of many million bases are stable. This stability in the genetic code carrying molecule is important, however the functioning of DNA is governed by weak interactions. Non-covalent electrostatic interactions between basepairs, the 'π-stacking' interactions and, to a lesser extent, hydrophobic effects play important roles in stabilising the duplex (Searle and Williams, 1993). The H-bonding surfaces allow highly specific interactions to be formed. The base-pairing scheme suggested by Watson and Crick provides

specific recognition of adenine by thymine, and of guanine by cytosine based on the hydrogen-bonding pattern shown (Figure 1-2). This is of fundamental importance since this governs DNA's templating role. Other base pairing surfaces are also available in the minor and major groove. These allow proteins and small molecules to be able to recognise a sequence without the need to disrupt the base pairing.

1.2.2 Polymorphism of DNA

DNA has been shown to be able to adopt a number of different varied conformations dependant on both environment and sequence. While maintaining the Watson-Crick base pairing, a number of stable conformations have been observed. The form DNA adopts in conditions similar to that found *in vivo* is B-form DNA, however if the water concentration is reduced a transition occurs to A-DNA. This transition changes groove widths and depths, and causes a tilt in the helical axis, leading to a characteristic 'hole' forming down the middle of the duplex (Figure 1-3). Z-DNA is an even more unusual structure formed by poly-CG sequences and adopts a left-handed helix (Wang et al. 1979). The polymorphic behaviour of DNA appears to be crucial to the biological function. A-DNA like transitions are observed on the binding of a number of proteins, and a Z-DNA conformation is believed to be important in strand separation (Stryer, 1988).

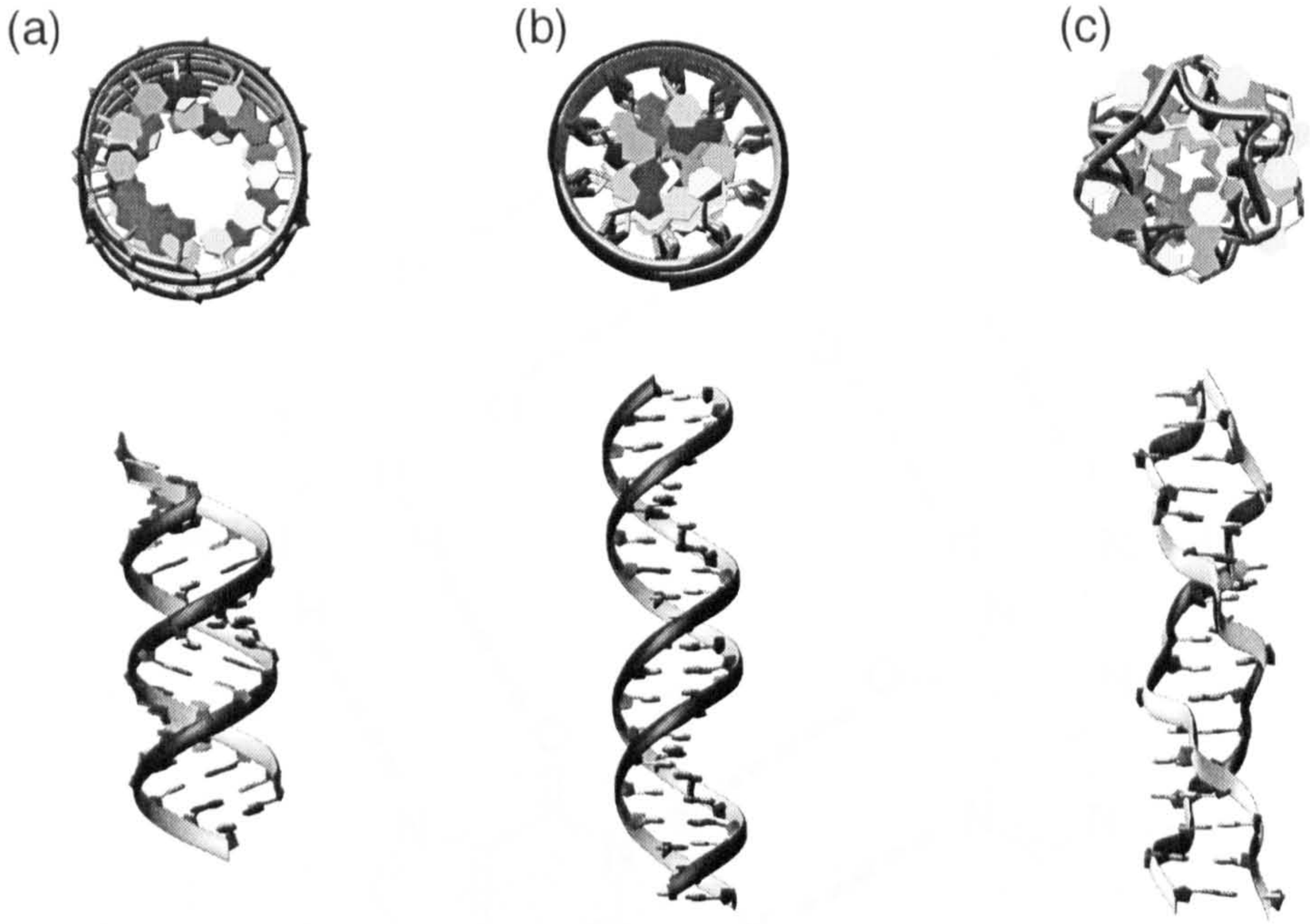


Figure 1-3: Three conformations of DNA. A-DNA (a) B-DNA (b) and Z-DNA (c) The top views are generated by projecting along the main helical axis.

DNA also adopts a number of conformations that do not follow the Watson-Crick base pairing. Telomeric repeats are sequences that occur at the ends of the chromosomes. These sequences form tetraplex structures, where four guanines associate forming a planar arrangement (Figure 1-4). Since telomeres are involved in cell replication, drugs that target these sequences are potentially useful anti-tumour agents (Haq et al. 1999).

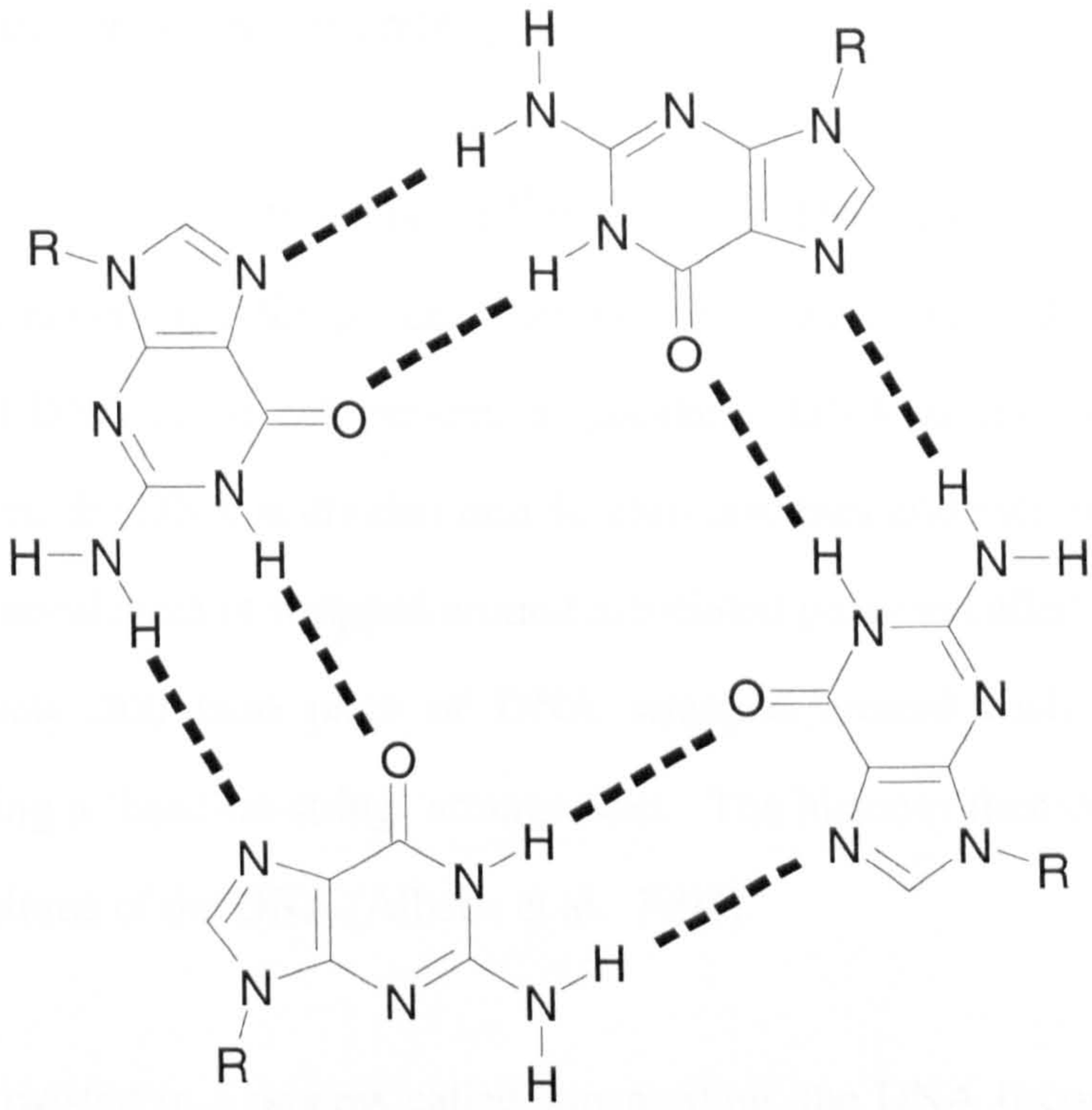


Figure 1-4: Tetraplex structure adopted at chromosomal telomeres. An alternative base-pairing strategy allows four guanines to associate to form this planar structure.

More recently, some single stranded sequences of DNA have been shown to adopt stable hairpins with as few as seven nucleotides being involved. These sequences adopt an unusual base pairing orientation in the loop region, with a G-A base pair (Hirao et al. 1994). Although it is not known if such sequences are important *in vivo*, the sequence shows resistance to 3' exonuclease, which has potential use as a protection sequence for anti-gene therapy. Such sequences also allow the creation of novel nicked sites by a combination of two loops, and some properties of these are investigated in Chapter 6.

1.2.3 DNA Topology and Supercoiling

The human cell is estimated to contain 10^9 basepairs of DNA and this would lead to a length of over 2 meters of DNA per cell assuming B-form structure. The storage of this large amount of DNA in the cell presents a quandary. DNA must exist in a packaged form. In humans, the DNA is divided into 46 chromosomes and each one consists of a single DNA molecule that is wrapped around associated proteins called histones. There are approximately 200 base pairs of DNA wrapped around each histone protein complex, creating a 'bead-on-string' arrangement. The histones then condense, further reducing the volume of the DNA (Alberts et al. 1994).

If the DNA is twisted in a process called supercoiling, the DNA becomes tangled and decreases in length. The phenomenon can be similarly observed by a demonstration with a telephone. If the handset is twisted a number of times, when the handset is replaced the cord will wrap around itself. An example of bacterial DNA plasmid can be seen in Figure 1-5. Breaking of the plasmid, twisting one end relative to the other, and religating introduces supercoiling. This can be reversed by nicking a strand, thus causing the DNA to relax back to the original loop. The supercoiling of DNA can also aid in strand separation.

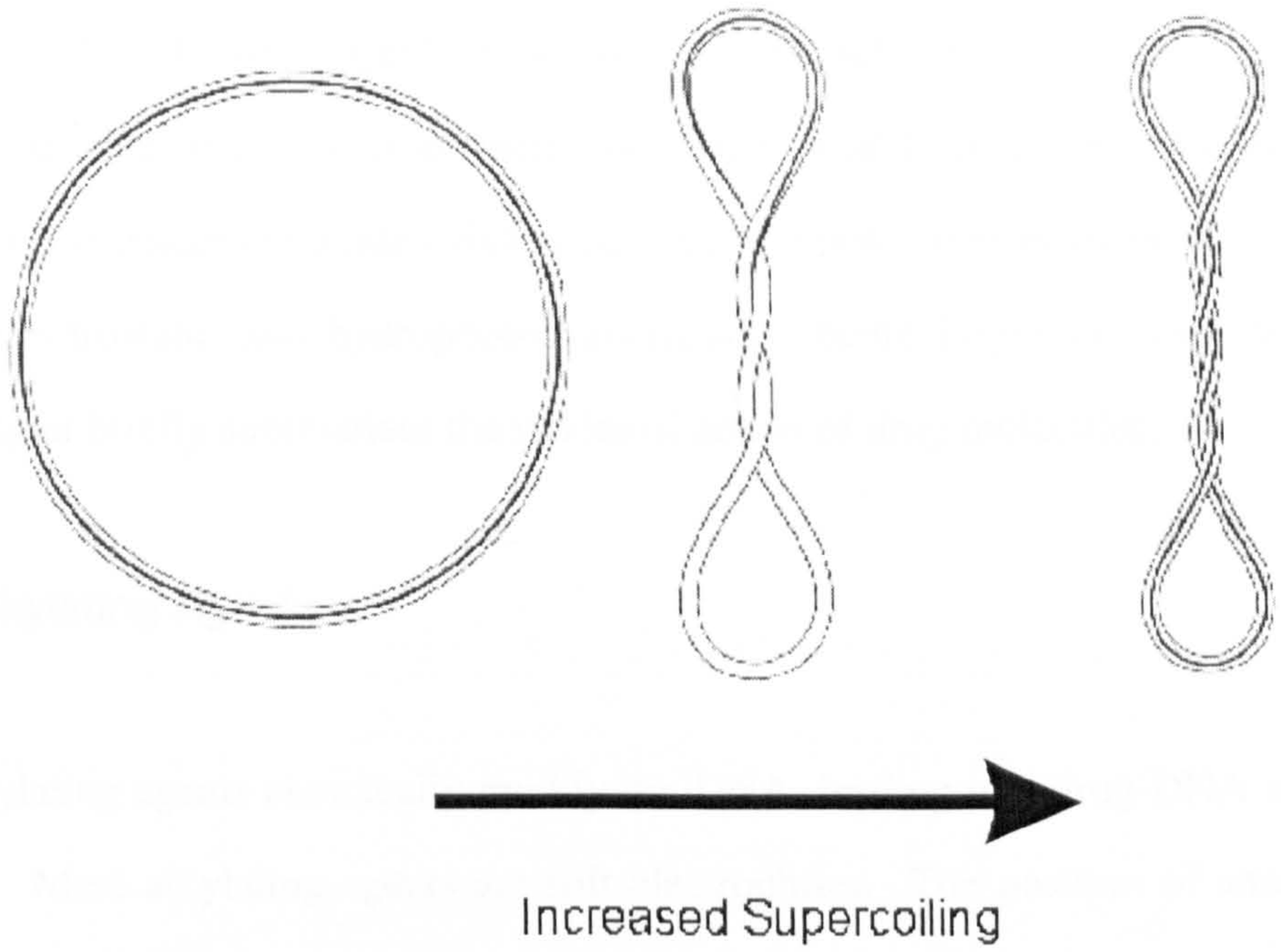


Figure 1-5: The effect of supercoiling on plasmid DNA.

Enzymes accessing DNA can change the supercoiled state, which can lead to tangling. The process of removing supercoiling is tightly controlled by enzymes called topoisomerases. As will be discussed in the following sections, drugs that interact with this class of enzyme and DNA have important clinical uses.

1.3. Drugs that Target DNA

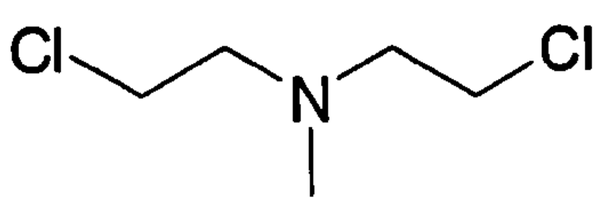
Drug molecules that interact with DNA can be divided into three categories. Examples from each of these categories are used clinically as cytotoxic agents, but there is large variety in the specificity and complexity of these drug molecules. There is also a wide variability in the amount of distortion that these drug molecules induce in the DNA

structure. The alkylating agents form covalent interactions with the DNA target, and even at multiple locations in an inter- and intra- strand sense. In contrast, both the minor groove binders and intercalators interact in a non-covalent manner, with van der Waals, electrostatic and hydrophobic interactions being important stabilising forces. This chapter briefly summarises the modes of action of drug molecules.

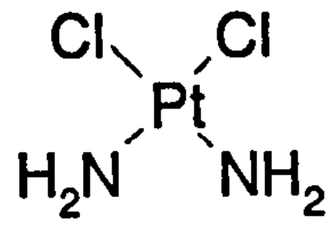
1.3.1 Alkylating Agents

The alkylating agents chemically modify the DNA, leading to a drug-DNA aduct being formed. Most alkylating agents are soft electrophiles. The position of attack for soft electrophiles on DNA base pairs progresses in the order G-N7 > A-N1 > C-N3 > T-N3. The specificity of these drug molecules is dependent on the nucleophilicity of the site for attack. The nitrogen mustard type compounds have been shown to cross link guanine N7, and similar patterns are seen with *cis*-diamminedichloroplatinum (II) (cisplatin). The cross-linking of guanine by cisplatin forces the nucleotides to adopt a perpendicular arrangement to fit the tetrahedral geometry of platinum, therefore disrupting the base pairing (Schwartz et al. 1989).

Nitrogen Mustard



cisplatin



CC-1065

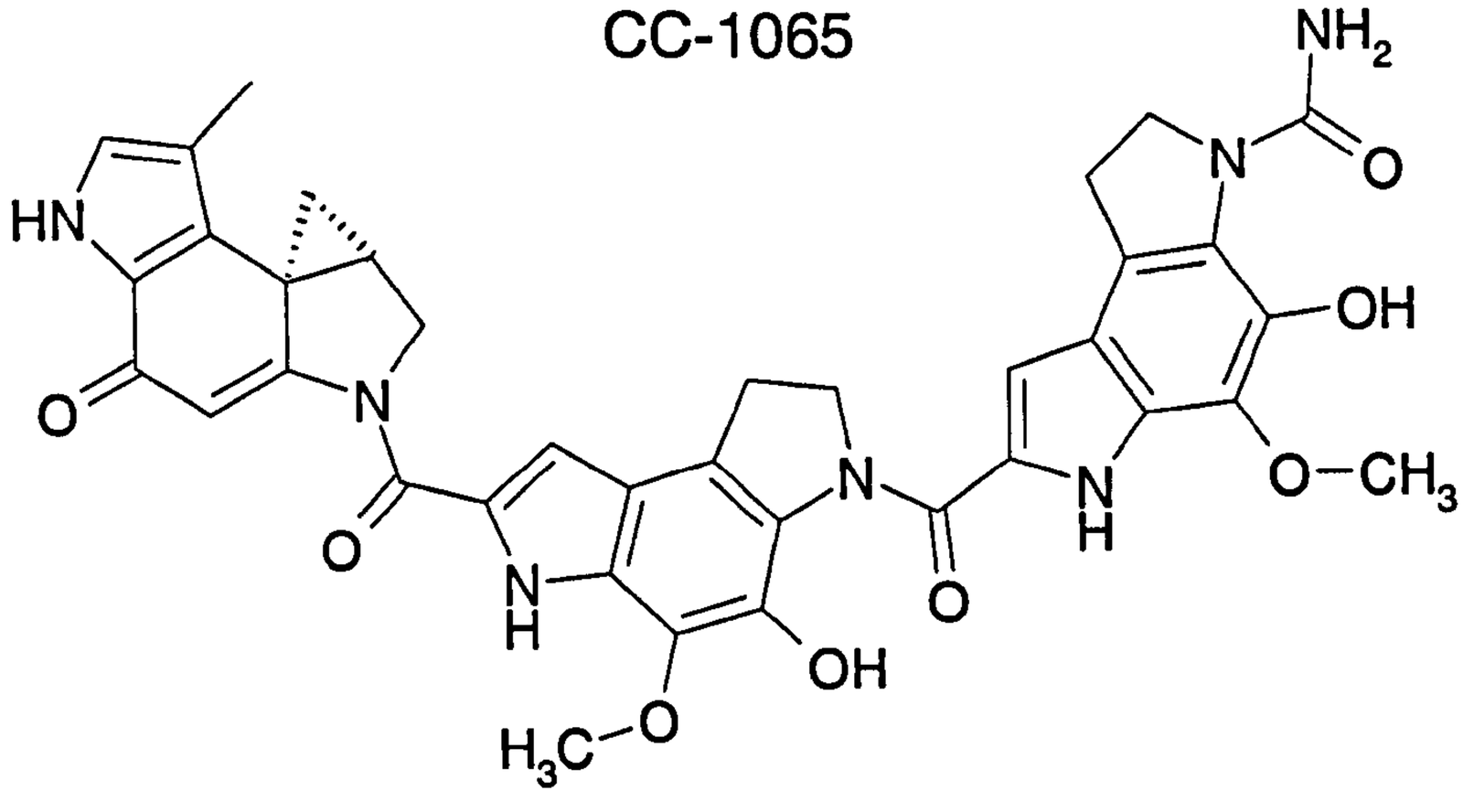


Figure 1-6: DNA alkalating agents vary in complexity. Most interact by nucleophilic attack on nucleotide bases.

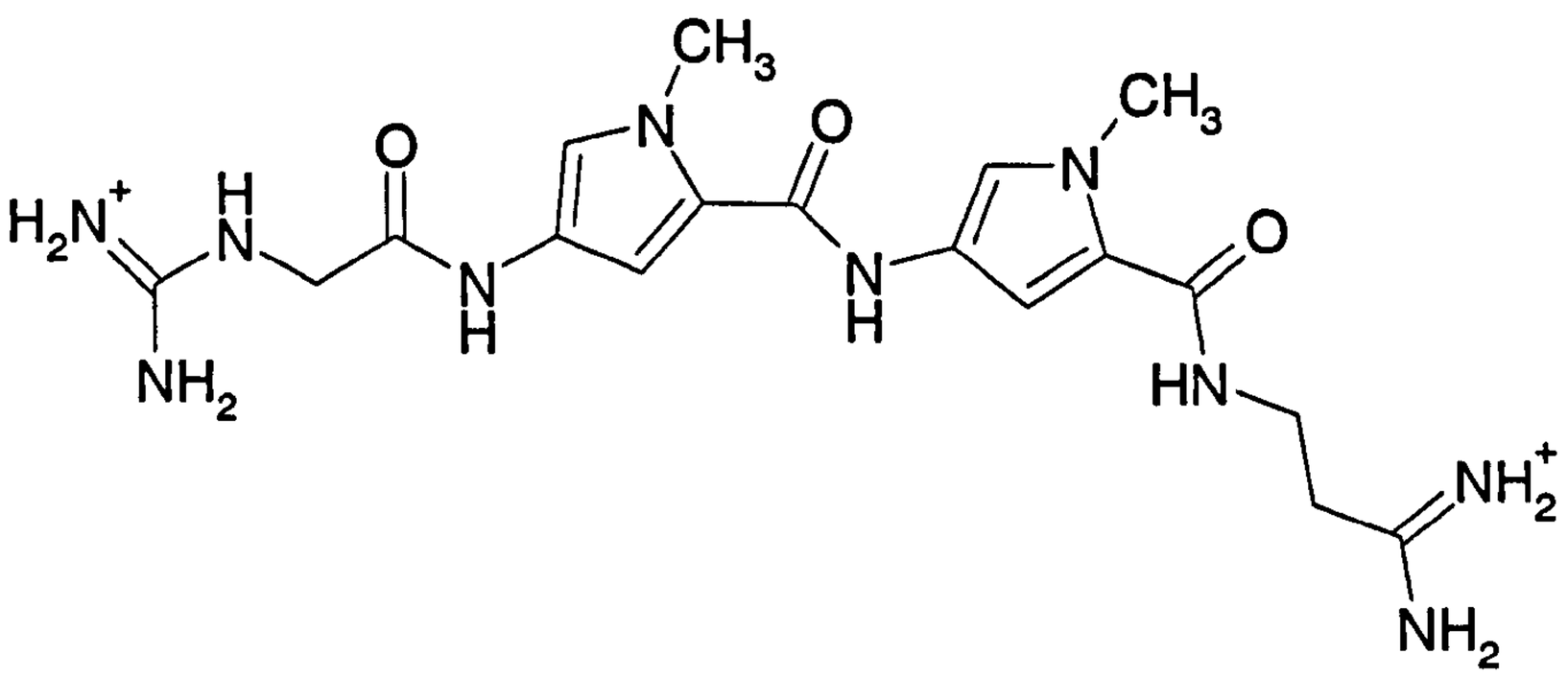
Not all alkylating agents are this simple, CC-1065 is isolated from a species of streptomyces and is a minor groove binder. The drug is selective for AT sequences where nucleophilic attack by an adenine N3 causes the cyclopropane ring to open. This leads to alkylation and destabilisation of the duplex.

1.3.2 Minor Groove Binders

The two grooves of DNA can act as binding pockets for interacting molecules. While the major groove is predominantly targeted by proteins, small ligands usually target the minor groove. The narrowness and depth of the minor groove provides close van der

Waals contacts and a hydrophobic pocket for small molecules that bind there. Minor groove binders share a number of common features including isohelicity, charge and hydrogen-bonding affinity. The isohelicity of these drugs with the DNA minor groove can be seen in Figure 1-7, with both molecules having the typical curvature of minor-groove binders. Minor groove binders usually possess some flexibility that allows the shape to better fit that of the minor groove. The positive charge is thought to be most important during the initial binding stages, but is able to form interactions with the phosphate backbone and floor of the groove.

Netropsin



Hoechst 33258

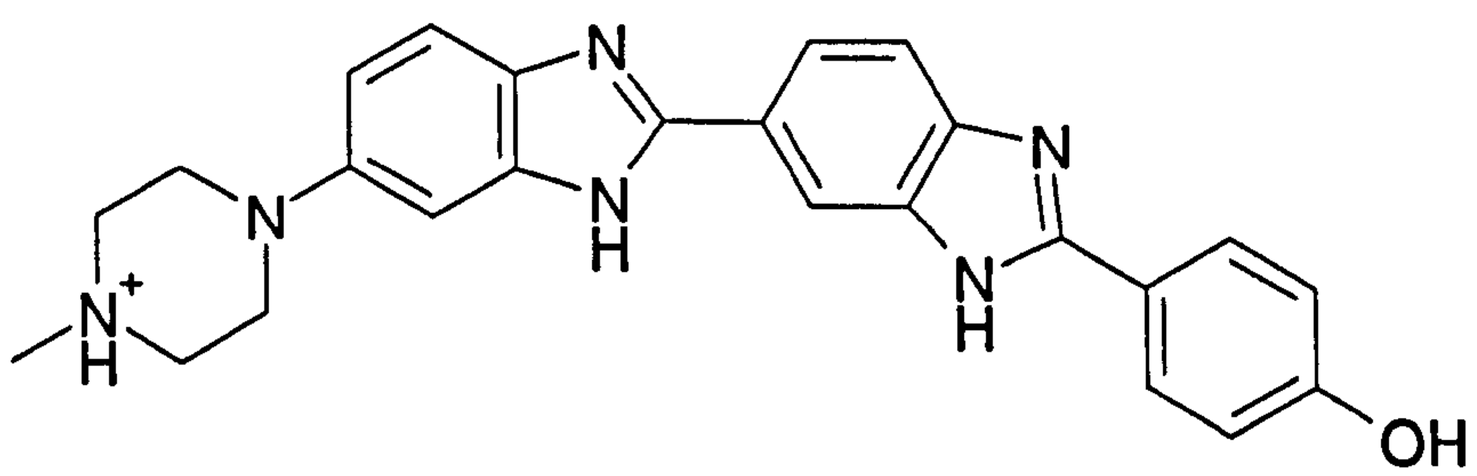


Figure 1-7 The minor groove binders netropsin and Hoechst 33258. Both drugs show a curved shape that is isohelical with DNA.

Many minor groove binders show preferences for runs of AT sequences, and there are a number of reasons for this. The AT base-pairs do not possess any functional groups that project into the minor groove; by comparison, the -NH_2 on guanine is projected into the minor groove. Further, these AT-tracts show a distinct narrowing in the minor groove, and this leads to an increase in van der Waals interactions for drug molecules bound at these sequences. Both distamycin and Hoechst 33258 have hydrogen bond donors on the concave face, and are able to interact with the hydrogen bond acceptors on the AT base-pairs minor groove edge.

Although there is a distinct bias towards AT sequences, switching the hydrogen bond donors on the molecule to hydrogen bond acceptors would bias the binding preference of these molecules. Work using overlapping drug molecules that consist of pyrrole and imidazole units, has been used to target a 16 base-pairs varied GC and AT sequence. The side-by-side arrangement of the drug molecules allows distinction of GC versus CG base pairs (Trauger et al. 1998). The actual 16 base-pair target used corresponded to a sequence found in an HIV promoter region, and the ability to selectively impede a gene's operation provides some exciting prospects for future therapeutic strategies.

1.3.3 Intercalators

The intercalators are the final binding model for DNA binding drugs. The classical intercalating model was proposed in the 1960's, following observations that some DNA associating drug molecules were observed to increase the viscosity of DNA (Blackburn and Gait, 1996), a property not associated with minor groove binders. For intercalation to occur the drug molecule must be able to insert between the DNA bases perpendicular to the long axis of DNA. This causes an increase in separation of base pairs by up to

3.4 Å to accommodate the drug. Intercalation also leads to a change in twist, which causes unwinding of the DNA. The amount of unwinding varies depending on the intercalator used but values as large as 26° are seen.

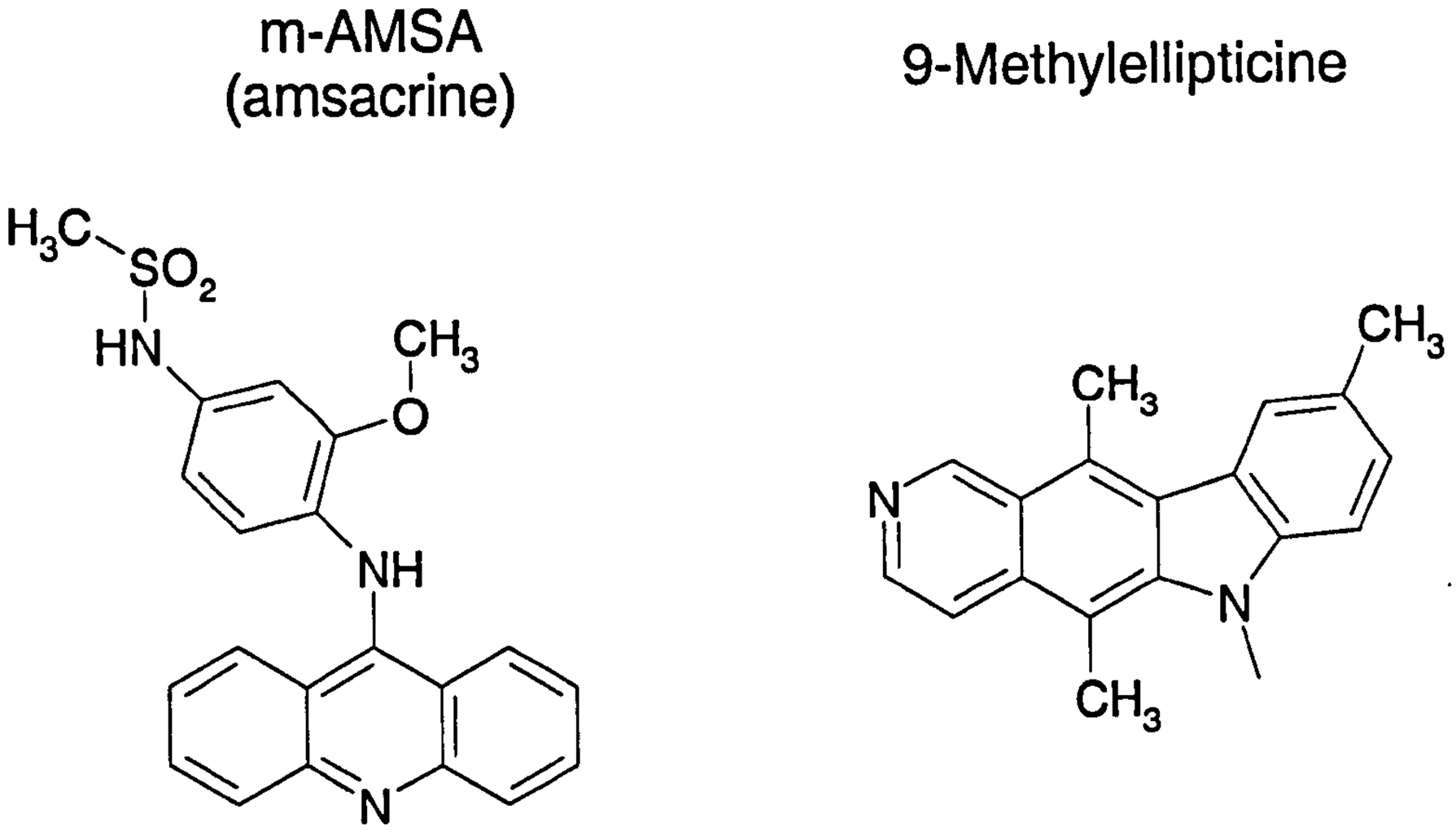


Figure 1-8: Examples of biologically active intercalators m-AMSA and 9-Methylellipticine. The planar aromatic groups on both of these molecules are able to insert between DNA base pairs. Both these intercalators are potent inhibitors of Topoisomerase II.

Molecules that interact with DNA in this manner perhaps unsurprisingly exhibit many biological effects, and have found uses in biochemistry as well as medicine. The molecules shown in Figure 1-8 consist of a planar aromatic moiety that is able to insert between DNA base pairs.

Although it could be envisaged that intercalation could occur between each base step, this is not usually seen even with simple intercalators. This 'neighbour exclusion principle' suggests that even minor changes in DNA structure can propagate to nearby sites, thus disfavouring binding at these sites.

The anthracyclines are a class of structurally related intercalators, and all possess a polyketide-derived aglycone with functionality (Kantola et al. 2000). Examples of molecules in this family are produced by a number of bacteria, and these intercalators are among the most widely used chemotherapeutic agents. The anthracyclines also have groove-binding functionalisation on either or both ends of the aglycone structure.

The difference in groove binding functionality affects the binding affinity and biological effects of the drugs. Nogalamycin has two groove binding sugars, and is the strongest binding and most potent. Nogalamycin is an excellent probe since it forms highly stable complexes with a plethora of interactions. Although nogalamycin is not used clinically, due to significant cardiotoxic effects, related compounds such as daunomycin are used. Synthetic analogues derived from nogalamycin such as 7(R)-O-methylnogarol (menogaril) are also finding clinical applications (Krueger et al. 1981).

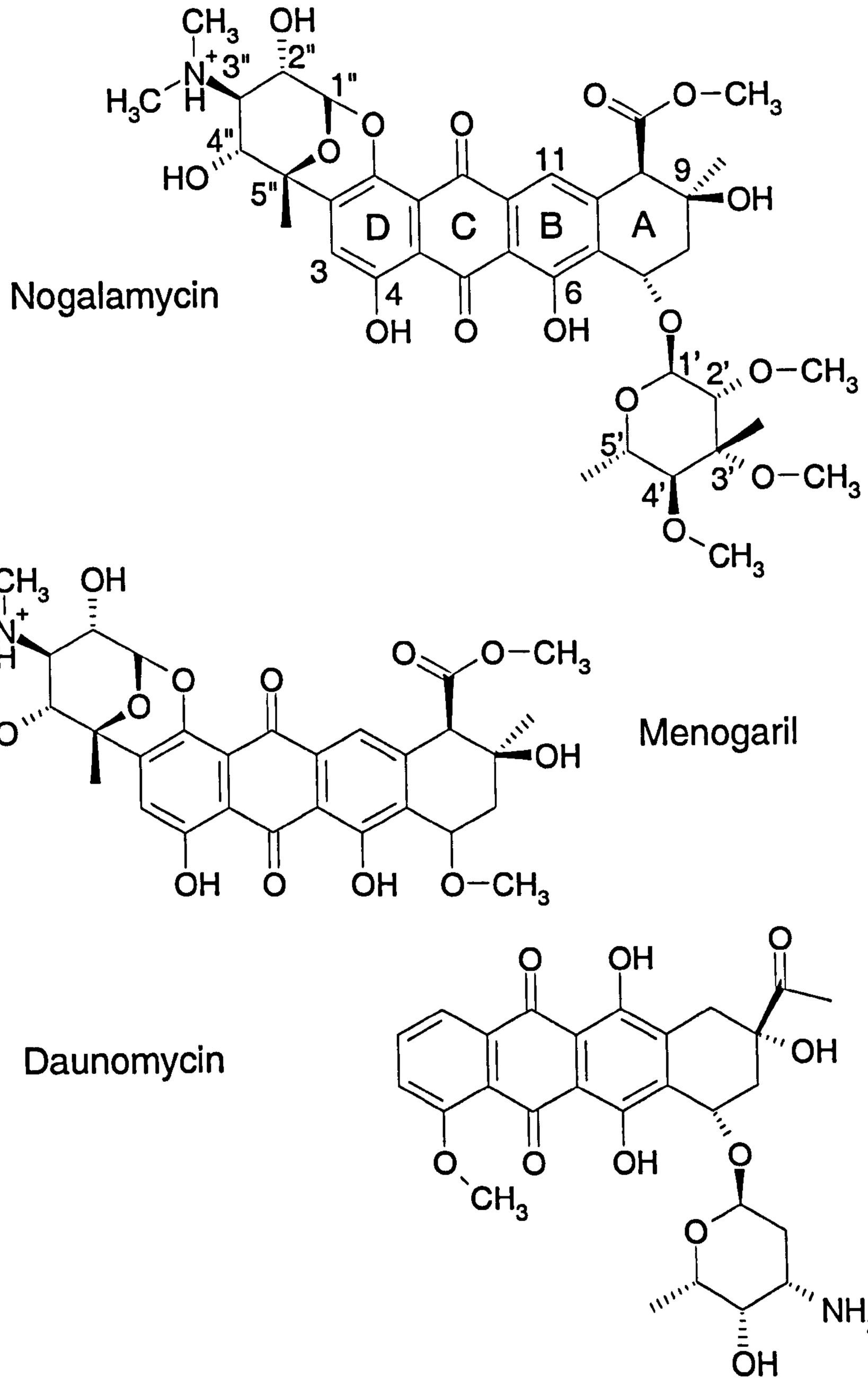


Figure 1-9: The anthracyclines. The four ring anthracine derived functionality is common to all members. Daunomycin has only a minor groove binding sugar. Nogalamycin possesses both major and minor groove binding. Menogaril is a semi-synthetic derivative of nogalamycin lacking the nogalose sugar.

1.4. Origin of Cytotoxicity for Anthracycline Antibiotics

The production of such complicated drug molecules represents a considerable investment, both in terms of energy and genetic information by the bacterium. These molecules must therefore instil a competitive advantage, and these compounds exhibit antibacterial as well as cytotoxic characteristics. What are the sequence of events that leads to cell death on exposure to the anthracyclines?

The drug molecules especially in the case of nogalamycin exhibit strong binding and significant distortion to DNA structure. These molecules possess extensive functionality, and have been shown to exhibit various biological effects. Three main routes of causing cell dysfunction have been suggested for the anthracycline antibiotics.

The simplest of these proposed mechanisms considers the role of the drug in template-blocking, similar to that suggested for some of the minor groove binders. Enzymes such as those involved in replication and transcription including RNA polymerase, associated transcription factors and DNA polymerase, require an association event in order to function. Blocking of the association or even the movement along the strands of DNA would seriously impair function of the cell. Evidence for this is seen in some gel assays that observe various effects that are consistent with enzymes being prevented from interacting with DNA (Welch et al. 1994). Although it is likely to be a factor, studies report fragmentation of DNA within cells which lead to cell death.

A well understood method of causing DNA damage is oxidative attack caused by the production of radicals close to the DNA. If the drugs promote the formation of active

radicals when intercalated to the DNA this would lead to damage of the cellular DNA. This is the route taken by calicheamicin, where reduction of a sulfur group leads to a Bergman cyclisation and formation of an active carbon radical 'warhead'. The remainder of this complex cytotoxic agent is important for forming interactions with the DNA minor groove, and maintaining the close proximity of the carbon radical and DNA (Ikemoto et al. 1995). The diquinone function of the anthracycline ring system has been shown to increase the formation of radicals in solution (Müller et al. 1997).

Studies to investigate this oxidative damage in doxorubicin (adriamycin), a related anthracycline to nogalamycin, showed differential effects depending on concentration. At concentrations between 1 and 5 μM DNA, strand cleavage was observed as well as inhibition of RNA synthesis and induction of programmed cell death, apoptosis, a system that causes cell termination when damage has occurred. Significantly, no increase in oxidation products of DNA bases were observed at these concentrations, which correspond to typical plasma values observed in patients during standard chemotheraputic regimens. Elevated concentrations of the 8-hydroxyguanine product were observed with concentrations of drug of about 100 μM. At physiological concentrations, however, oxidative damage does not provide the main kill mechanism of these drugs (Müller et al. 1997).

The principal kill mechanism comes from the interaction of the drugs with topoisomerases. The topoisomerases are an ubiquitous class of enzymes that are involved in altering the morphology of DNA by removing supercoiling. Since DNA exists in this 'packaged' form, these enzymes are essential in any function of DNA. The enzymes exist in two classes, type 1 and type 2 (TOP1 and TOP2 respectively). Both enzyme classes function by breaking the DNA backbone. TOP1 causes a single

nick in the backbone and rotation about this break; this is a nick-swivel-close mechanism. In contrast TOP2 breaks both strands and moves another region of duplex through the open 'gate', using a double-strand passage mechanism. During this phase the enzymes are covalently linked to the DNA (Berger et al. 1996). The final stage is the religation to reform the contiguous backbone and dissociation of the topoisomerase. In eukaryotic cells, TOP2 is associated with the nuclear matrix and associated with DNA synthesis, chromosome condensation and separation (Berrios et al. 1985). TOP1 has been associated with chromatin assembly, recombination, cell division and DNA template reading by RNA polymerase.

Most anthracyclines and analogues interact with the cleavage complex, the covalently attached form of topoisomerase to DNA. Stabilising the ternary complex between drug DNA and topoisomerase can result in dissociation of the enzyme before the religation stage of the catalytic cycle. The drug molecules cause cellular systems to bring about DNA damage, ultimately leading to apoptosis in dividing cells.

Nogalamycin has been shown to be active against TOP1 (Sim et al. 1997). This is consistent with the observation that nogalamycin leads to a marked decrease in RNA production (Ray et al. 1996), since TOP1 is involved with RNA polymerase function. In similarity with other topoisomerase I active drug molecules, such as camptothectin, nogalamycin is most toxic during the S-phase of the cell cycle when mRNA production is at a maximum.

The interactions formed with the ternary complex are yet unknown, however biochemical studies are revealing aspects of topoisomerase poisoning. Nogalamycin forms a good model system for the study of the differential effects, since the closely

related synthetic analogue menogaril completely lacks TOP1 inhibition, but is a potent TOP2 inhibitor. The only difference between these two molecules is that menogaril lacks a minor groove binding sugar. The role of the sugar is unclear, but other TOP1 inhibitors possess some form of minor groove binding moiety. Topoisomerase I has an increased affinity for curved DNA sequences. A possible mechanism is that the distortion in the DNA on binding these molecules enhance the association of topoisomerase I (Sim et al. 2000).

1.5. Bibliography

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. and Watson, J.D. (1994) *Molecular Biology of the Cell*, Garland Publishing Inc.

Berger, J.M., Gamblin, S.J., Harrison, S.C. and Wang, J.C. (1996) Structure and Mechanism of DNA Topoisomerase II. *Nature* **379**, 225-232.

Berrios, M., Osheroff, N. and Fisher, P.A. (1985) Insitu Localization of DNA Topoisomerase-II, A Major Polypeptide Component of the Drosophila Nuclear Matrix-Fraction. *Proc. Nat. Acad. Sci. USA* **82**, 4142-4146.

Blackburn, G.M. and Gait, M.J. (1996) *Nucleic Acids in Chemistry and Biology, 2nd Edition*,

Haq, I., Trent, J.O., Chowdhry, B.Z. and Jenkins, T.C. (1999) Intercalative G-tetraplex stabalization of telomeric DNA by a cationic porphyrin. *J. Am. Chem. Soc.* **121**, 1768-1779.

Hirao, I., Kawai, G., Yoshizawa, S., Nishimura, Y., Ishido, Y., Watanabe, K.A. and Miura, K.-I. (1994) Most Compact Hairpin-Turn Structure Exerted by a Short DNA Fragment, d(GCGAAGC) in Solution: an Extraordinarily Stable Structure resistant to Nucleases and Heat. *Nuc. Acids. Res.* **22**, 576-582.

Ikemoto, N., Kumar, R.A., Ling, T.T., Ellestad, G.A., Danishefsky, S.J. and Patel, D.J. (1995) Calicheamicin-DNA Complexes: Warhead Alignment and Saccharide Recognition of the Minor Groove. *Proc. Nat. Acad. Sci. USA* **92**, 10506-10510.

Kantola, J., Kunnari, T., Hautala, A., Hakala, J., Ylihonko, K. and Mantsala, P. (2000) Elucidation of anthracyclinone biosynthesis by stepwise cloning of genes for anthracyclines from three different Streptomyces spp. *Microbiology* **146**, 155-163.

Krueger, W.C., Pschigoda, L.M., Schpok, S.L.F., Moscowitz, A., McGovren, J.P. and Neta, P. (1981) The Interaction of Nogalamycin and Analogs with DNA and Other Bio-Polymers. *Chemico-Biol. Interac.* **36**, 1-18.

Müller, I., Jenner, A., Bruchelt, G., Niethammer, D. and Halliwell, B. (1997) Effect of Concentration on the Cytotoxic Mechanism of Doxorubicin - Apoptosis and Oxidative DNA damage. *Biochem. Biophys. Res. Comm.* **230**, 254-257.

Ray, R., Chakraborty, B.K., Ray, K., Mukheriji, S., Chowdhury, J.R. and Panda, C.K. (1996) Effect of anthracycline antitumor antibiotics (adriamycin and nogalamycin) and cycloheximide on the biosynthesis and processing of major UsnRNAs. *Mol. Cell Biochem.* **162**, 75-82.

Schwartz, A., Marrot, L. and Leny, M. (1989) Conformation of DNA modified at a (dGG) or a (dAG) site by the antitumour drug *cis*-diamminedichloroplatinum(II). *Biochem.* **28**, 7979-7984.

Searle, M.S. and Williams, D.H. (1993) On the stability of nucleic-acid structures in solution - enthalpy entropy compensations, internal rotations and reversibiliy. *Nuc. Acids. Res.* **21**, 2051-2056.

Sim, S.-P., Gatto, B., Yo, C., Liu, A.A., Li, T.-K., Pilch, D.S., LaVoie, E.J. and Liu, L.F. (1997) Differential Poisoning of Topoisomerases by Menogaril and Nogalamycin Dictated by the Minor Groove-Binding Nogalose Sugar. *Biochem.* **66**, 13285-13291.

Sim, S.-P., Pilch, D.S. and Liu, L.F. (2000) Site-specific Topoisomerase I-Mediated DNA Cleavage Induced by Nogalamycin: a Potential Role of Ligand-Induced DNA Bending at a Distal Site. *Biochem.* **39**, 9928-9934.

Stryer, L. (1988) *Biochemistry, 3rd Edition* , W. H. Freeman and company.

Trauger, J.W., Baird, E.E. and Dervan, P.B. (1998) Recognition of 16 basepairs in the minor groove of DNA by a pyrrole-imidazole polamide dimer. *J. Am. Chem. Soc.* **120**, 3534-3535.

Wang, A.H.J., Quigley, G.J., Koipak, F.K., Crawford, J.L., van Boom, J.H., van der Marel, G.A. and Rich, A. (1979) *Nature* **282**, 680

Watson, J.D. and Crick, F.H.C. (1953) Molecular Structure of Nucleic Acids. A Structure for Deoxyribose Nucleic Acid. *Nature* **171**, 737-738.

Welch, J.J., Rauscher, F.J. and Beerman, T.A. (1994) Targeting DNA-Binding Drugs to Sequence-Specific Transcription Factor DNA Complexes - Differential-Effects of Intercalating and Minor-Groove Binding-Drugs. *J. Biol. Chem.* **269**, 31051-31058.

Chapter 2. Methods Used in the Study of DNA Intercalators

2.1. Structural Elucidation

The initial work on DNA structure perhaps best exemplifies the relationship between structure and function. The structure revealed the interactions present, explaining the specificity of one strand of DNA for another. This in turn explained the mechanism of semi-conservative replication, leading to an explanation of how DNA can act as an information carrying molecule and pass information on during replication with high fidelity. Knowledge of the structure and interactions present between DNA and drug molecules is important for understanding the function and design of important classes of DNA targeting molecules

Spectroscopic methods have provided valuable insights into drug DNA interactions. The strong UV absorbance of DNA at 260 nm and the fluorescence of some drug molecules provide useful 'handles' for investigating DNA and DNA complexes.

X-ray crystallography has made a large impact into the structural study of DNA complexes. The solving of the single crystal structure of DNA (Drew et al. 1981) provided details of the sequence dependent conformation. Similar techniques have been applied to many drug-DNA complexes revealing details of interactions (Yang and Wang, 1999).

With the advent of more sensitive spectrometers and the use of multi-dimensional experiments, nuclear magnetic resonance has added an extra tool to investigate these important molecules. Experiments reveal distances between hydrogen atoms in an intra- and inter- molecular sense giving information about the conformation of DNA duplexes and contacts with interacting drugs. Since NMR can be carried out in solution some exciting insights regarding the dynamics of molecules in solution and details of the role of solvent molecules have also been revealed (Schwabe, 1997).

Molecular modelling is often used in conjunction with NMR; the models employed are progressively becoming more sophisticated. Information determined experimentally by NMR can be used to drive models towards certain regions of conformational space, thus overcoming difficulties inherent with structural prediction. With the increasing computer power, models that are more complex are being applied to larger biological systems. Molecular modelling has enabled experimenters to approximate aspects of systems that have been hitherto difficult or impossible to measure.

As will be discussed in Chapter 5 all these methods provide complementary and sometimes overlapping information. The combination of techniques provides valuable methods for studying interactions at the molecular level.

2.2. NMR

2.2.1 Introduction

Nuclear magnetic resonance, as a phenomenon, was first described in 1945, but it was in the early 1980s before it was used in structural determination. Interestingly one of the first uses of biological NMR was the investigation of hydration of DNA rather than the structure (Jacobson and Anderson, 1954). These initial experiments were carried out with electromagnets with a B_0 field strength of 0.705 Tesla which corresponds to a proton frequency of just 30MHz. The authors noticed a broadening of the water resonance signal on addition of DNA, which was attributed to an increase in order of the water lattice, or possible exchange phenomena.

Since the early 1960s, NMR has been used increasingly within the field of organic chemistry, where it is used to aid the identification of synthetic products. The technological advances that have been made within NMR over the last 40 years have increased the usefulness of the technique. Increases in magnet strength from simple electro-magnets to liquid helium cooled superconducting magnets, which function at temperatures of 4K, and more recently 2K have improved the sensitivity and resolution. Commercial spectrometers are currently available at fields of up to 21 Tesla corresponding to proton frequencies of 900MHz. With the increase in resolution and sensitivity, NMR has increasingly been applied to large biological systems.

2.2.2 Origin of an NMR Signal

The origin of magnetic resonance stems from the physical property of nuclear spin. Nuclei are observed in a magnetic field to adopt a number of different orientations of differing energies. The number of orientations that can be adopted within a magnetic field is 2I+1 where I represents the spin quantum number for the nucleus. Most NMR active nuclei of interest within biological systems posses a spin number of ½. A notable exception are ^{12}C nuclei that have a spin of 0, the most abundant form of carbon therefore gives no NMR signal. The spin-½ nuclei can adopt one of two orientations within a magnetic field referred to as + ½ and – ½. The nuclei that adopt an orientation coincident with the magnetic field are of a lower energy then those aligned against, this energy difference is the basis of the NMR signal (Figure 2-1).

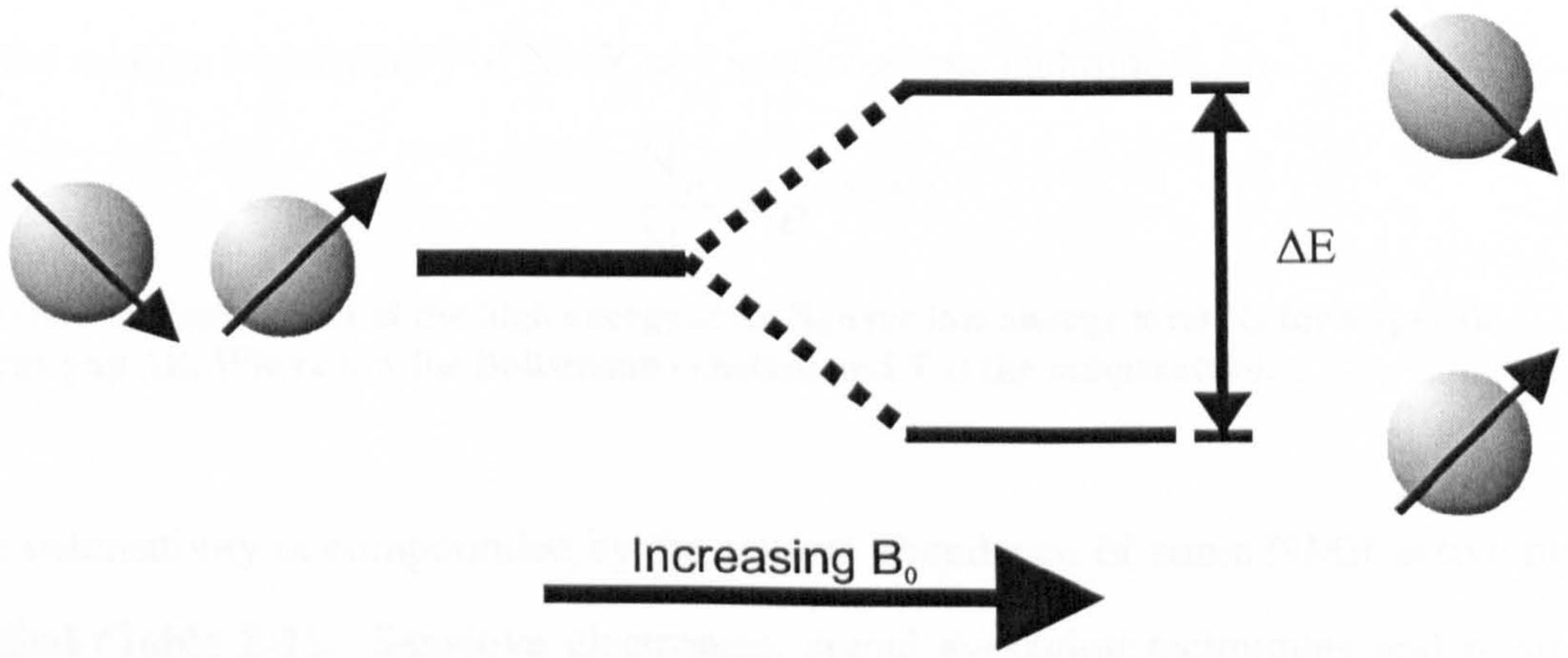


Figure 2-1: Energy levels adopted by spin-½ nuclei in the presence of a magnetic field B_0.

This difference can be quantified by Equation 1 where B_0 is the magnetic field strength typically in Tesla, h is Plank's constant, and γ is the magneto-gyro ratio, a constant specific to the isotope being observed.

$$\Delta E = \frac{\gamma h B_0}{2\pi}$$

Equation 1: Energy seperating the high and low energy spin states.

The energy required to excite the transition between the two energy levels occurs in the radio frequency (rf) region of the electromagnetic spectrum. Magnet strengths are often described in terms of the frequency at which protons resonate, for example a 2.35 Tesla magnet causes proton resonance at about 100 MHz. The distribution between the energy levels can be determined by a Boltzmann equation. This shows the excess of nuclei resident in the lower energy state at thermal equilibrium (Equation 2). The equation reveals there is only an excess of one ^1H nuclei in the low-energy state for every hundred thousand nuclei in the high-energy state at 298 K and a field strength of 2.35 Tesla (100 MHz). This small excess even in very strong magnetic fields accounts for the relative insensitivity of NMR as a spectroscopic technique.

$$\frac{N_\beta}{N_\alpha} = e^{(-\Delta E/kT)}$$

Equation 2: Populations of the high energy state N_β over low anergy level N_α for a specific energy gap ΔE. Where κ is the Boltsmann constant and T is the temperature.

The insensitivity is compounded by the natural abundance of some NMR active nuclei studied (Table 2-1). Sensitive electronics, signal averaging techniques and relatively high concentrations of sample, compared with other spectroscopic methods, are required to be able to obtain spectra of an interpretable quality.

Table 2-1: Sensitivities of commonly used isotopes

Nucleus	Spin	γ	Natural Abundance (%)	Relative Sensitivity	Absolute Sensitivity
^{1}H	½	$2.68\times10^{+08}$	99.98	1.00	1.00
^{2}H	1	$4.11\times10^{+07}$	0.015	9.65×10^{-3}	1.45×10^{-6}
^{13}C	½	$6.73\times10^{+07}$	1.108	1.59×10^{-2}	1.76×10^{-4}
^{15}N	½	$2.71\times10^{+07}$	0.37	1.04×10^{-3}	3.85×10^{-6}
^{31}P	½	$1.08\times10^{+08}$	100	6.63×10^{-2}	6.63×10^{-2}

For reasons of sensitivity ^{1}H nuclei are the principal nuclei detected by NMR experiments. For an 11.74 Tesla magnet the ^{1}H nuclei comes into resonance around 500MHz. However, the actual resonance frequency is dependent upon the environment of the nucleus. Magnetic fields created by the electrons near the nucleus serve to modulate the effective magnetic field that is experienced.

Although it is possible to record a spectrum by scanning through the frequency range and noting the value at which each nuclei comes into resonance, this is seldom done in practice. Pulse Fourier Transform NMR (FT-NMR) is the preferred experimental method. FT-NMR offers a number of advantages and is essential for multidimensional NMR experiments. The use of one pulse to excite all resonances simultaneously reduces the experiment time. This reduction in experiment time means many repetitions can be performed and the signal averaged leading to an increase in the signal to noise ratio. This is especially important considering the inherent insensitivities of NMR.

The use of pulse experiments also allows the magnetisation of the sample to be manipulated more precisely. Sequences of pulses can be arranged so that properties such as through bond and through space couplings evolve during the course of the experiment.

Although a more rigorous explanation is needed to describe some of the multidimensional experiments, a vector model explanation gives a good description of a simple pulse NMR experiment. The nuclei in the magnetic field, B_0 precess around the axes parallel to the field. The nuclei precess at different frequencies around the Z-axis, and the nuclei can be imagined as being oriented around a cone pointing along the Z-axis (Figure 2-2a). The NMR signal is a bulk property and the net magnetisation can be considered as aligned along the Z-axis.

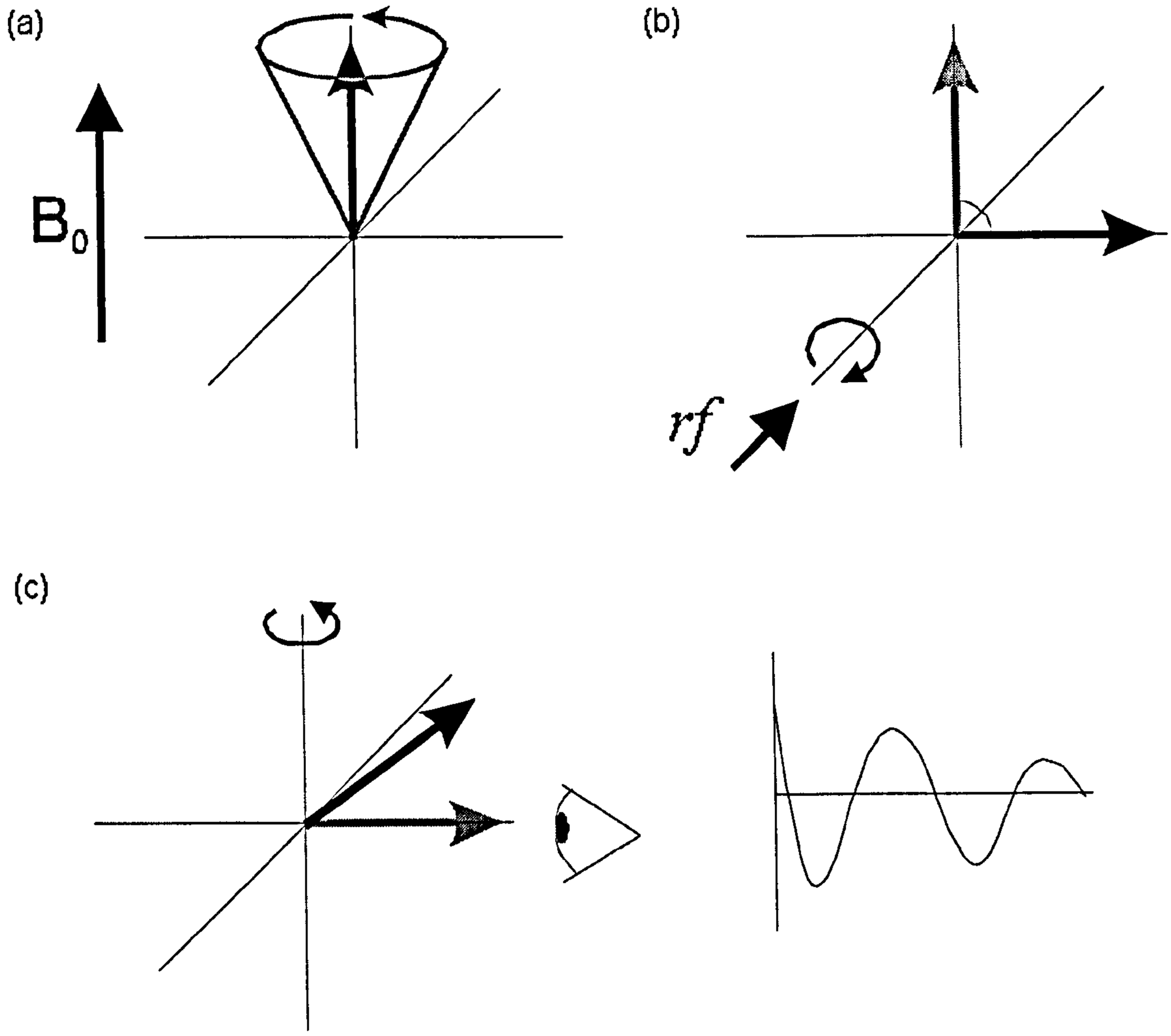


Figure 2-2: Vector model of pulse NMR. Magnetisation undergoes precession around the magnetic field (B_0), which is oriented along the z-axis by convention. The net magnetisation is represented by the black arrow oriented with the z-axis (a). A pulse causes a rotation of the net magnetisation and is timed such that the magnetisation is rotated into the XY plane (b). Precession occurs in the XY plane and is detected as a sine wave by a detector placed at X (c). The magnetisation gradually relaxes back to the equilibrium position coincident with the B_0 field.

Upon an application of a rf pulse, the net magnetisation is rotated into the XY plane. The actual angle rotated is dependent upon the strength (power) and time that the pulse is applied for, and these are under the control of the experimenter. To position the magnetisation ready for acquisition a $\pi/2$ pulse is normally applied, which is sufficient

to cause rotation by $\pi/2$ radians or 90°. This causes magnetisation to be rotated from equilibrium into the XY plane. The nuclei undergo precession around the B_0 field giving rise to a sinusoidal reading of intensity versus time. This signal gradually decays as the thermal equilibrium is restored. The plot of intensity versus time is referred to as the free induction decay or FID. The frequency of the FID is the frequency of the proton resonance. In practice, the FID obtained comprises of the sum of many overlapping sine waves and the individual frequencies are obtained by a Fourier transform.

2.2.3 Two Dimensional Techniques

The concept of two-dimensional NMR was first proposed in 1971 (Jeener, 1971). Since that time a number of experimental techniques have been developed. All two-dimensional experiments contain a variable time delay (t1 delay) that modulates some observed property. FIDs are recorded for each value of this delay. Fourier transform of the FIDs obtained gives the familiar intensity versus frequency plot, however each frequency point is also modulated by the t1 delay. This time domain signal can be subjected to the normal Fourier transform analysis and a plot of intensity at two frequencies obtained. The pulse sequences of two-dimensional experiments consist of three sections (Figure 2-3). Preparation sets up the magnetisation transfer, mixing allows transfer to occur, and acquisition causes magnetisation to enter the XY plane for detection.

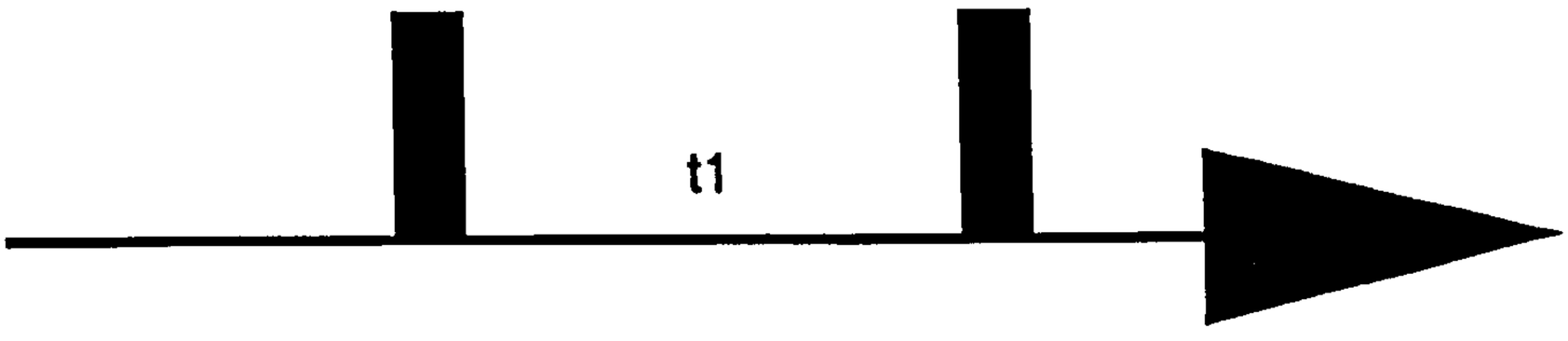


Figure 2-3: Stages of a two-dimensional experiment. Pulses (90°) are represented by rectangles and acquisition by the arrow head. All experiments consist of the three main stages however the number of pulses in each stage varies.

There has been an explosion in the number of techniques available in NMR, and a large number of experiments are available that reveal different components of the molecular structure. The next sections describe some of the important experiments used for investigating structures of DNA and DNA-drug complexes.

2.2.4 J-Correlated Experiments

The above explanation into the origin of an NMR signal considered only one isolated nucleus, but such a system occurs rarely. The presence of a nearby nucleus that can exist in one of two states causes a fluctuation in the magnetic field experienced by the nucleus being observed. This is referred to as a scalar coupling or J coupling, and the effect is mediated through bonds. The limit to the range of coupling is about three σ-bonds, however, longer range couplings are observed if π-bonds are involved. The ability of one nucleus to affect another allows information about connectivity to be determined.

The Correlation Spectroscopy (COSY) (Bax and Freeman, 1981) experiment is the most basic two-dimensional experiment, consisting of just two pulses separated by the t1 delay. The COSY reveals direct through-bond coupling between protons. Cross peaks appear at the chemical shifts of the two protons involved (Figure 2-4). The COSY experiment suffers from a number of experimental difficulties, but is improved by the application of a double quantum filter (DQF-COSY) (Rance et al. 1983; Piantini et al. 1982).

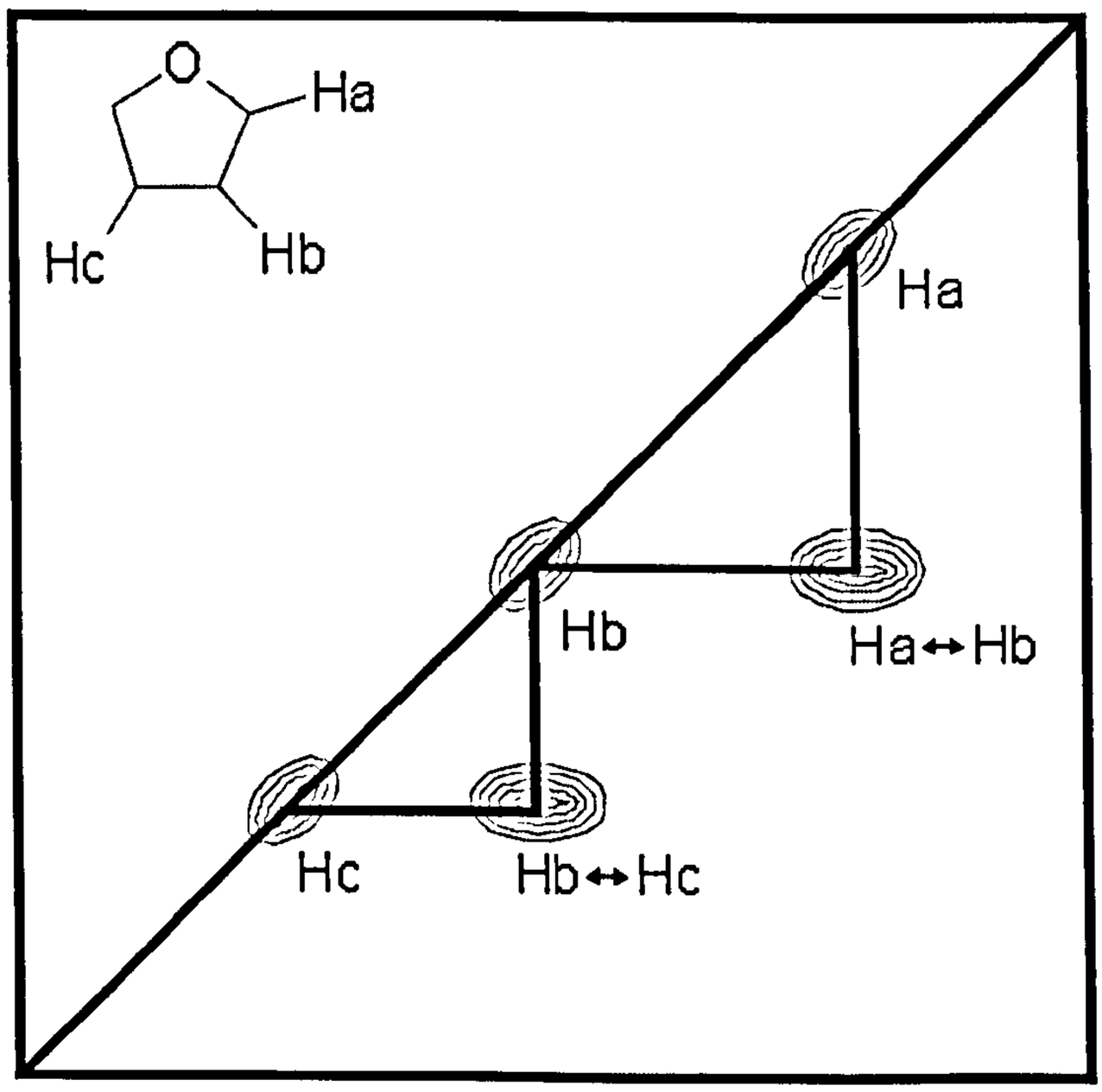


Figure 2-4: Schematic of a COSY experiment with only peaks below the diagonal shown. Cross-peaks occur at the frequencies of coupled protons (3 bonds away).

The DQ-filter eliminates non-coupled resonances from the spectrum and also has much more defined diagonals, allowing peaks which are closer to the diagonal to be resolved. The peak shapes observed with the DQF-COSY are more complex than the basic

COSY, being anti-phase in nature and give rise to a distinctive peak consisting of alternating positive and negative levels.

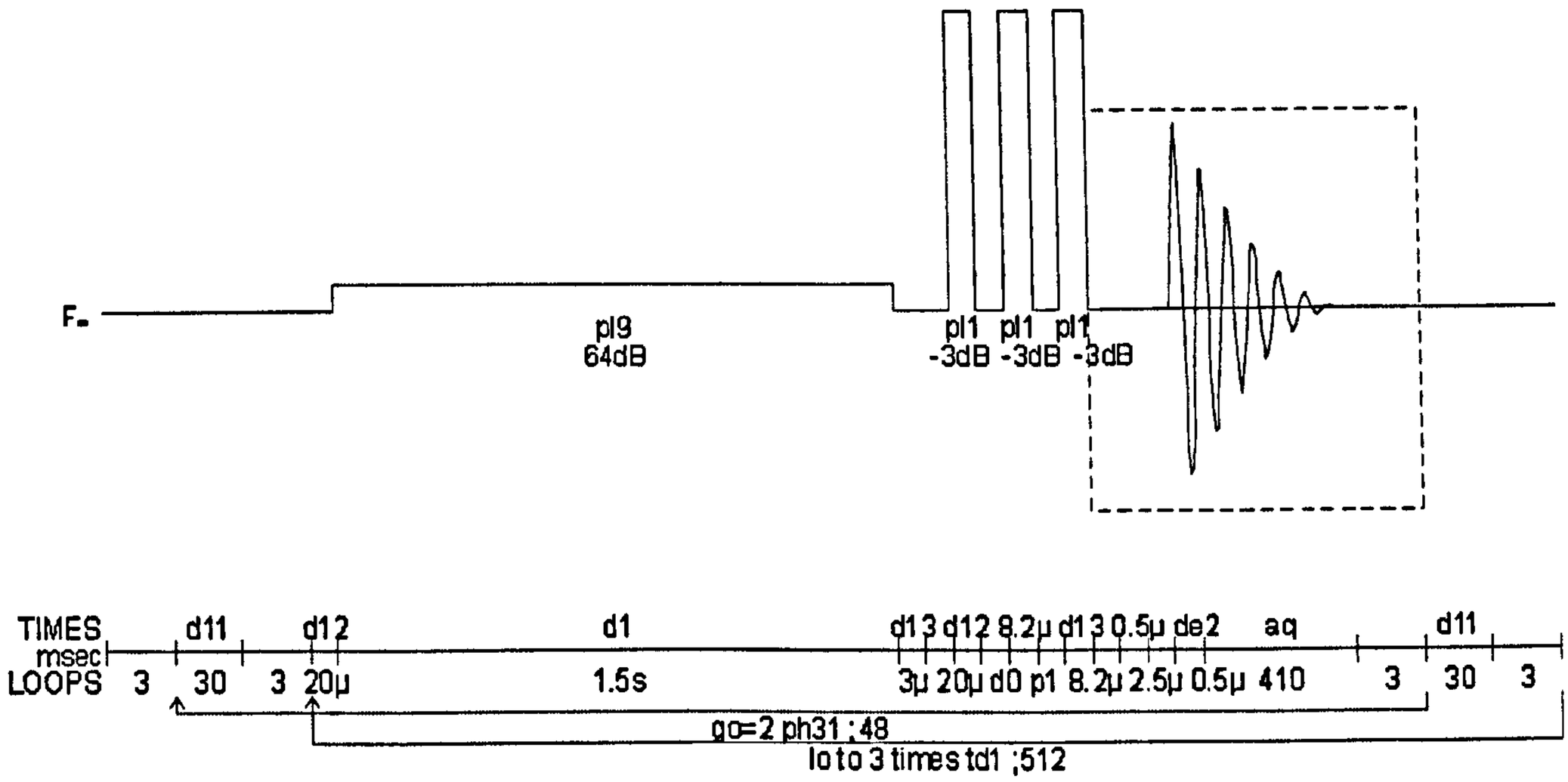


Figure 2-5: Pulse sequence to acquire a DQF-COSY experiment. The initial pulse is a saturation pulse to reduce residual signals from solvent. The remaining three pulses are 90° pulses that effect transfer between coupled nuclei. The signal is then acquired.

The Total Correlation Spectroscopy (TOCSY) experiment (Bax and Davis, 1985a) complements the DQF COSY by highlighting J couplings through magnetisation transfer. Cross-peaks are observed of gradually decreasing intensity between directly coupled protons and protons that share a mutually coupled proton. The TOCSY thus highlights protons related through J-coupling (spin-systems) and is especially useful when used to investigate couplings in nucleotide sugars (Figure 2-6). Cross-peaks are absorptive, giving all positive peaks. This removes the chance of cancellation effects that can occur in DQF-COSY experiments when positive and negative levels are overlapping.

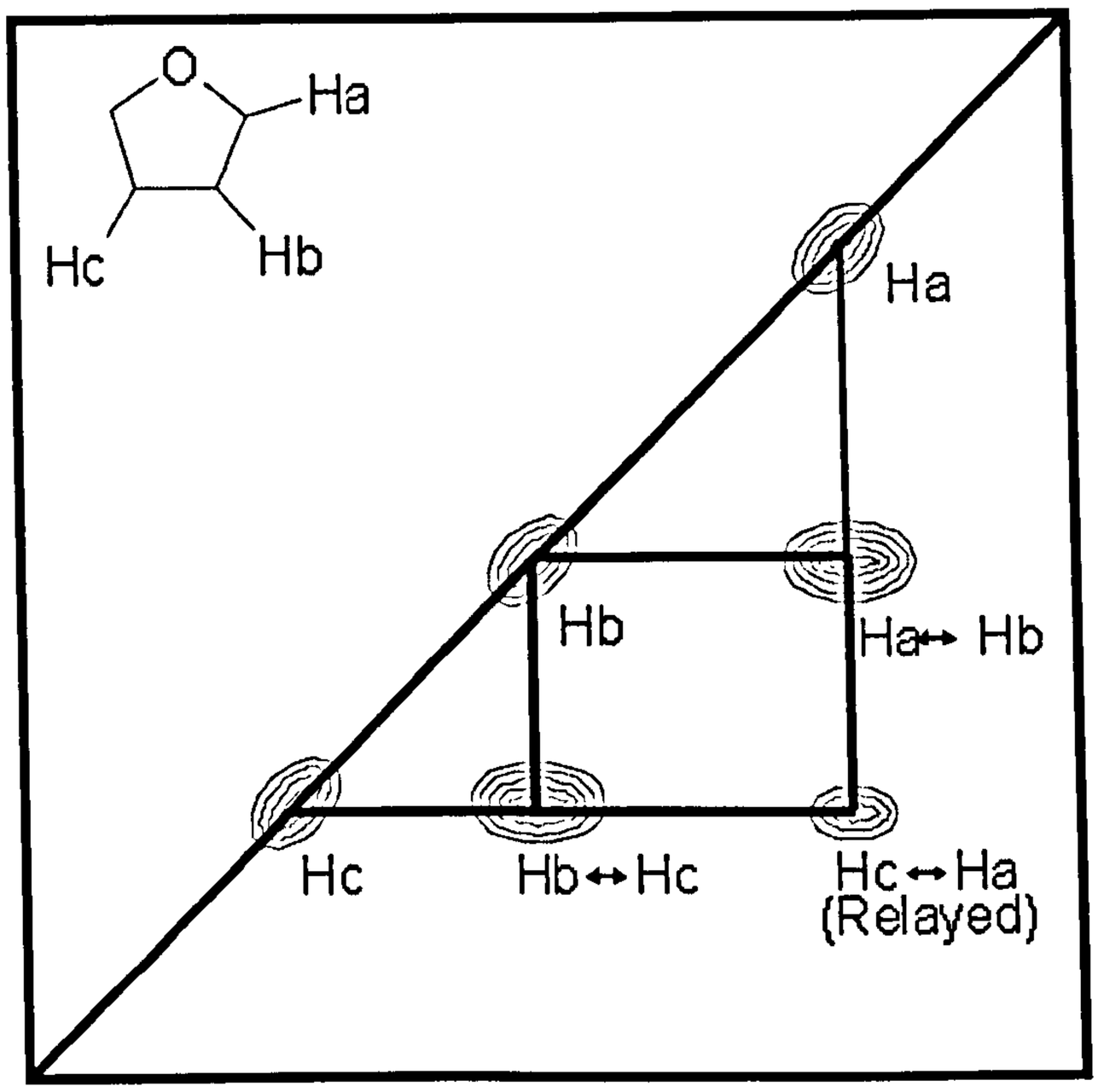


Figure 2-6: Schematic of a TOCSY experiment with only peaks below the diagonal shown. Cross-peaks occur at the frequencies of coupled protons and between protons that share a mutually coupled intermediate. Thus highlighting entire spin-systems.

2.2.5 Heteronuclear J-Correlated

Heteronuclear J-correlated spectra provide information about proton connectivity typically to carbon, nitrogen or phosphorus. Cross-peaks are observed which occur at the frequency of the ^{1}H and the hetronucleus being studied. Traditionally this was termed a ^{13}C-^{1}H COSY however the advent of new techniques, such as the Heteronuclear Multi-Quantum Coherence (HMQC) experiments (Bax et al. 1983) allow the direct information to be recorded for the ^{1}H nucleus, thus affording the sensitivity benefits of acquiring on the most sensitive nucleus. A related experiment, the Heteronuclear Multi-Bond Connection (HMBC) experiment (Bax and Summers,

1986), is able to highlight coupling over a longer range, with protons showing couplings to heteronuclei that are two or sometimes more positions away.

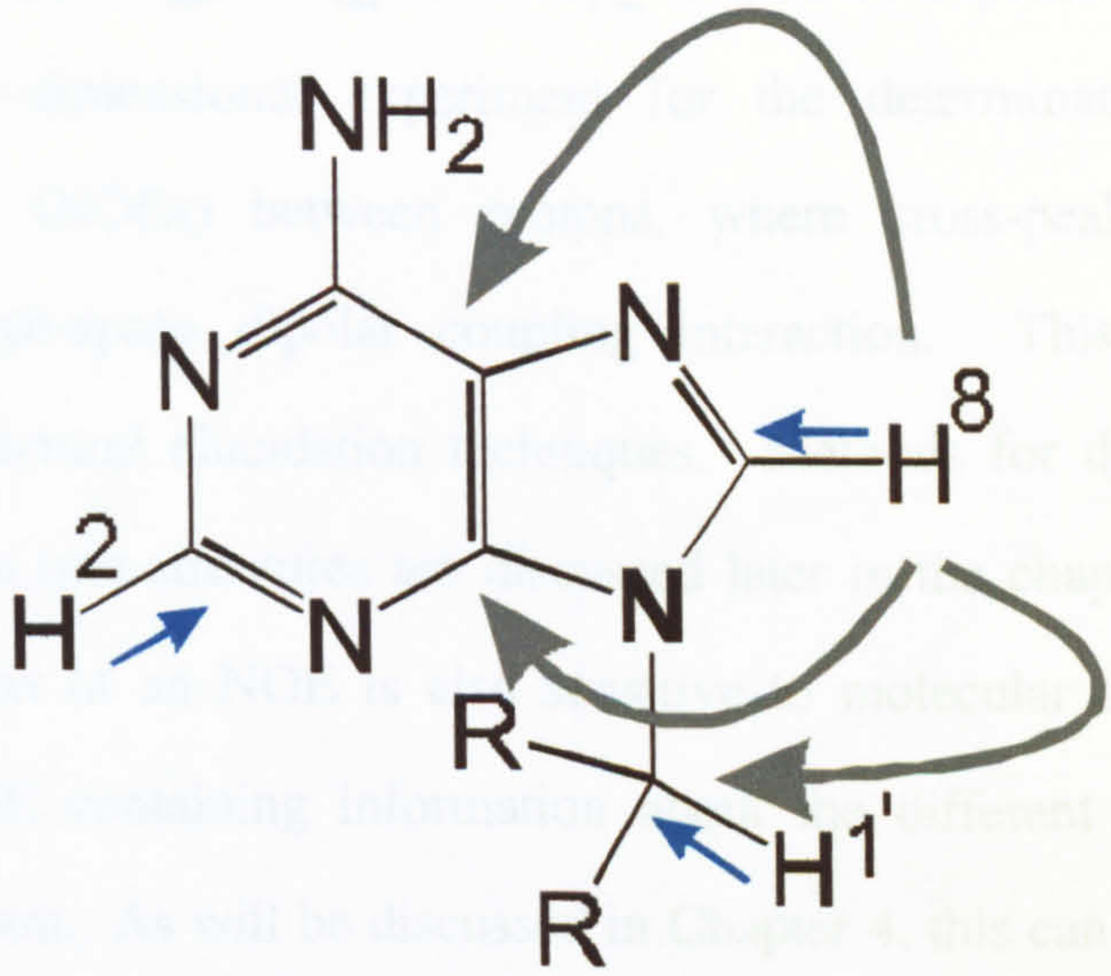


Figure 2-7: Connectivities revealed by a combination of HMQC (blue arrows) and HMBC (grey arrows) experiments for an adenine base. Not all HMBC peaks are shown for clarity, however carbons two or three positions away from the proton are visible. Connection between the H1' and C1' carbon can be linked via the HMBC connection to the H8. Similar pathways allow resolution of ambiguous drug protons.

In combination, these two techniques provide an extremely powerful way of determining the backbone connectivity of a molecule. Although more typically employed in drug or natural product assignment, a natural abundance HMQC ^{1}H–^{13}C can be applied to DNA, and gives useful information for the localisation of A-H2. An example of this A-H8 -> A-H1' pathway can be seen (Figure 2-7).

2.2.6 The NOE and NOESY experiments

The Nuclear Overhauser Effect Spectroscopy (NOESY) experiment (Bax and Davis, 1985b) is a two–dimensional experiment for the determination of the nuclear Overhauser effect (NOEs) between protons, where cross-peaks between protons represent a through-space dipolar coupling interaction. This experiment is the cornerstone of structural elucidation techniques. Methods for the conversion of the NOE intensity data into structures are discussed later in the chapter. As is described below the evolution of an NOE is also sensitive to molecular tumbling rates. This results in the NOE containing information about the different correlation times of molecules in solution. As will be discussed in Chapter 4, this can provide insights into the nature and kinetics of solvent interactions formed between drug molecules and DNA.

One-dimensional techniques that determine NOE are in widespread use, but in macromolecular systems, the ability to selectively irradiate specific protons is usually hampered by overlap.

Spin-Spin relaxation is the origin of the nuclear Overhauser effect (NOE). The ability of one nucleus to relax another is highly distance dependant, only being effective within approximately a 4.5 Å radius. Since the effect is proportional to $1/r^6$, quantification of the NOE allows an accurate determination of the distance and a means of studying through-space interactions.

If we look at an energy level diagram (Figure 2-8), ignoring scalar coupling we can see that an excitation of nucleus A would stimulate the transitions W_{1A} (Figure 2-8 levels

(1) to (3) and (2) to (4)). This serves to increase the population of the levels labelled (4) and (3), the excited states of nucleus A. This is the effect of a normal pulse experiment.

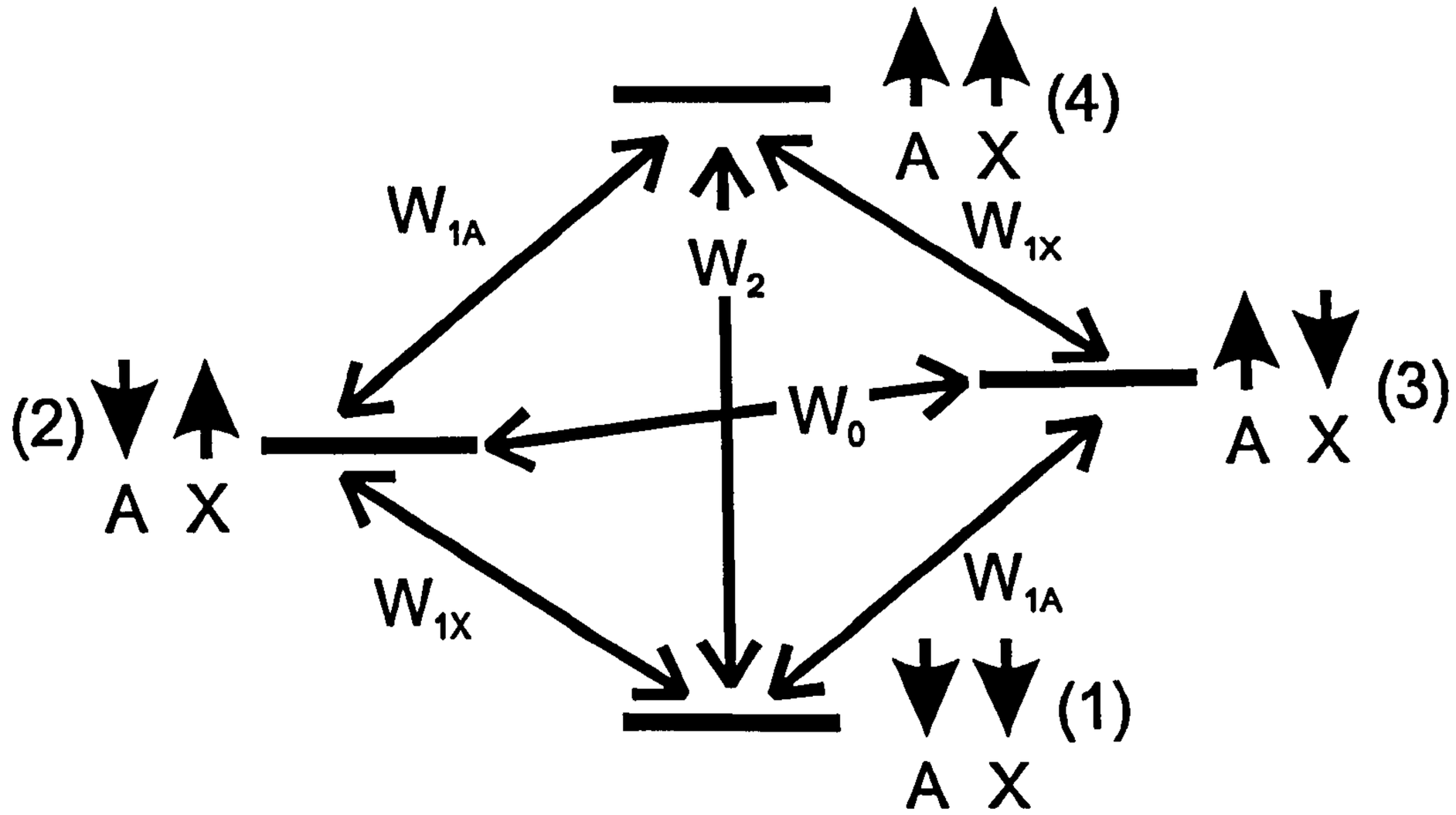


Figure 2-8 Transitions that give rise to the nuclear Overhauser effect (NOE) in two isolated spins. Energy levels for two spatially close but not necessarily coupled nuclei A and X. Energy levels (1) and (3) represent ground states for nucleus X, and (1) and (2) represent ground states for nucleus A. Transitions W_1 caused by irradiation at the correct frequency leads to the relevant transition to the excited states. Relaxation pathways W_0 and W_2 lead to a perturbation in the populations of each energy level and affect the intensity of transition for the non-irradiated nuclei.

Only transitions which obey the selection rule are visible; for those satisfying $\Delta M=1$. However, the relaxation has no such selection rule, and although not directly observable, relaxation affects the populations of the energy levels. The system will endeavour to restore the equilibrium and can do this by employing a transition between level 1 and level 4 (W_2) which is a double quantum transition or between levels 2 and 3 (W_0) a zero-quantum transition.

If the W_2 path is taken then there is a population transfer from an excited state for nucleus 'A' to a ground state of nucleus 'A' (level 4 to level 1). The relative populations of the 'X' nucleus excited states (level 2 and 4) then decrease relative to the ground states (levels 1 and 3). Thus, there is an enhancement of the detected intensity of the 'X' nucleus, this is a positive nuclear Overhauser effect (+NOE). The Zero-quantum transition, W_0, causes a transfer from the other excited state for nucleus 'A' to the other ground state of nucleus 'A' (level 3 to level 2). However, this leads to a decrease in the relative populations of the ground states of nucleus 'X' (levels 1 and 3) compared to the excited states (levels 2 and 4). There is a decrease in the intensity of the X nucleus on the irradiation of the 'A' nucleus, this is a negative nuclear Overhauser effect (-NOE). The irradiation of a nucleus 'A' therefore affects the intensity of a spatially close nucleus 'X'.

If both relaxation pathways operated equally there would be no net gain or loss and therefore no observable nuclear Overhauser effect. However, both pathways are not favoured equally. The W_2 pathway is favoured by high frequency fast tumbling molecules, typically small and in non-viscous solvents. For this class of molecules, positive NOEs are observed. Larger molecules are relaxed by the W_0 pathway and show negative NOEs. The relative proportions are determined by not only molecule size and shape but also by magnet strength and temperature. These are important considerations since at the cross over point the net effect is zero and little information can be gleaned.

The Rotational Overhauser Effect Spectroscopy experiment (ROESY) (Schleucher et al. 1995) is a related experiment to the NOESY and also gives cross-peaks between protons that are close in space, relayed through dipolar couplings. The ROE obtained

evolves via a different mechanism to the NOE and is always positive regardless of molecular size and tumbling rate. This experiment does not suffer from a null point in its build up like NOE detecting experiments, and is suitable for molecules that are in this intermediate zone of the NOE build up curve. The ROESY experiment is also useful in the determination of chemical exchange effects. Since protons undergoing chemical exchange result in a positive peak in the ROESY spectrum, this is clearly distinguishable from the negative ROE peaks. ROESY experiments are very difficult to interpret in terms of distance since the intensities of ROE cross peaks are modulated by a number of factors other than inter-proton separation. These difficulties mean ROESY experiments are seldom used to determine distance restraints.

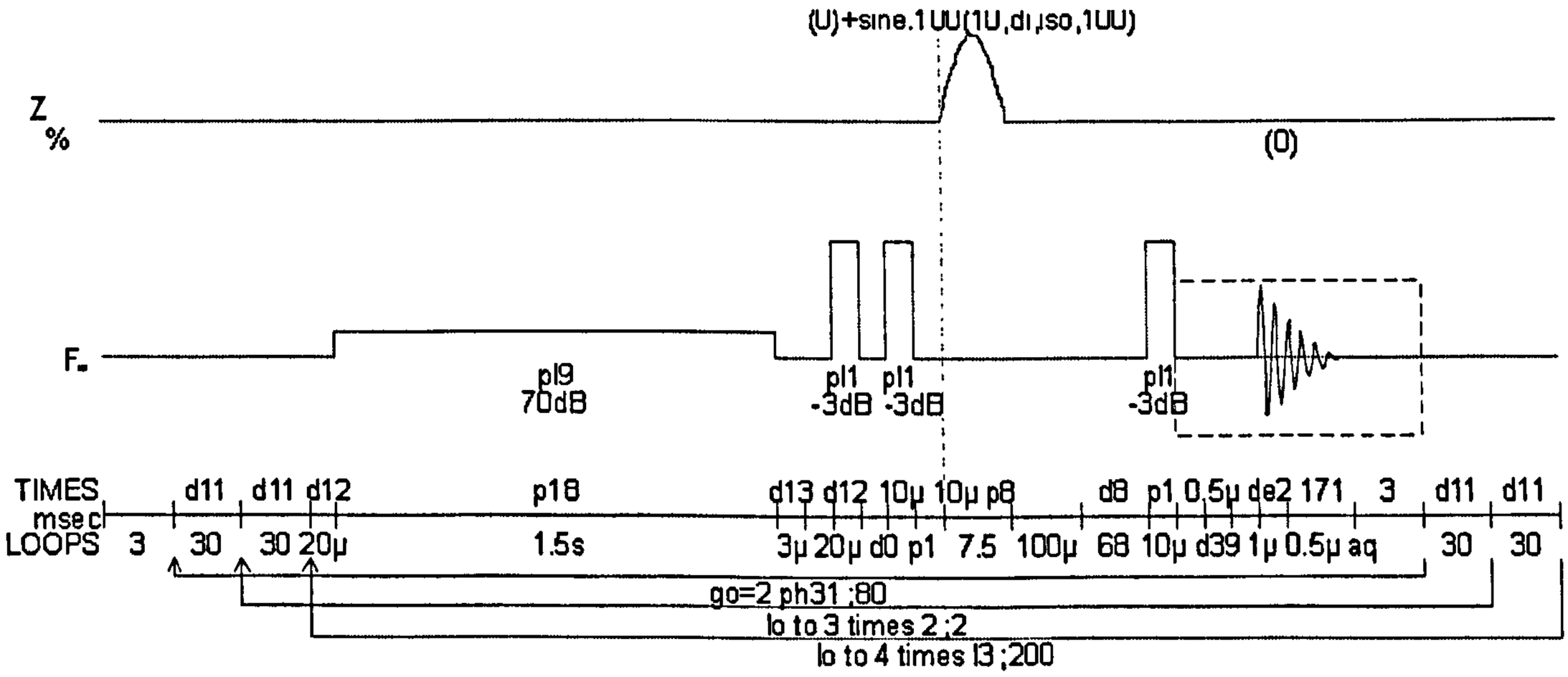


Figure 2-9: Improved NOESY using a gradient pulse during the mixing time to remove zero-quantum artefacts.

The original NOESY sequence suffers from a number of experimental difficulties. The most problematic is the presence of the zero-quantum effects. This is caused when strongly coupled peaks are present and coupling artefacts are superimposed upon the

normal cross-peaks, distorting intensities. These effects are more noticeable when the NOE mixing time is short (<100 ms), a condition required for integration to minimise spin-diffusion effects. The use of gradients provides one of the best methods for removing these artefacts (Figure 2-9). The application of a gradient during the mixing time destroys magnetisation aligned along the Z-axis that gives rise to these artefacts (Otting et al. 1990).

2.2.7 Investigation of Exchangeable Protons

The imino protons at thymine N3 and guanine N1 are useful probes into base pairing, since the formation of a Watson-Crick hydrogen bonding pattern protects these protons from exchange with solvent. To study these protons and also the amino protons it is necessary to use ^1H_2O as a solvent. The usefulness of being able to study the imino protons and the solvent structure itself can be seen in Chapter 4 and Chapter 5.

The use of ^1H_2O as a solvent presents an obvious difficulty due to the concentration of protons in the solvent. There are a number of strategies to effect the removal of an excessively large solvent peak. Residual HOD signal is normally removed by pre-saturation at the solvent frequency. This technique is not useful in ^1H_2O if the aim is to study imino protons since saturated protons exchange with the imino protons reducing the intensity to near zero.

The jump-return sequence (Griffey et al. 1985) consists of two 90° pulses separated by a delay. The first pulse moves the magnetisation onto the X-axis and allows precession for a specified time. The delay is set so the frequency of interest will align along the Y-axis. A second pulse then returns the magnetisation on the X-axis, which includes the

on-resonance water signal, to the Z-axis leaving any magnetisation oriented on the Y-axis to be acquired. The intensity of the spectra is modulated by a sinusoidal pattern, with the on-resonance solvent signal at zero intensity and the frequency of interest at maximum intensity (Figure 2-10).

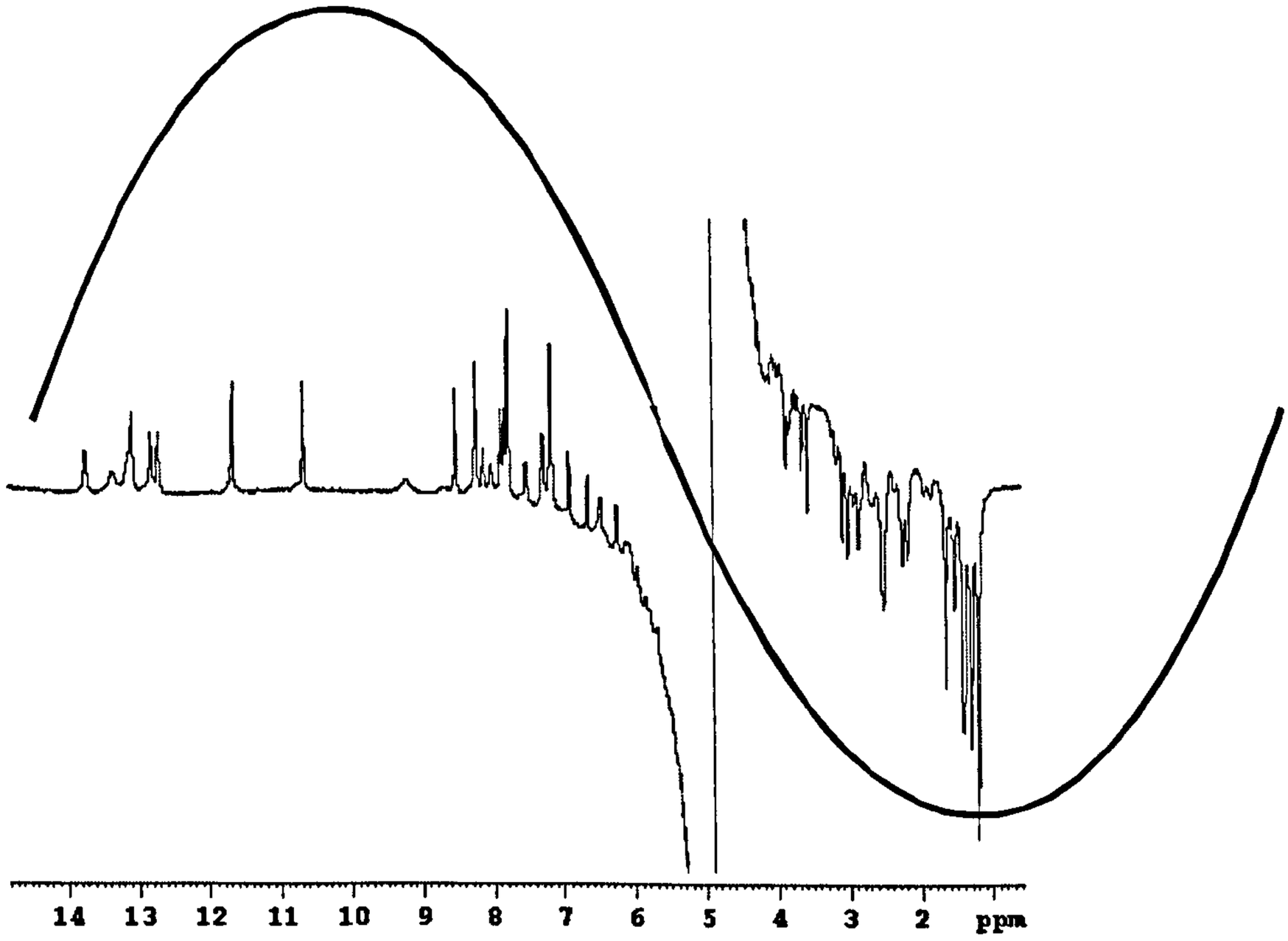


Figure 2-10: 1D ^{1}H experiment for a sample in H_2O acquired with a 1:1 Jump-return filter to suppress solvent. The intensity modulation is indicated by the overlaid sinusoidal plot. Although the intensity around the H_2O signal at 4.8 ppm is zero, significant distortion is observed making resolution of protons in this region difficult.

The development of gradient techniques has improved the method of solvent suppression. Gradients provide a reversible method of dephasing (un-aligning) the magnetisation vector and therefore abolishing an NMR signal. The **Water** Gradient Attenuation by Tailored Excitation (WATERGATE) sequence (Piotto et al. 1992) is particularly efficient at solvent suppression and generates few artefacts. A gradient

dephases the magnetisation and a pulse sequence applies a selective 180° pulse that affects all but the on-resonance signal. A second gradient reverses all dephased signals apart from that of the on-resonance solvent. Gradients apply a position specific magnetic field that is only reversible if the molecules are not diffusing rapidly; this is not the case for water molecules, and the effect serves to further reduce the magnetisation of the solvent.

2.3. Derivation of Distance Restraints

The conversion of information from NOESY spectra into distance restraints presents a number of complicating issues. The NOESY experiment reveals protons that are close in space and the intensity is proportional to the distance separating the two protons. However, the intensity can be modified by additional cross-relaxation to a different nucleus. The NOE is allowed to build up and transfer during the mixing period of the experiment and this is under the control of the experimenter. The mixing time is usually in the millisecond range, and for large biological molecules typically ranges from 50-300 ms. Although the mixing time is under experimental control the rates of build up and transfer vary for each proton pair, complicating the analysis of intensity, depending on local motional anisotropes.

Table 2-2: Fixed distances between easily identifiable protons.

Interacting Nucleotide Protons	Reference Distance (Å)
Cytosine: H5→H6	2.45
Thymine: H6→CH_3	3.00
Any deoxyribose: H2'→H2"	1.84

A number of factors affect the rate of the NOE build-up and therefore the intensity of the observed cross-peak. The relaxation rates observed for methyl groups are generally faster than those to methine groups since an individual proton is already close to two others in a methyl environment. In DNA some fixed distances defined by the covalent structure are available that correspond to readily resolvable NOESY cross peaks that allow normalisation of intensities in similar environments (Table 2-2). Equation 3 can be used to normalise the measured intensities against the NOE intensities for a proton pair with a known separation that are in similar environments, giving improved distance measurements (Searle, 1993). As will be discussed next this assumes the two protons involved are isolated. This is often referred to as the isolated spin pair approximation (ISPA) and in more complicated situations this rarely holds, and further refinement is needed.

$$D = \sqrt[6]{\frac{D_{ref}^{6}}{(I/I_{ref})}}$$

Equation 3: Determination of a distance between two protons, *D* for an cross-peak intensity *I*. Calibrated for a known proton seperation D_{ref} with a measured intensity I_{ref}.

2.3.1 Relaxation Matrix Approach

The above approach fails to completely consider the build up and transfer of the NOE between proton-pairs in a complex system (spin diffusion). One approach to determine structures from NOE intensities is the relaxation matrix approach. Equations can be determined for each proton that calculates the spin-diffusion and network relaxation of the NOE on each nucleus. This requires an estimate of the correlation time of the molecule and assumes isotropic motion, that is the whole molecule acts as one with, one global correlation time.

Determination of the full relaxation matrix is experimentally difficult, since not all the cross peaks and diagonal peaks are resolved in the NOESY experiment and a substantial part of the total magnetisation is below the noise level of the experiment. The approach adopted has thus been an iterative one based on use of well resolved signals from the NOESY and a knowledge of the initial structure which can be used to calculate the remaining terms in the matrix. Determination of the distances based on the analysis allows the structure to be refined and the process is repeated until the structural-derived intensities match the experimental data.

Structures are then normally refined by a molecular dynamics calculation using the determined distance restraints to modify the structure until a match between theoretical and experimental intensities is reached. Typical protocols use high force constants and high temperatures to overcome local energy barriers and force the structure to adopt a

specific conformation (Bertini et al. 1999). With such an approach, the high energies used override the force constants involved in the force field and produce a structure that fits the NOE data.

Since the NOE intensity represents the average overall available conformations sampled during the mixing times, the distances obtained also represent an average structure that is weighted by the $1/r^6$ term. Distances obtained from 2D NOE intensities can therefore be inconsistent with any one physically realistic structure. (Brandan and Thomas, 1989; Behling et al. 1987). This is particularly important when the dynamic nature of these molecules is considered (Boltan and James, 1979)

2.3.2 Distance Extrapolation

A variation on the above method is to consider that the distance restraints may not completely describe a single realistic structure. The use of more sophisticated modelling techniques allows the computed structures to evolve under the influence of a forcefield, but restrict conformations using the experimentally derived restraints. The use of lower temperatures and lower energy restraints allows unrealistic structures that may be described by the averaged NOE data to be avoided, since restraints violation by the forcefield is possible. Restraints can also be used as upper-bound restraints, this approach leads to flat energy wells that give room for dynamic movement, while still using the experimental data to restrict the conformations available.

Distances for restraints can be determined by an approach described by Baleja et al. (Baleja et al. 1990) that compensates for spin-diffusion by extrapolation of the distance to zero mixing time.

Distances are determined according to Equation 3 above. The distances determined as a function of mixing time can be approximately fitted to a linear equation, and this holds especially for short mixing times (less than 150 ms). The slope of the line represents the rate of gain or loss by spin-diffusion of magnetisation relative to the reference used to calculate the distance. Extrapolation to a 'zero-mixing' time yields distances corrected for spin-diffusion effects.

2.4. Molecular Modelling

2.4.1 Quantum Mechanical Methods

With the advent of powerful computers molecular modelling has become a realistic technique for investigating the structure of large biological macromolecules. Modelling allows us to probe and even quantify many properties of a system but the data obtained are only as good as the model used. The increase in sophistication is yielding data closer to experimental results.

In order to describe completely a system it is necessary to know its wave equation since only this quantifies a molecule's energy. This is only completely achievable for single electron systems such as the hydrogen atom (Grant and Richards, 1995). For any more complex systems, approximations are required. The methods for approximating these equations are well documented, however as will be detailed later it is not always necessary to determine the wave function in order to resolve issues of molecular structure.

Ab initio methods for determining molecular orbitals tend to revolve around the methods developed by Hartree and Fock using a self-consistent field method. Starting values are determined and used for the coefficients for the molecular wave equation, the results of which can be applied in an iterative manner to improve the value of the various coefficients. This results in an orbital with the lowest possible energy. Although such approaches can give an accurate picture, the complexity of the calculation is too great to be applied to the thousands of atoms present in biological systems. Such approaches do allow us to characterise small subsets of biological macromolecules, such as the amino acids and nucleotide bases of DNA and RNA. Parameters such as atomic charge density along with experimental information, can be gathered with only a small expenditure of computer time and this information can be utilised in other computational approaches.

2.4.2 The Role of the Forcefield

If the aim of a computational method is to yield structures, we can make a number of simplifying approximations. Since we are only interested in one facet of a molecule, the detail obtained by *ab initio* methods can discarded. One method employed to the solution of molecular structure is molecular mechanics. The approach ignores the computationally expensive quantum mechanical methods and instead uses a classical mechanical model. The problem of interacting nuclear and electronic functions in *ab initio* methods is sidestepped by considering only nuclear motion with a fixed electron distribution. Parameters can be derived for all possible connections of these nuclei. This approach has been termed a 'ball and stick' approach because each nucleus is connected to others by a highly characterised spring. Molecular mechanics calculations rely on a set of parameters derived for each nuclei or groupings of nuclei and these are

referred to collectively as the forcefield. There are a number of terms to consider, although these vary subtly with different implementations. A widely used forcefield for biological systems is the AMBER 94 forcefield (Cornell et al. 1995), which contains parameters for all naturally occurring amino acids and nucleotides.

2.4.3 Non-Bonding Terms

Non-bonding interactions consist of the van der Waals and electrostatic interactions. The energy for the van der Waals interaction is typically calculated from the Lennard-Jones potential function. This is sometimes referred to as the 6-12 function since repulsive terms are accounted for by a r^{-12} term and attractive by a r^{-6} component. This function reproduces the van der Waals term well with low computational overheads. The two constants A and B are calculated for each atom pair and are additive in nature.

$$E_{vdw} = \sum \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}$$

Equation 4: Non-bonding van der Waals energy calculated using a Lennard-Jones (6-12) potential. Where r_{ij} is the separation of two protons and A and B are constants based on the atoms involved.

Non-bonding electrostatic interactions comprise the remainder of the non-bonding forces to be considered, and are normally calculated using Coulomb's Law. Electrostatic interactions are based on partial charges assigned to individual nuclei, q in Equation 5 the distance separating the nuclei, r and the dielectric scaling constant D. The force decays at only r^{-1} and, as will be discussed later, this causes some difficulty due to the slow decay and long range nature of this non-bonding force.

$$E_d = \sum q_i q_j / D r_{ij}$$

Equation 5: Electrostatic non-bonding interaction between a pair of nuclei based on Coulomb's Law. Where $q_i q_j$ is the product of the partial charges and r_{ij} is the separation, D is the dielectric constant.

The electrostatic interactions are perhaps the most difficult to model accurately in the forcefield. The partial charges used are normally derived from *ab initio* calculations and these have been shown to be dependent on conformation (Bayly et al. 1993). Furthermore, the charges obtained are simple point charges and therefore spherical in nature. The detail of the potential surface observed in some systems is lost. The ability of these point charges to reproduce, especially out of plane electrostatics, has been questioned (Hunter, 1993). Stacking interactions, for example, between nucleotide base pairs, are sensitive to the fluctuations in the electrostatic energy surface and the simple approach fails to model these systems accurately. Methods have been suggested that overcome these problems without increasing complexity too much, however the simple approach reproduces the electrostatic interactions crudely. The final problem in the modelling of electrostatics is how these charges fall off with distance. This is a particular problem in the *in vacuo* modelling of polyanionic species, such as DNA, that give rise to large repulsive forces. This problem is reduced by the use of explicit solvent modelling which will be discussed later.

2.4.4 Bonding Terms

The bonding terms are made up of bond stretching, bending and torsion angles. The force constant and equilibrium value for each of these is associated with an atom type. Each atom in the system can be assigned to a particular atom type. This generalisation reduces the number of parameters and facilitates a rapid parameterisation, since atom

types describe easily identifiable chemical features such as 'Hydrogen attached to an sp^3 carbon'. Bond stretching is modelled by means of a simple harmonic potential (Equation 6).

$$E_{bond} = \sum k_{eq}(l - l_{eq})^2$$

Equation 6: Bonding energy approximated by a simple harmonic potential, where k_{eq} is the energy constant for a bond, l is the current bond length and l_{eq} is the equilibrium bond length.

This function gives a good description of energy around the equilibrium position, however the function does not tail off at extended bond lengths, so is poor at calculating dissociation energies. Poor quality structures, with longer bond distances, have much higher energies but this results in structures being more rapidly restored to equilibrium during a minimisation. Bond angles are treated in a similar way to bond stretching and are modelled by use of a simple harmonic potential (Equation 7).

$$E_{angle} = \sum k_{\theta}(\theta - \theta_{eq})^2$$

Equation 7: Bond angle approximation also using a simple harmonic potential, where energy constant for the bond is k_{θ}, θ is the actual bond angle and θ_{eq} is the equilibrium bond angle.

The final bonding term is the dihedral angle. The energy term for this is based on an additive series of cosine functions and reproduces the rotational preferences for staggered versus eclipsed conformations, and the preference for cis- trans- isomerism (Equation 8).

$$E_{DH} = \sum V_n(1 + S\cos n\omega)$$

Equation 8: Dihedral angle energy where V_n is the barrier height, S controls the phase and is either 1 or –1, n is the periodicity of rotation and ω is the torsion angle. Multiple functions for the same angle can be summed to allow a more accurate description of the torsion angle.

This term does not represent the full rotational barrier and can thus be misleading. Non-bonding interactions play a significant role in determining this energy. Thus, the 1-4 van der Waals and electrostatic interactions must also be evaluated to determine an overall energy for the rotation about a dihedral angle.

2.4.5 Variations Between Force Fields

The above terms cover the principal functions used in molecular modelling. Some other terms are also used but these vary on an implementation basis. Perhaps the most common additional term is a specific term for hydrogen bonding. This was essential in biological modelling *in vacuo*, but the advent of explicit solvent modelling has reduced the need for artificially high forces to maintain the structural integrity. Current versions of the AMBER forcefield incorporate this hydrogen-bonding term into the electrostatic interactions.

2.4.6 Molecular Mechanics

Obtaining a parameterised molecule and a set of equations is the initial step, but this information can be utilised to generate structures of key importance. The concept of structure for a large biological molecule is difficult to define. There is probably no one conformation that generates an overall minimum structure. Although some structures will obviously be of unfavourably high energy there are undoubtedly a number of low energy conformations all of which are accessible at physiological conditions. This concept makes determination of structures more challenging since we are required to talk about sets of conformations rather than an individual structure. Two molecular mechanics methods allow us to probe aspects of structure.

2.4.7 Energy Minimisation

Molecular minimisation aims to adjust molecular geometries to ideally find the global energy minimum of the structure. This is achieved by determining the first and in some cases the second derivatives of the potential function with respect to co-ordinates. The problem exists of finding an energy minimum and determining if this is the overall minimum for the molecule. For large systems the latter question is much more difficult to answer, since the only way of determining the global minimum is to evaluate the energy of all available conformations. This is an impossible task with such a large number of possible structures. The presence of a single energy minimum also suggests a single structure. This is not the case, and there are probably many minima of comparable low energy. The realistic role of energy minimisation is therefore to remove high energy conformations.

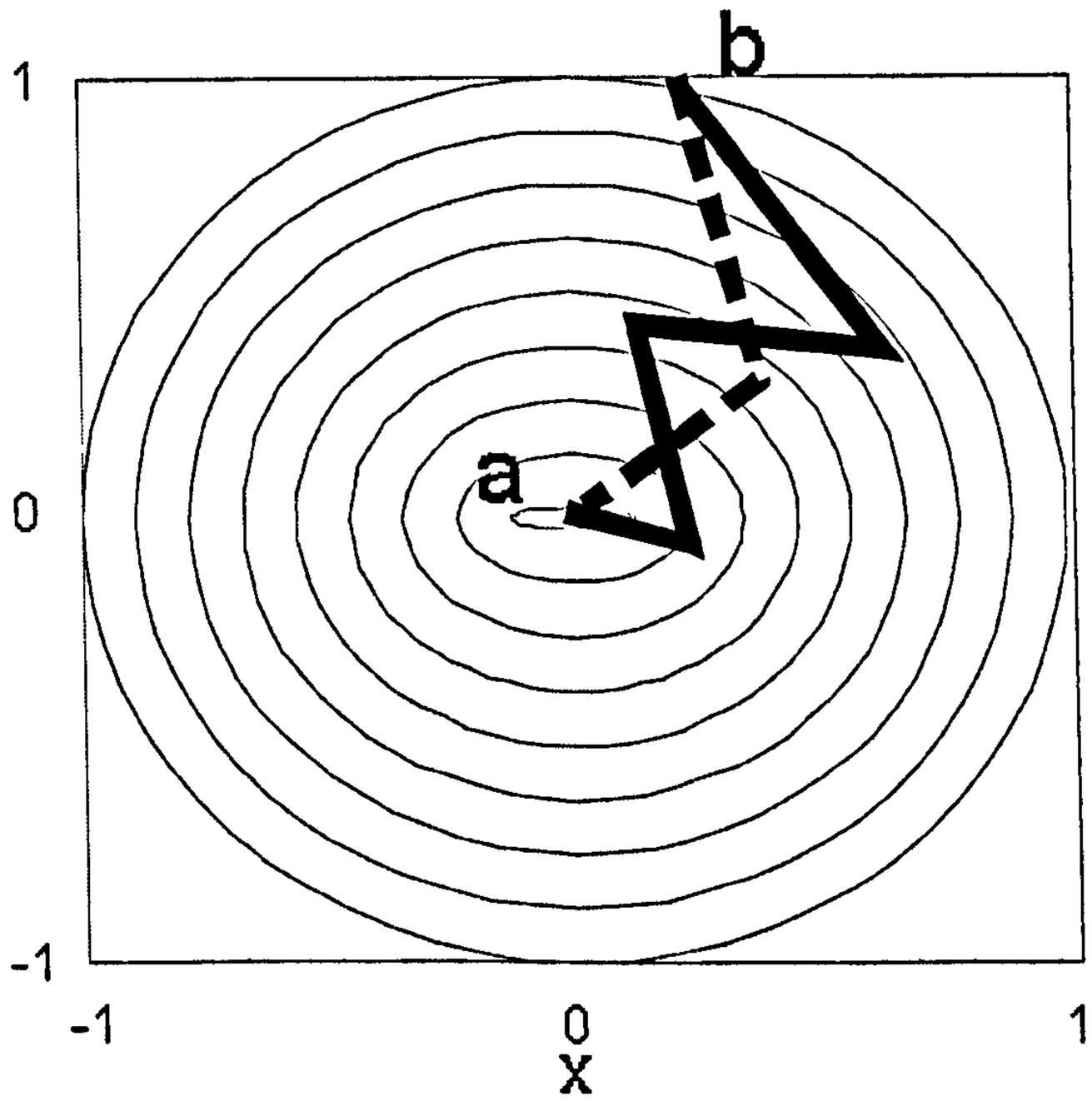


Figure 2-11: Energy minimisation from a structure (b) to a minimum (a). The route taken by steepest decent takes steps towards a but overshoot is common (solid line). Conjugate gradient follows a similar principle but is moderated by past steps and locates the minimum quicker (dashed line)

Two common methods are in use, and both of these only calculate the first derivative of the energy function. Both methods iteratively determine co-ordinates that lead to a decrease in the energy of the system. For both, success is defined as concurrent structures of similar energy, with any change only serving to increase the energy of the system, which is consistent with being located in a minimum. The steepest descent method moves the molecule towards the part of the energy surface that has the steepest downward gradient. The method suffers from overshoot as the base of the energy well is approached (Figure 2-11). This can lead to difficulties in missing local minima especially if the energy well is shallow. The conjugate gradient method is an enhancement of the steepest descent, algorithm by considering previous steps towards

the minima. Such an approach dampens the tendency for overshoot. It is common to employ both of these methods to achieve a minimisation of starting structures where high energies from incorrect geometries and bond lengths might be present. A limited number of cycles of steepest descent followed by conjugate gradient minimisation gives rapid convergence to the local minima.

2.4.8 Molecular Dynamics

Molecular dynamics attempts to determine conformation by considering the energy available to a molecule. To be able to move from one low energy structure to another, of perhaps lower energy, potential energy barriers must be surmounted. If the energy available to a system in the form of thermal energy is considered, there may be sufficient energy available to overcome the various transitional barriers (Figure 2-12).

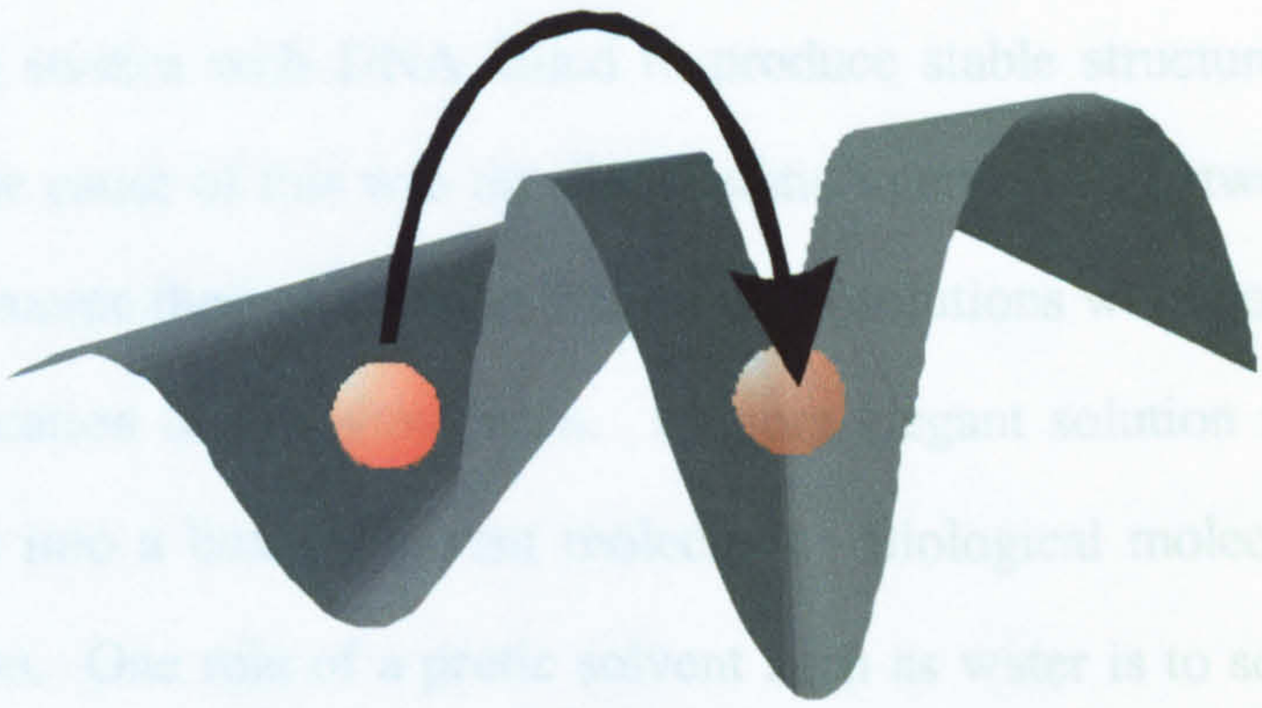


Figure 2-12: Potential energy surface for a molecule. A structure in a trough on the surface must overcome an energy barrier if it is to be able to access other areas of the potential energy surface.

Molecular dynamics operates by considering atom velocities as well as co-ordinates. Initial velocities are assigned based on a random distribution derived from the initial temperature. The system is then allowed to evolve by simulating the energies and allowing motion. For the simulation to function correctly the resolution must be in the order of the fastest motion, this means that energies and velocities are re-calculated typically after one femtosecond of molecular motion. The dynamic behaviour of a molecule means that unless there is a particularly steep energy well, many minima can be sampled. This has important considerations when considering the conformation of a molecule, since we are forced to consider ensembles of structures and averaged values for key structural parameters.

2.4.9 Explicit Solvent Modelling

Early modelling studies with DNA failed to produce stable structures over long MD simulations. The cause of this was the electrostatic interactions between the phosphate groups. To overcome these difficulties a number of solutions were postulated including artificial modification of the phosphates. A more elegant solution is to immerse the solute molecule into a box of solvent molecules. Biological molecules are normally found in solution. One role of a protic solvent such as water is to solvate and 'shield' polarised sites on a molecule. Solvent molecules are normally added in a pre-equilibrated box to surround the solute, along with counter ions sufficient to cause neutrality. The use of such a system introduces a number of new difficulties. There is an obvious increase in the complexity of the system, however a common simplification in the treatment of solvent molecules is to make all bonding interactions constant, and only evaluate the non-bonding interactions for solvent molecules. This reduces the number of energy evaluations required.

The use of a box of solvent would lead to artificial forces being present at the boundary. The use of periodic boundary conditions in modelling neatly avoids this problem. Molecules that cross the boundary are relocated to the opposite sides, and the same for non-bonding interactions. Such a system creates an infinite array of "unit cells". The use of a suitably set cut-off, above which distance non-bonding interactions are not evaluated, has been shown to prevent imaging effects where a molecule would interact with itself across the periodic boundary (Figure 2-13).

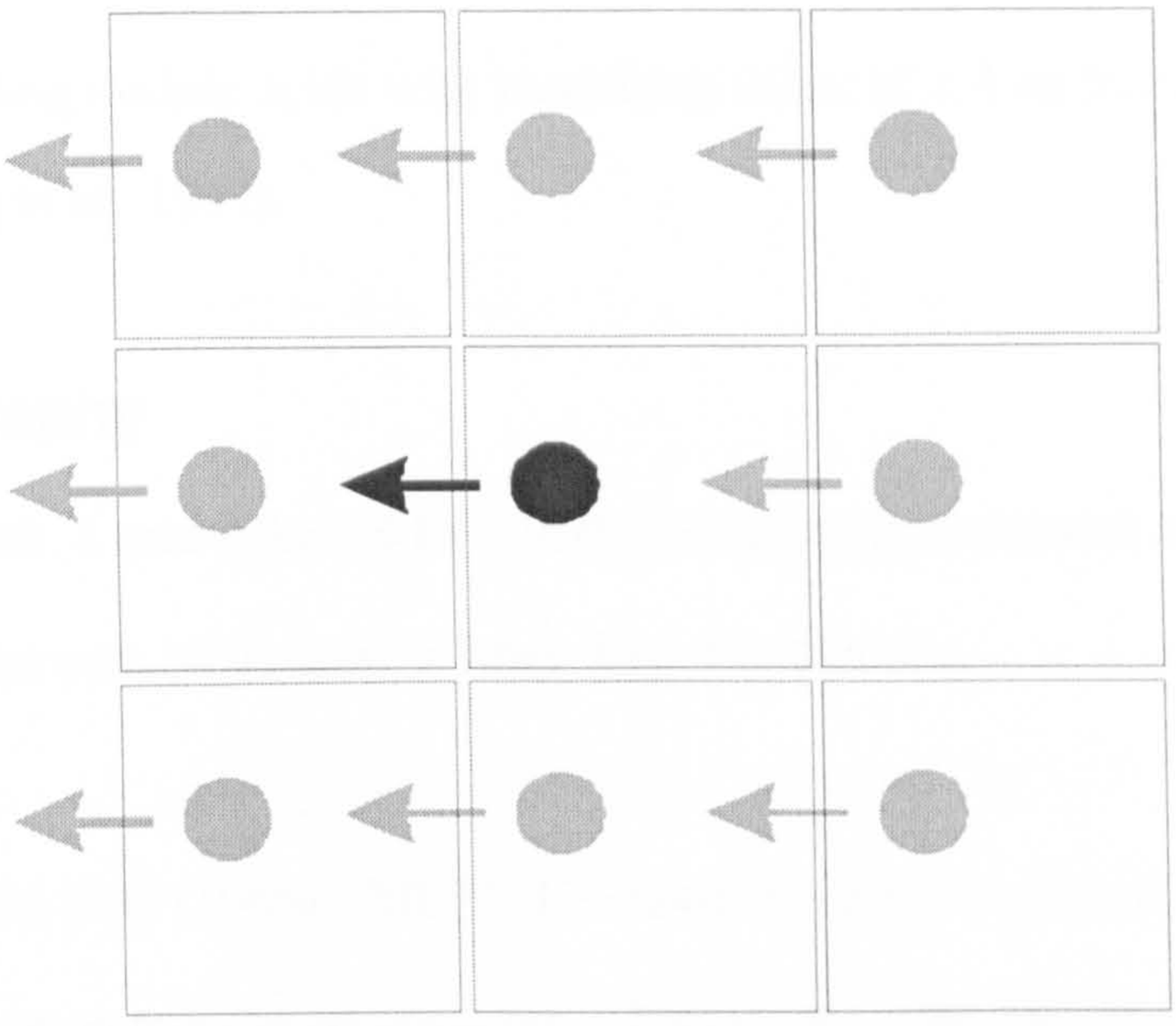


Figure 2-13: Periodic boundary conditions. Translation of a molecule out of a unit cell is prevented by wrapping the co-ordinates so the molecule is relocated to the other side of the unit cell. This creates an 'infinite' array of cells removing any possible effects caused by the boundary.

The final modification required to achieve long term stability in the modelling of biomolecules considers the problems inherent with electrostatics. Since these are long range in nature the effect of the cut-off is to truncate the electrostatic forces, and this leads to instability (Lee et al. 1995). The use of the particle mesh Ewald (PME) function to simulate electrostatics overcomes this difficulty (Darden et al. 1993; York et al. 1993; York and Yang, 1994). This method calculates the charge at set points in space and allows for the periodic boundary conditions. Although initial modelling appeared to function correctly without the use of PME, long simulations of about 100 ps of molecular time showed instabilities due to these truncations (York et al. 1995). With the use of these methods it is possible to achieve high stability of biological

molecules including nucleic acids with modelling times of 1-5 ns being reported in the literature (Young et al. 1997).

2.5. Bibliography

Baleja, J.D., Moult, J. and Sykes, B.D. (1990) Distance measurement and structural refinement with NOE data. *J. Mag. Res.* **87**, 375-384.

Bax, A. and Davis, D.G (1985a) MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy. *J. Mag. Res.* **65**, 355-360.

Bax, A. and Davis, D.G (1985b) Practical aspects of two-dimensional transverse NOE spectroscopy. *J. Mag. Res.* **63**, 207-213.

Bax, A. and Freeman, R. (1981) Investigation of Complex Networks of Spin-Spin Coupling by Two-Dimensional NMR. *J. Mag. Res.* **44**, 542-561.

Bax, A., Griffey, R.H. and Hawkins, B.L. (1983) Correlation of proton and N-15 chemical-shifts by multiple quantum NMR. *J. Mag. Res.* **55**, 301-315.

Bax, A. and Summers, M.F. (1986) ^{1}H and ^{13}C Assignments from Sensitivity-Enhanced Detection of Heteronuclear Multiple-Bond Conectivity by 2D Multiple Quantum NMR. *J. Am. Chem. Soc.* **108**, 2093-2094.

Bayly, C.I., Cieplak, P., Cornell, W.D. and Kollman, P.A. (1993) A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints For Determining Atom-Centered Charges: The RESP model. *J. Phys. Chem.* **97**, 10269

Behling, R.W., Rao, S.N., Kollman, P.K. and Kearns, D.R. (1987) Molecular Mechanics and Dynamics Calculations on $(dA)_{10}$ $(dT)_{10}$ Incorporating Distance Constraints Derived from NMR Relaxation Measurements. *Biochem.* **26**, 4674-4681.

Bertini, H., Clemente, A., Rombeck, I., Rosato, A., Turano, P., Lippert, B. and Quadrifoglio, F. (1999) Three-dimensional solution structures of two DNA dodecamers through full relaxation matrix analysis. *Mag. Resn. Chem.* **37**, 564-572.

Boltan, P.H. and James, T.L. (1979) Dynamics in DNA Molecular Motions in RNA and DNA Investigated by Phosphorous-31 and Carbon-13 NMR Relaxation. *J. Phys. Chem.* **83**, 3359-3366.

Brandan, A. and Thomas, L.J. (1989) Two Dimensional Nuclear Overhauser Effect: Complete Relaxation Matrix Analysis. *Meth. Enzym.* **176**, 169-183.

Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M.Jr., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W. and Kollman, P.A. (1995) A 2nd Generation Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules. *J. Am. Chem. Soc.* **117**, 5179-5197.

Darden, T.A., York, D.M. and Pedersen, L.G. (1993) Particle Mesh Ewald - An N.LOG(N) Method for Ewald Sums In Large Systems. *J. Chem. Phys.* **98**, 10089-10092.

Drew, H.R., Wing, R.M., Takano, T., Broka, C., Tanaka, S., Itakura, K. and Dickerson, R.E. (1981) Structure of a B-DNA Dodecamer - Conformation and Dynamics .1. *Proc. Nat. Acad. Sci. USA* **78**, 2179-2183.

Grant, G.H. and Richards, W.G. (1995) *Computational Chemistry*, Oxford Science Publications.

Griffey, R.H., Redfield, A.G., Loomis, R.E. and Dahquist, F.W. (1985) Nuclear Magnetic Resonance Observation and Dynamics of Specific Amide Protons in T4 Lysozyme. *Biochem.* **24** , 817-823.

Hunter, C.A. (1993) Sequence-dependent DNA Structure: The Role of Base Stacking Interactions. *J. Mol. Biol.* **230**, 1025-1054.

Jacobson, B. and Anderson, W.A. (1954) Proton Magnetic Resonance Study of the Hydration of Deoxyribonucleic Acid. *Nature* **173**, 772-773.

Jeener, J. (1971) Ampere International Summer School, Basko Polje, Yugoslavia.

Lee, H., Darden, T.A. and Pedersen, L.G. (1995) Accurate crystal-molecular dynamics simulations using particle mesh Ewald - RNA dinucleotides ApU and GpC. *J. Chem. Phys.* **243**, 229-235.

Otting, G., Orbons, L.P.M. and Wüthrich, K. (1990) Suppression of Zero-Quantum Coherence in NOESY and soft NOESY. *J. Mag. Res.* **89**, 423-430.

Piantini, U., Sorensen, O.W. and Ernst, R.R. (1982) Multiple quantum filters for elucidating NMR coupling networks. *J. Am. Chem. Soc.* **104**, 6800-6801.

Piotto, M., Saudek, V. and Sklenar, V. (1992) Gradient-tailored excitation for single quantum NMR spectroscopy of aqueous solutions. *J. Bio. Mol. NMR* **2**, 661-665.

Rance, M., Sorensen, O.W., Bodenhausen, G., Wangner, G., Ernst, R.R. and Wüthrich, K. (1983) Improved spectral resolution in COSY H-1-NMR spectra of proteins via double quantum filtering. *Biochem. Biophys. Res. Comm.* **117**, 479-485.

Schleucher, J., Quant, J., Glaser, S.J. and Griesinger, C. (1995) A Theorem Relating Cross-Relaxation and Hartmann-Hahn Transfer In Multiple-Pulse Sequences - Optimal Suppression of TOCSY Transfer in ROESY. *J. Mag. Resn. A* **112**, 144-151.

Schwabe, J.W.R. (1997) The role of water in protein-DNA interactions. *Cur. Op. Struct. Biol.* **7**, 126-134.

Searle, M.S. (1993) NMR studies of Drug-DNA interactions. *Progress in Nuclear Magnetic Resonance* **25**, 403-480.

Yang, X.L. and Wang, A.H.J. (1999) Structural studies of atom-specific anticancer drugs acting on DNA. *Pharmacology and Therapeutics* **83**, 181-215.

York, D.M., Darden, T.A. and Pedersen, L.G. (1993) The Effect of Long-Range Electrostatic Interactions in Simulations of Macromolecular Crystals - a Comparison of the Ewald and Truncated List Methods. *J. Chem. Phys.* **99**, 8345-8348.

York, D.M. and Yang, W.T. (1994) The Fast Fourier-Poisson Method for Calculating Ewald Sums. *J. Chem. Phys.* **101**, 3298-3300.

York, D.M., Yang, W.T., Lee, H., Darden, T.A. and Pedersen, L.G. (1995) Toward the Accurate Modeling of DNA - The Importance of Long-Range Electrostatics. *J. Am. Chem. Soc.* **117**, 5001-5002.

Young, M.A., Ravishanker, G. and Beveridge, D.L. (1997) A 5-nanosecond molecular dynamics trajectory for B-DNA: Analysis of structure, motions, and solvation. *Biophys. J.* **73**, 2313-2336.

Chapter 3. Structure of Nogalamycin Bound at a High Affinity TpG Site

3.1. Introduction

The anthracycline antibiotics interact with double-stranded DNA by intercalating an aglycone ring system between the DNA base pairs. As described in Chapter 1, these compounds are amongst some of the most widely used and effective chemotherapeutic agents available. The molecular basis for their mode of action lies in their ability to inhibit transcription factor binding, or interfere with topoisomerase activity, largely through the formation of a highly stable drug-DNA complex (Fisher and Aristoff , 1988). Nogalamycin (Figure 3-1) has a unique structure that permits the antibiotic to thread through the DNA helix and interact with both the major and minor grooves simultaneously (Searle et al. 1988). The antibiotic effectively clamps the two strands of DNA together, with interactions in both grooves imparting a high binding affinity but also slow binding kinetics. This requires major structural disruption of the DNA for both association and dissociation events to occur (Fox et al. 1985). These factors have been linked to both its potent biological activity (Bhuyan and Dietz, 1965; Bhuyan and Reusser, 1970; Li et al. 1979; Ennis, 1981), and also its higher level of cytotoxicity than observed for other members of the anthracycline family (Fox et al. 1985).

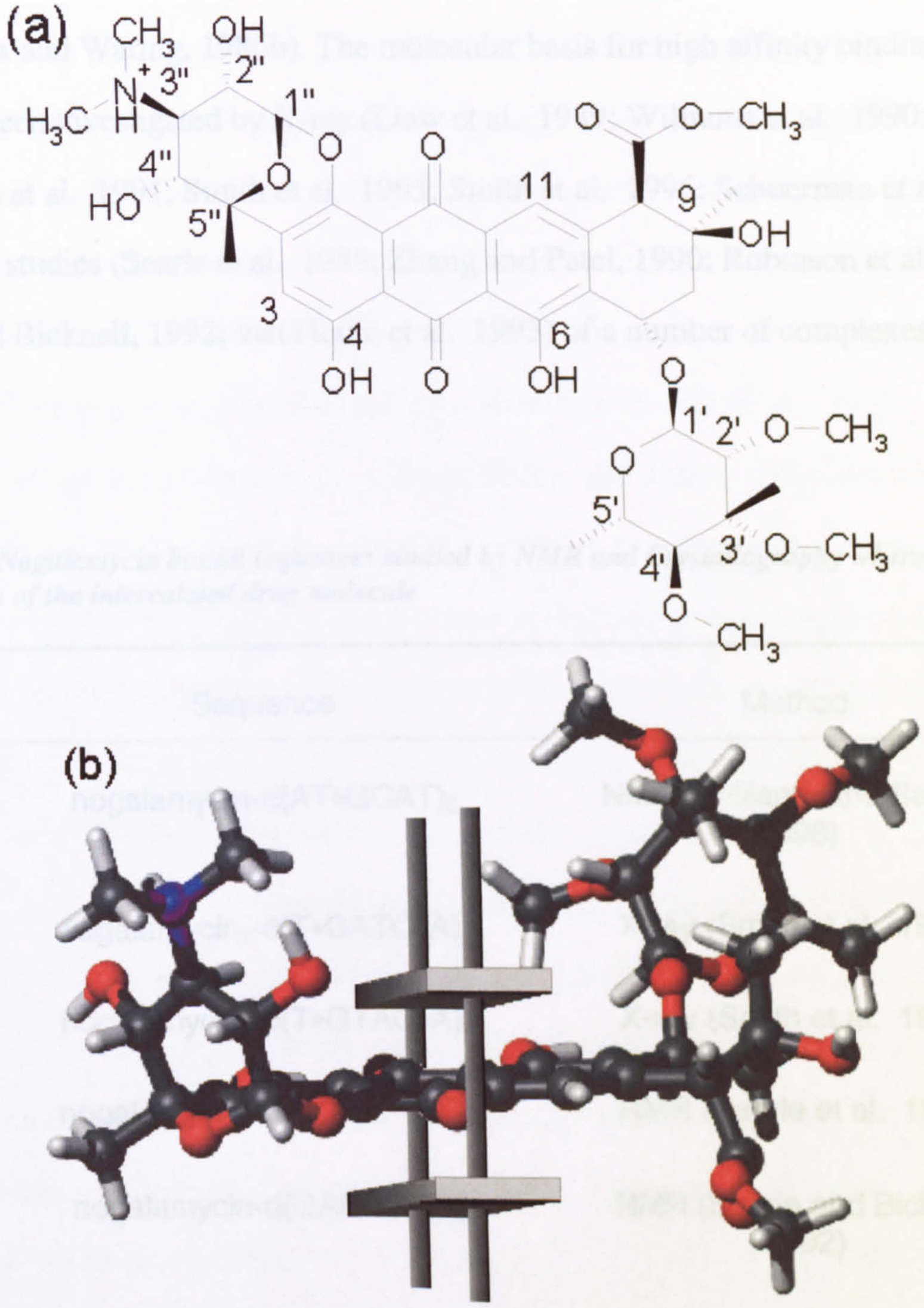


Figure 3-1: Structure of nogalamycin showing the atomic numbering scheme (a). The 'U' shaped structure of the antibiotic permits intercalation with the sugars on the A and D rings able to form interaction with both grooves of DNA simultaneously (b).

Footprinting analysis shows nogalamycin to have a sequence selectivity for the 5'-pyrimidine-purine steps 5'-TpG (and equivalent 5'-CpA) and 5'-CpG (Fox and Waring,

1986a; Fox and Waring, 1986b). The molecular basis for high affinity binding to these sites has been investigated by X-ray (Liaw et al. 1989; Williams et al. 1990; Gao et al. 1990; Egli et al. 1991; Smith et al. 1995; Smith et al. 1996; Schuerman et al. 1996) and NMR studies (Searle et al. 1988; Zhang and Patel, 1990; Robinson et al. 1990; Searle and Bicknell, 1992; van Houte et al. 1993) of a number of complexes (Table 3-1)

Table 3-1: Nogalamycin bound sequences studied by NMR and Crystallography where • represents the position of the intercalated drug molecule

Sequence	Method
nogalamycin-d(AT•GCAT)$_2$	NMR (Williams and Searle, 1998)
nogalamycin$_2$-d(T•GATC•A)$_2$	X-ray (Smith et al. 1995)
nogalamycin$_2$-d(T•GTAC•A)$_2$	X-ray (Smith et al. 1996)
nogalamycin$_2$-d(GC•AT•GC)$_2$	NMR (Searle et al. 1988)
nogalamycin-d(GAC•GTC)$_2$	NMR (Searle and Bicknell, 1992)
nogalamycin$_2$-d(AGC•AT•GCT)	NMR (van Houte et al. 1993)
nogalamycin-d(GC•GT) d(AC•GC)	NMR (Zhang and Patel, 1990)
nogalamycin$_2$-d(C•GTAC•G)$_2$	NMR (Robinson et al. 1990)
nogalamycin$_2$-d(^{5Me}C•GT(pS)A^{5Me}C•G)$_2$	X-ray (Egli et al. 1991)

In the majority of structures reported to date, antibiotic molecules are bound at the terminal intercalation sites (TpG and CpG), leading to some ambiguity regarding the orientational preference of the antibiotic at these sites. The nogalose sugar in these complexes lies in the minor groove rather than overhanging the end of the duplex where it might interfere with crystal packing (Figure 3-2a). It has been suggested that this binding orientation is largely dictated by end-effects (Smith et al. 1995). In contrast, the NMR structures (Searle et al. 1988; Zhang and Patel, 1990) accommodate two bound drug molecules in close proximity such as to suggest that steric interactions between nogalamycin molecules dictate their bound orientation (Figure 3-2b). Smith et al. (1995) concluded that in longer sequences of DNA, with intercalation sites situated further apart, the alternative (180° related) orientation may be observed for the drug bound at the 5'-TpG site. To address the question of the orientational preference of the antibiotic, and probe its molecular basis, this chapter describes NMR studies of the nogalamycin complex with the hexamer duplex $d(ATGCAT)_2$, which contains adjacent 5'-TpG and 5'-CpA high affinity sites. Titration of drug with this DNA sequence clearly indicates a preference for the formation of the 1:1 complex. This is confirmed by the pattern of through space NOEs that characterise the binding location and orientation with some accuracy.

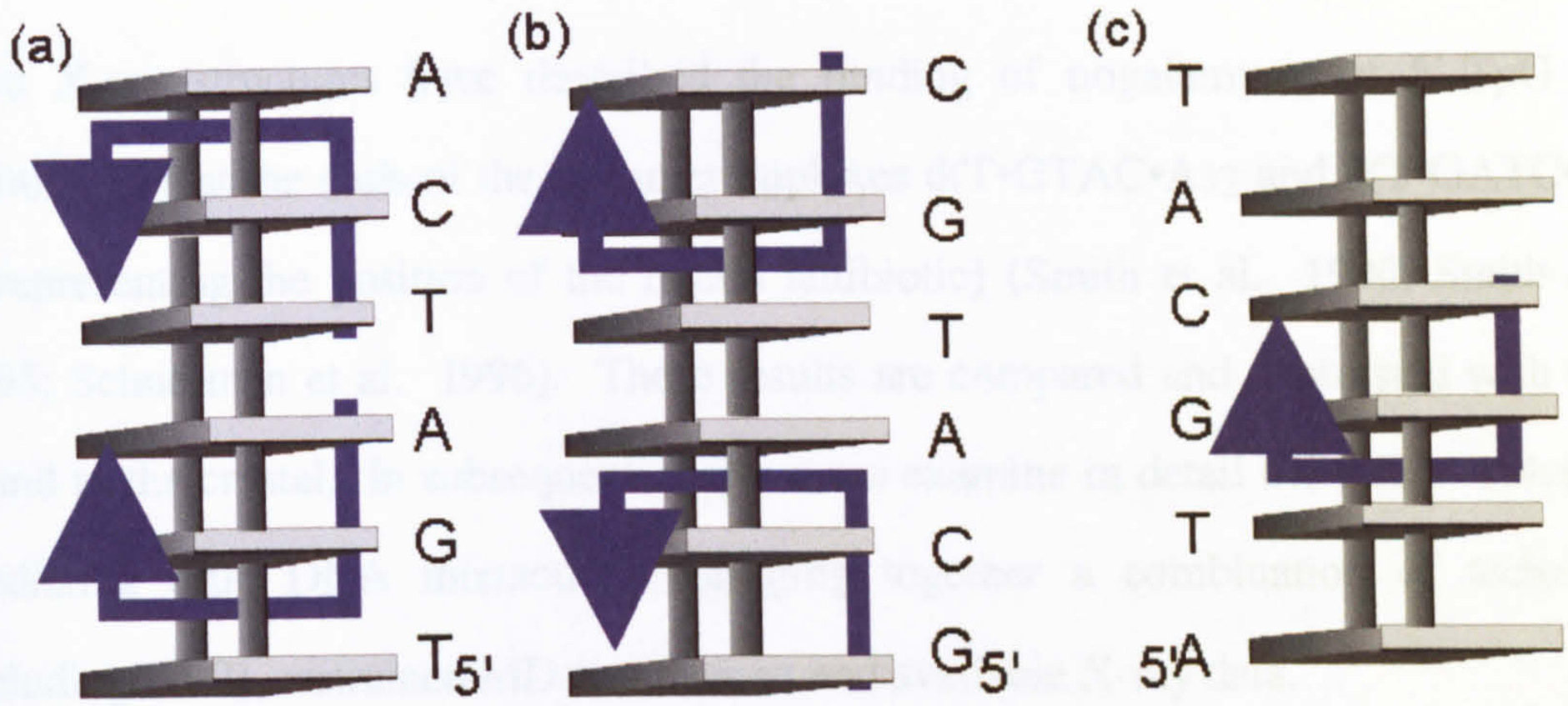


Figure 3-2: Schematics of nogalamycin complexes. Sequence studied by X-ray crystallography (Smith et al. 1995) (a) has a potential bias due to the terminal binding positions of the two drug molecules, which may prevent alternative orientations. NMR sequence (Searle et al. 1988) (b) overcomes difficulties with terminal binding sites, but the close proximity of drug molecules may prevent other configurations being adopted. The proposed structure of the sequence used in this study (c) is free from all such effects and is unhindered in its binding orientation.

We have used NOE-restrained molecular dynamics (rMD) simulations employing the AMBER 94 forcefield (Pearlman et al. 1995) to determine the structure and dynamics of the nogalamycin-d(ATGCAT)$_2$ complex over a ~ 1 ns simulation. We have used an explicit solvation model in which long range electrostatic interactions have been treated using the particle mesh Ewald summation method (York et al. 1993; York et al. 1995; Darden et al. 1993; Lee et al. 1995; Cheatham and Kollman, 1996) This has already been shown to lead to stable trajectories over long simulation times (> 1 ns) while reproducing many sequence specific DNA structural features for a variety of sequences (Cheatham et al. 1995; Cheatham and Kollman, 1996; Young et al. 1997; Duan et al. 1997; Duan et al. 1997). Time-dependent analysis of the dynamics simulation has provided insights into the fluctuations of key structural parameters not apparent from the static X-ray models.

Two X-ray structures have described the binding of nogalamycin at 5'-TpG high affinity sites at the ends of the hexamer duplexes d(T•GTAC•A)$_2$ and d(T•GATC•A)$_2$. [• representing the position of the bound antibiotic] (Smith et al. 1996; Smith et al. 1995; Schuerman et al. 1996). These results are compared and contrasted with those found in the crystal. In subsequent chapters we examine in detail the role of solvent in mediating drug DNA interactions, bringing together a combination of techniques including NMR, restrained MD simulations and available X-ray data.

3.2. NMR Studies of d(ATGCAT)$_2$ and its Complex with Nogalamycin

3.2.1 NMR Studies of the Hexamer Duplex d(ATGCAT)$_2$

The hexamer duplex was fully assigned using a combination of through space and through bond interactions using NOESY and TOCSY data sets (Figure 3-3). Duplex formation under the conditions used (100 mM NaCl, 10 mM phosphate buffer at pH 7.0) was confirmed.

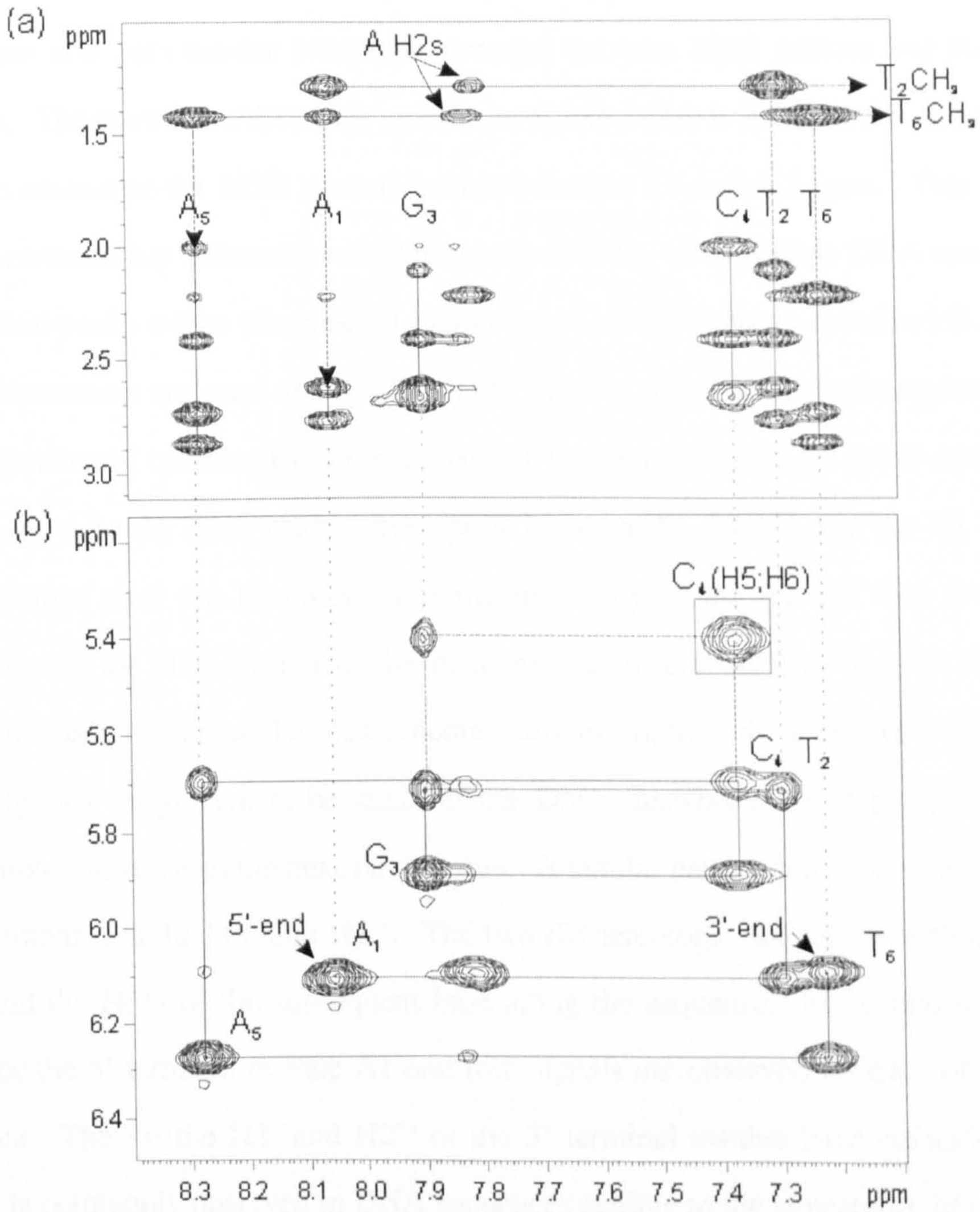


Figure 3-3: NOESY spectrum of d(ATGCAT)$_2$ recorded at 288K pH 7.0 showing the H6/8→H2';H2'' assignment (a) and H6/8→H1' "backbone" pathway with only the inter-residue cross-peaks labelled (b). The boxed peak is confirmed by TOCSY to be the J-coupled cytosine H5→H6. Spectra are calibrated relative to the methyl signal of trimethylsilylpropanoate.

Sequential assignment of DNA base pairs was achieved via standard sequential pathways (Chazin et al. 1986). At mixing times of 300ms for NOESY experiments,

both inter and intra residue NOEs are revealed between H6/8 protons and H1' sugar protons. The chemical shift of H1' proton in B-like DNA is between 5.0 and 6.5 ppm and the resonance for H6/8 protons occurs between 7.0 and 8.5 ppm. This spectral region contains key information regarding the identity of important DNA resonances. Additional peaks within this spectral region are observed between cytosine H6 and H5. These resonances are good markers being the only J coupled peaks in the region and aid in the location of cytosine H6 protons. As can be seen in Figure 3-3 the 5'-end residue A1 shows only the intra-residue cross peak between the A1 H8 and the A1 H1'. A second cross peak can be observed at the frequency of the A1 H1' that reveals the frequency of the H6/8 proton of the next residue, in this instance the T2 H6. This sequential assignment can be seen schematically in Figure 3-4. This method allows an unambiguous assignment to be made of the DNA 'backbone' by stepping from base H6/8 proton to sugar to the next base proton. A similar pattern can be seen between the base protons and the H2' and H2''. The two diastereotopic protons give NOEs to the base and the H6/8 of the subsequent base along the sequence. Thus, two signals are seen for the 5' terminal residue A1 and four signals are observed for each of the other residues. The T6 the H2' and H2'' of the 3' terminal residue have coincident shifts which is commonly observed in DNA sequences leading to the appearance of only three peaks for the terminal residue. Further indications to residue assignments can be obtained from the T2 and T6 CH_3 to preceding H8 protons of A1 and A5, respectively.

Determination of the remaining drug resonances confirm the assignment of the $d(ATGCAT)_2$ sequence which was consistent with a right-handed B-like structure.

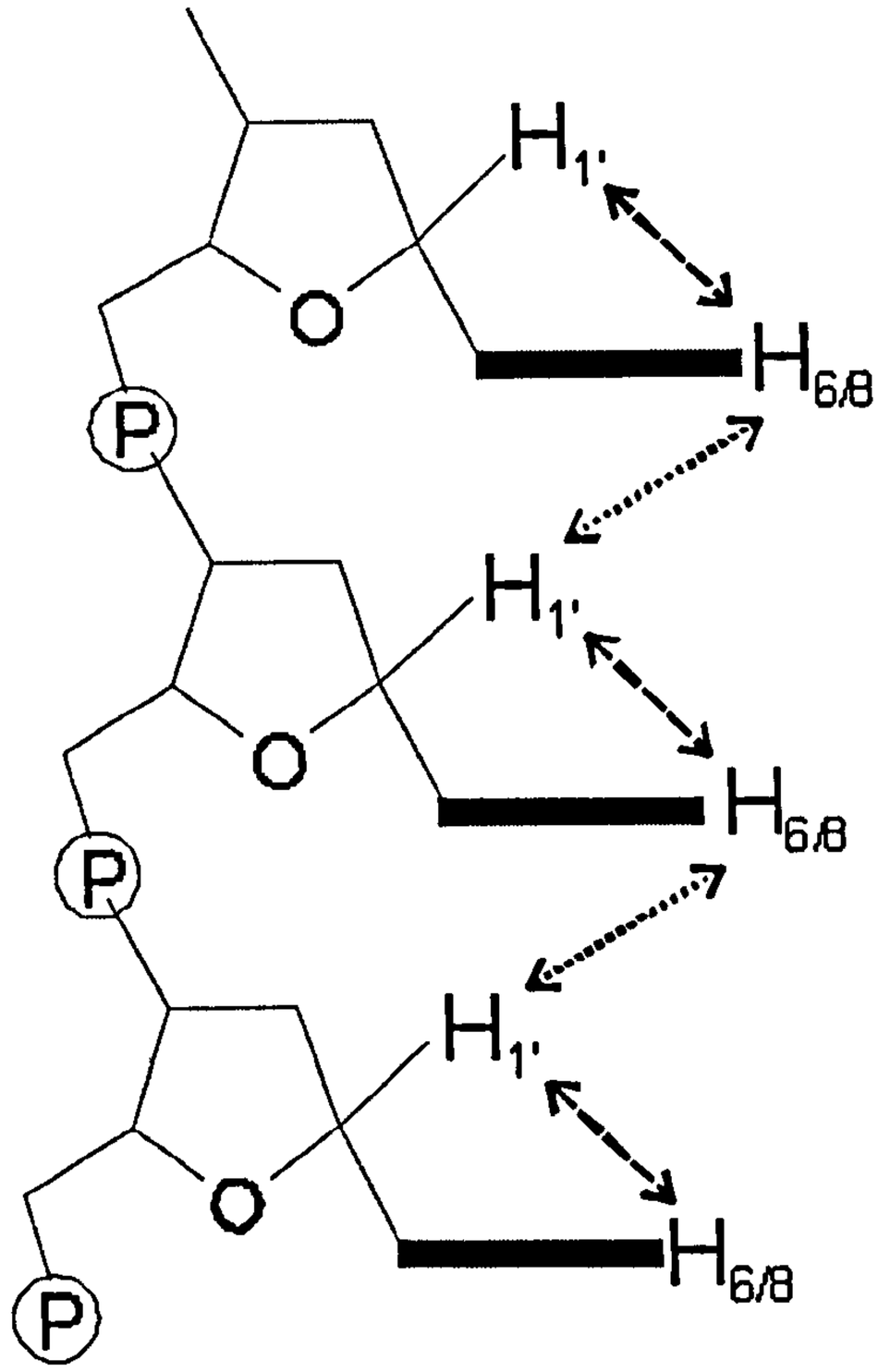


Figure 3-4: Schematic showing sequential assignment methodology between H6/8 base protons and H1' sugar protons in B-DNA. Double headed arrows represent NOEs expected to be observed between protons.

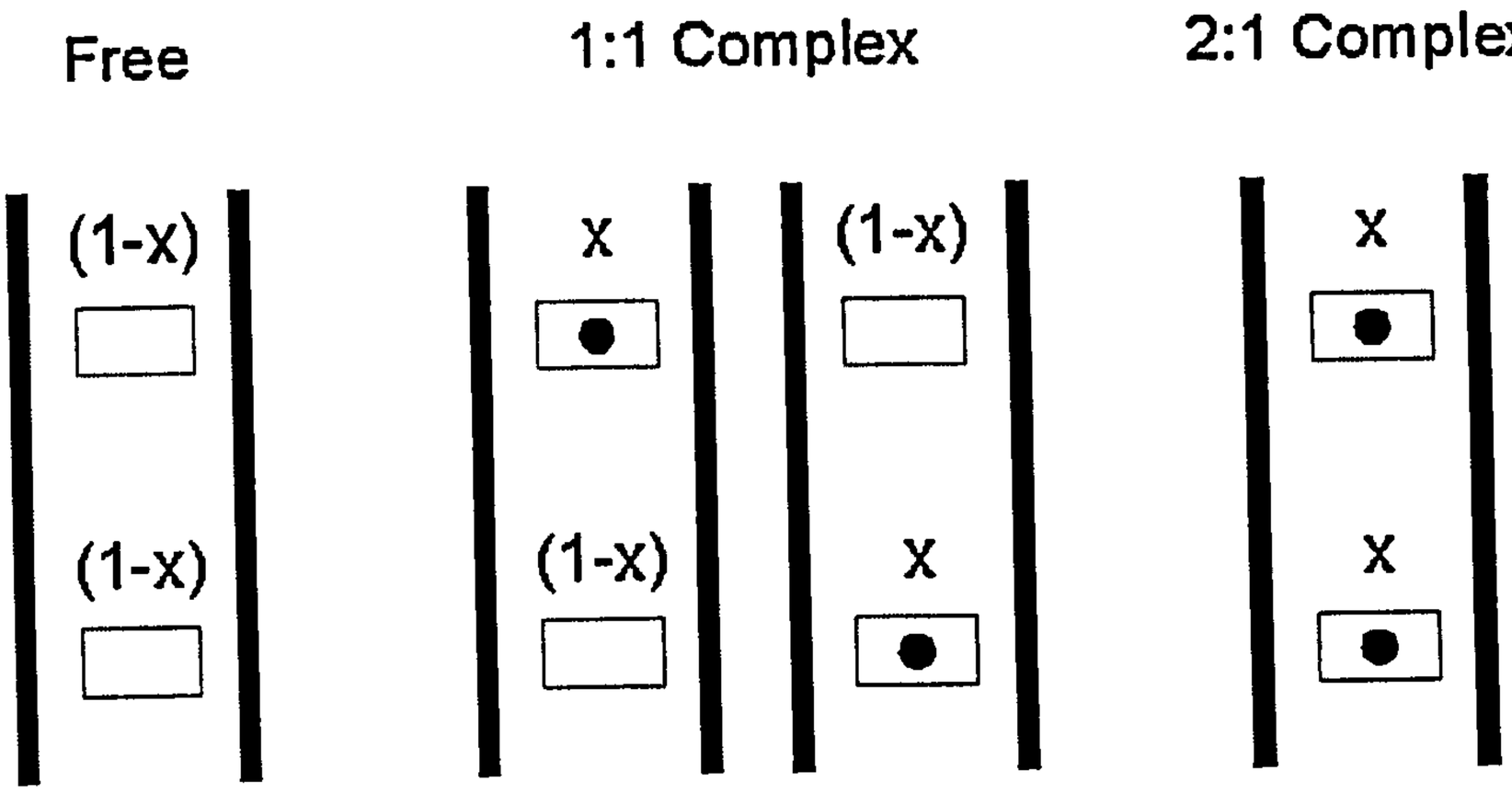
3.2.2 The 1:1 Binding Stoichiometry for the Nogalamycin-d(ATGCAT)$_2$ Complex

As was discussed previously on page 69, the deoxyribonucleotide sequence d(ATGCAT)$_2$ possesses two high affinity binding sites (TpG and CpA). Binding at only one of these sites would remove the symmetry of the nucleotide, however, if both sites become occupied the symmetry would be restored. The complexity of the spectra alone can therefore aid in the determination of the binding mode that is present. A simple model of binding allows a prediction of the relative proportions of free DNA

versus 1:1 and 2:1 drug:duplex species. Assuming an equal probability of binding to either site, and that free uncomplexed drug is unlikely to be observed at drug:duplex ratios less than 1:1, the partitioning of drug between each possible complex can be calculated (Figure 3-5), where x represents the mole fraction of drug to duplex binding sites.

This simple model suggests that if all binding sites are equal, at a 1:1 ratio (x=0.5) there would be equal proportion of free to 2:1 complex. Titration of the antibiotic into a 3 mM solution of $d(ATGCAT)_2$ (100 mM NaCl, 10 mM phosphate buffer at pH 7.0) leads to the lifting of the dyad symmetry of the hexamer duplex with each signal in the ^1H NMR spectrum of the free duplex being cleanly replaced by two corresponding resonances for the complex. The end-point of the titration was the formation of the 1:1 complex with a single antibiotic bound per duplex. The complex formed is highly stable over many months in an NMR tube.

At the 1:1 binding ratio, there is no evidence for other bound species or free DNA. The observable DNA complexes must therefore be associated with only a single bound drug molecule. This suggests a strong anti-co-operative behaviour with binding at one site preventing accommodation of the other drug molecule.



Total Binding:

$1 - 2x + x^2$ $\quad 2(x-x^2)$ $\quad x^2$

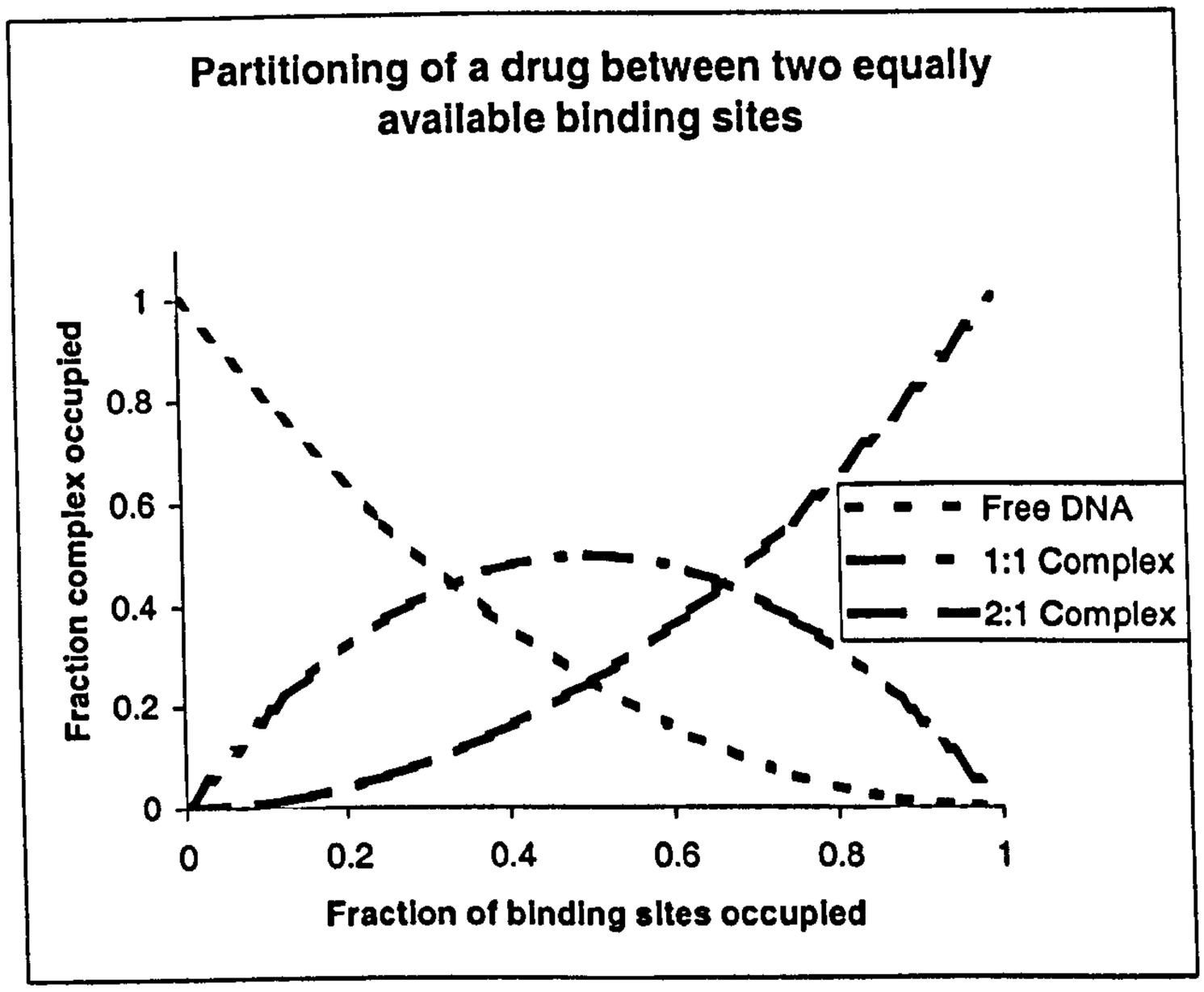


Figure 3-5: Theoretical partitioning of a drug molecule between two equally probable binding sites. The possible binding modes are shown, top, with occupied binding sites being represented with a filled circle. The partition is calculated where x represents the mole fraction of drug per duplex binding site. From the graph, the free duplex is seen to be gradually replaced by 1:1 and 2:1 complex. At an equimolar ratio of drug to duplex (x=0.5) the free DNA and 2:1 complex are of an equal intensity with the 1:1 complex concentration being a maximum.

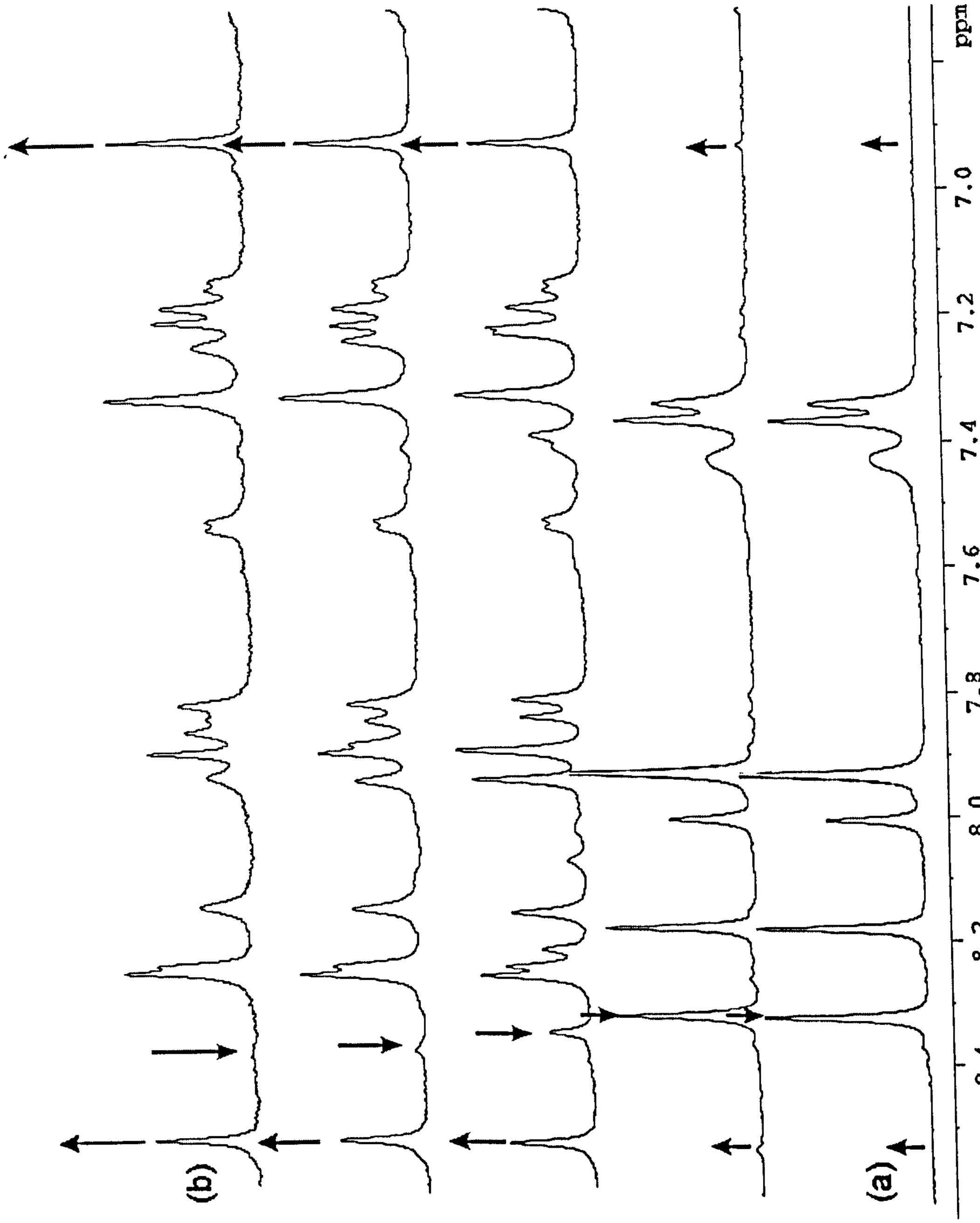


Figure 3-6: Titration of d(ATGCAT)$_2$ with nogalamycin recorded at 298K. Increasing concentration of nogalamycin versus duplex from 0:1 (a) to 1:1 (b). Arrows highlight a number of clearly defined signals due to free DNA that decrease and increasing signals from the complex. No evidence is seen of other species in solution.

3.2.3 2D NMR analysis and Chemical Shift Assignments in the Nogalamycin-d(ATGCAT)$_2$ Complex

Several regions of the 500 MHz NOESY spectrum of the complex (mixing time 250 ms, 288K) are highlighted in Figure 3-7. Resonance assignments were carried out using well established procedures (see section 3.2.1), and as outlined for NMR studies of several other nogalamycin-DNA complexes (Zhang and Patel, 1990; Searle and Bicknell, 1992) using a combination of TOCSY and NOESY data sets. DNA and antibiotic assignments are presented in Table A4-2 and Table 3-3. The sequential assignment pathway along the two DNA strands is highlighted using deoxyribose H1' (5.4–6.6 ppm) to base H6/H8 (6.5–8.5 ppm) correlations to identify neighbouring nucleotides, and to localise the drug-intercalation site (Figure 3-7).

The pattern of NOE connectivities along strand 1–6 is interrupted at the T2H1'→G3H8 step where this interaction is too weak to detect, implying a significant increase in inter-proton separation compared with standard B-DNA. This is consistent with DNA unwinding and intercalation at this site. In contrast, NOE connectivities are contiguous along strand 7–12. It has previously been reported (Searle and Bicknell, 1992; Williams and Searle, 1998) that base pair buckling at both TpG and CpG intercalation sites can lead to a high degree of asymmetry, which significantly distorts the pattern of NOEs along one strand but not the other.

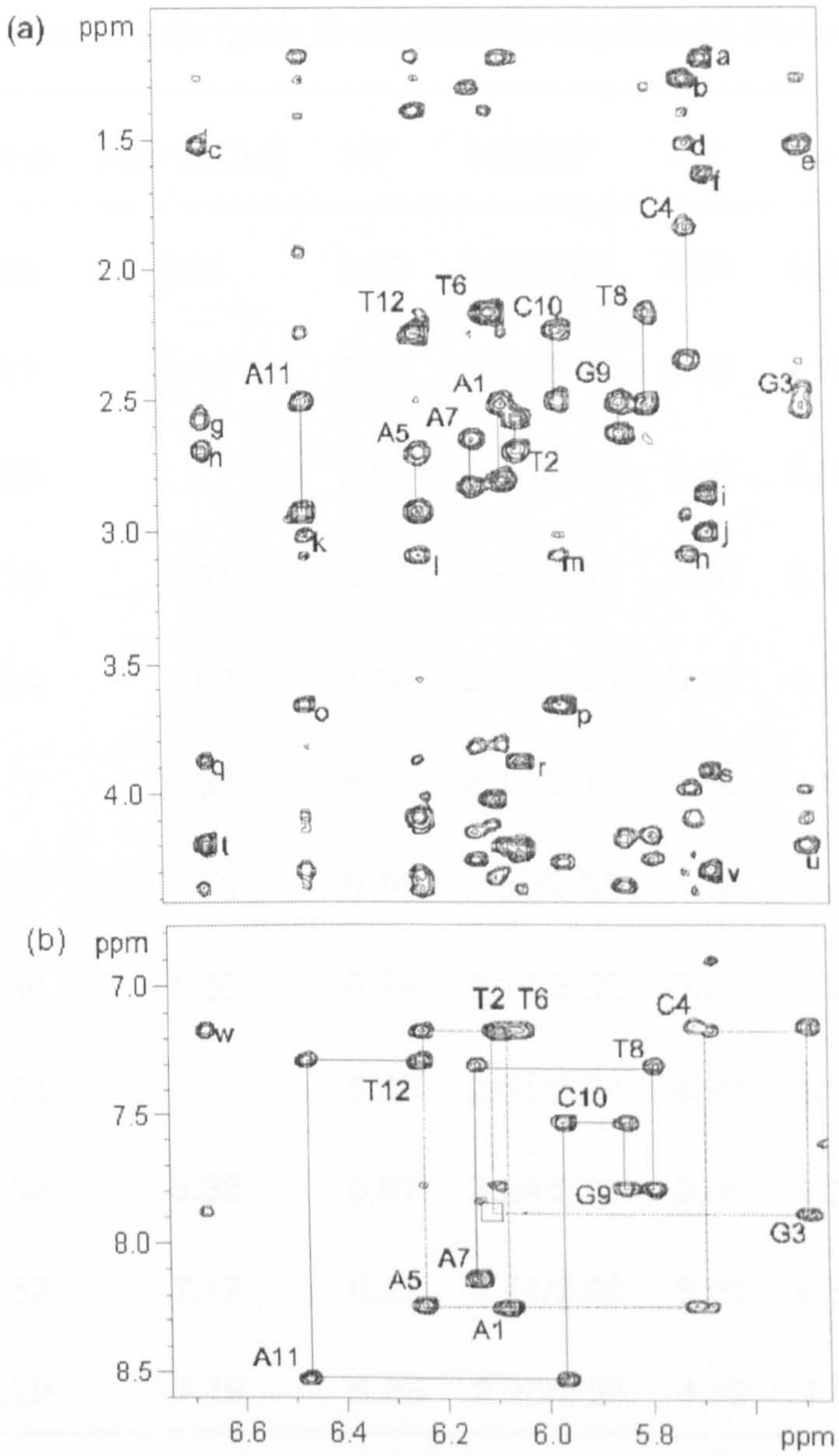


Figure 3-7: Expanded portions of the 250 ms NOESY spectrum of the 1:1 nogalamycin:d(ATGCAT)$_2$ complex at 288K. In (a) the H1' → H2'/H1'/H2" pairs are identified for the 12 unique deoxyribose spin systems labelled A1 through to T12. (b) Sequential assignment pathway highlighting H1' → H6/H8 connectivity which are contiguous along each strand except at the T2–G3 step corresponding to the drug intercalation site. Intranucleotide connectivities are labelled. A number of additional NOEs of structural significance (a)–(w) are highlighted as follows: (a) T2 CH_3↔H1", (b) C4 H1'↔3'-CH_3, (c) 9-CH_3↔H11, (d) C4 H1'↔9-CH_3, (e) G3 H1'↔9-CH_3, (f) 5"-CH_3↔H1", (g) T2 H2"↔H11, (h) T2 H2'↔H11, (i) NCH_3↔H1", (j) NCH_3↔H1", (k) A11 H1'↔H8, (l) A5 H1'↔3'-OCH_3, (m) C10 H1'↔3'-OCH_3, (n) C4 H1'↔3'-OCH_3, (o) A11 H1'↔2'-OCH_3, (p) C10 H1'↔2'-OCH_3, (q)10-OCH_3↔H11, (r) T2 H1'↔10-OCH_3, (s) H4"↔H1", (t) H10↔H11, (u) G3 H1'↔H10, (v) H3"↔H1", (w) T2 H6↔H11.

Table 3-2: Assigned chemical shifts (ppm) for d(ATGCAT)$_2$-nogalamycin DNA protons.[a]

	H6/H8	H2/H5/CH_3	H1'	H2'/H2"	H3'	H4'	N1/N3
A1	8.25	6.91	6.08	2.52/2.84	4.80	4.20	-
T2	7.17	1.19	6.06	2.58/2.70	5.02	4.22	13.72
G3	7.88	-	5.49	2.47/2.49	5.00	4.55	13.07
C4	7.16	5.01	5.72	1.84/2.36	4.76	3.97	-
A5	8.24	7.83	6.24	2.71/2.93	5.04	4.37	-
T6	7.17	1.39	6.11	2.17/2.17	4.56	4.01	13.35
A7	8.13	7.78	6.14	2.66/2.83	5.03	4.25	-
T8	7.30	1.30	5.79	2.17/2.52	5.03	4.24	13.72
G9	7.78	-	5.85	2.51/2.64	4.99	4.35	13.10
C10	7.52	5.38	5.97	2.24/2.51	5.05	4.26	-
A11	8.52	7.17	6.47	2.51/2.90	5.03	4.56	-
T12	7.29	1.19	6.25	2.25/2.25	4.62	4.08	13.35

a. All data referenced to internal trimethylsilylpropionate at pH 7.0, 288K, 100 mM sodium chloride, 10 mM sodium phosphate buffer.

Table 3-3: Assigned chemical shifts (ppm) for nogalamycin in the d(ATGCAT)$_2$-nogalamycin complex.[a]

Proton	δ	Proton	δ	Proton	δ	Proton	δ
H3	6.92	4-OH	11.65	6-OH	10.67	H7	4.76
H8	3.02/1.93	9-CH_3	1.53	H10	4.14	10-OCH_3	3.86
H11	6.67	H1'	5.21	H2'	3.65	2'-OCH_3	3.65
3'-CH_3	1.27	3'-OCH_3	3.09	H4'	3.16	4'-OCH_3	3.56
H5'	3.83	5'-CH_3	1.41	H1"	5.68	H2"	4.29
H3"	4.76	$N(CH_3)_2$	3.00/2.86	H4"	3.90	5"-CH_3	1.65

a. All data referenced to internal trimethylsilylpropionate at pH 7.0, 288K, 100 mM sodium chloride, 10 mM sodium phosphate buffer.

There are asymmetric perturbations to the chemical shift of DNA residues when drug binding occurs (Table 3-4). As expected drug association causes the most pronounced shift changes in the binding region, however DNA protons within this area are affected differently suggesting an asymmetry in binding at the intercalation site. The base H6 protons of pyrimidines and H8 protons of purines, located in the major groove, show this effect markedly. With A11H8 and C10H6 being 'down-field' shifted from the free complex. This is in contrast to an 'up-field' shift for T2H6 and no change for the G3H8. This seems consistent with the drug aglycone stacking preferentially with the T2 and G3 bases. The deoxyribose sugar protons all show similar trends in their chemical shift pattern with the data for H1' protons being the best resolved. The largest changes in value of chemical shift for these minor groove oriented protons are localised to the base pairs on either side of the intercalation site. T2H1', C10H1' and A11H1'

are shifted downfield while the G3H1' moves up-field relative to the corresponding chemical shift in free DNA. With the exception of the terminal T12H1' no other proton experiences a significant shift. The observed chemical shift changes are consistent with an intercalation at the 5'dTpdG site with the structural changes being introduced remote to the intercalation site.

Table 3-4: Change in chemical shift (ppm) for H6/8 and H1' protons upon drug binding. Negative values represent a down field shift upon binding.

H6/8					
A1	**T2**	**G3**	**C4**	**A5**	**T6**
-0.21	0.12	0.00	0.21	0.03	0.05
T12	**A11**	**C10**	**G9**	**T8**	**A7**
-0.07	-0.25	-0.15	0.10	-0.01	-0.09

H1'					
A1	**T2**	**G3**	**C4**	**A5**	**T6**
0.02	-0.35	0.39	-0.02	0.03	-0.02
T12	**A11**	**C10**	**G9**	**T8**	**A7**
-0.16	-0.20	-0.27	0.03	-0.08	-0.04

3.2.4 Intermolecular NOEs Define Drug TpG Intercalation Site Orientation

An essentially complete assignment of drug and DNA resonances provides the starting point for identifying intermolecular NOEs that define the position and orientation of the bound antibiotic. The antibiotic aglycone H11 provides an unambiguous indicator of the position and orientation of the drug. NOEs are identified to deoxyribose protons T2H1' and G3H1' and can only be satisfied if the drug is intercalated at the 5'-TpG step such that the nogalose sugar points towards the centre of the duplex (Figure 3-8). NOEs from the exchangeable 6-OH of the drug to C10H2" and A11H1' are similarly consistent with this orientation. Many other intermolecular NOEs support this

conclusion; for example, the methoxy and methyl groups on the nogalose and ring A sugar provide evidence for hydrophobic interactions in the minor groove and determine with some precision the position and orientation of the nogalose with respect to the intercalation site. As shown in Figure 3-8, NOEs are detected from 3'-OCH_3 to A5H1', C4H1' and C10H1', from 2'-OCH_3 to C10H1' and A11H1', and from OCH_3 of the C10-methylester of ring A to T2H1'. Additional NOEs from 9-CH_3 to G3H1', and from 3'-CH_3 to C4H1' are also identified unambiguously. The pattern of interactions shows that the nogalose and ring A sugars are extended along the minor groove, forming an extensive array of hydrophobic contacts encompassing the four base sequence $T^2{\bullet}G^3C^4A^5$ and those of the complementary strand. By comparison, the number of contacts in the major groove is rather limited; NOEs from 3"-CH_3 and H1" on the bicyclic sugar to T2CH_3 are consistent with the orientation of the antibiotic suggested by the pattern of NOEs in the minor groove.

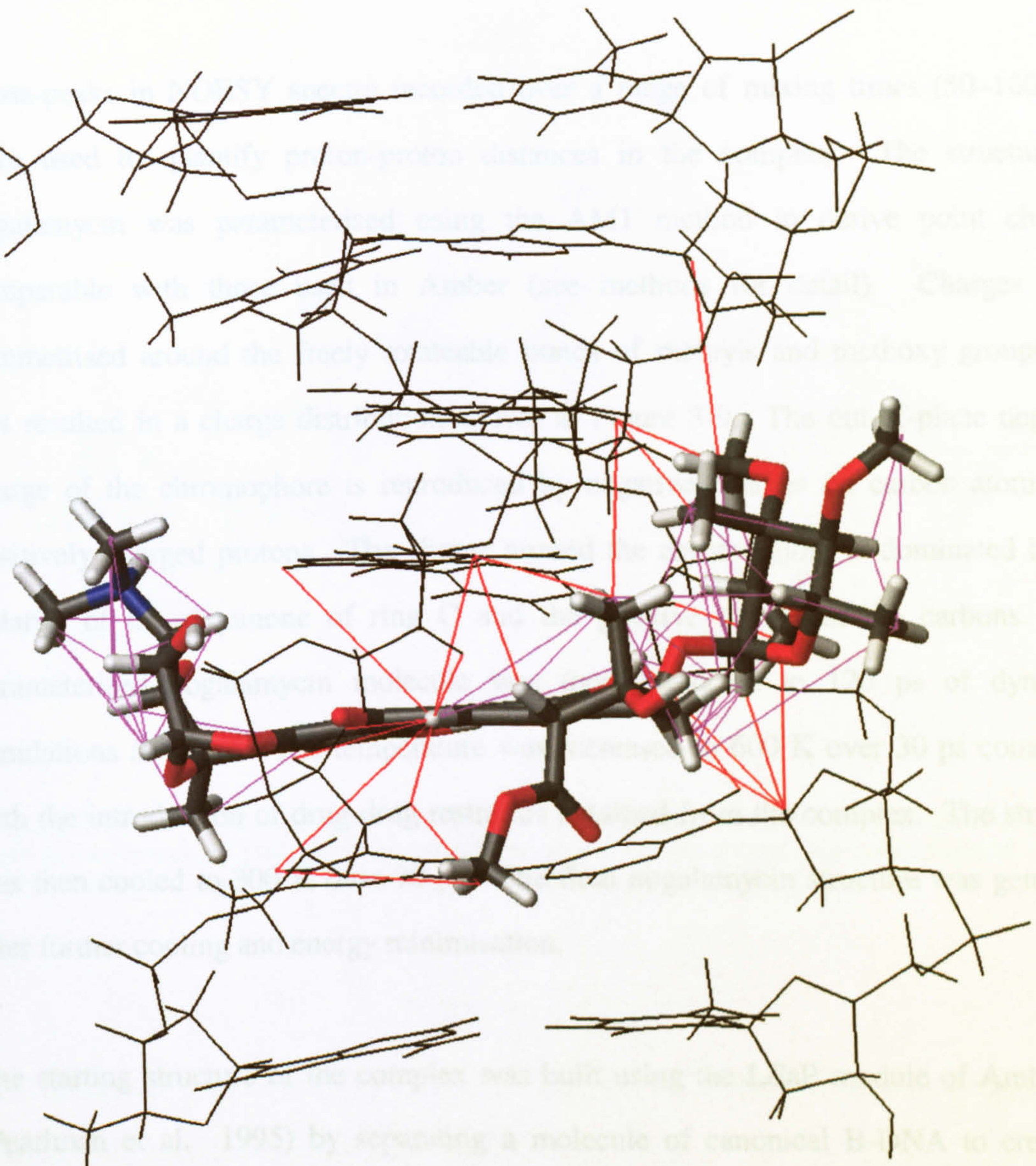


Figure 3-8: Graphical representation of NOE contacts. Nogalamycin is intercalated with the major groove located on the left. DNA is represented in black. Intra-drug restraints are shown in purple and drug-DNA restraints are shown in red.

3.2.5 Quantitative studies of Intermolecular NOEs and Structure Calculations

Cross-peaks in NOESY spectra recorded over a range of mixing times (50–100 ms) were used to quantify proton-proton distances in the complex. The structure of nogalamycin was parameterised using the AM1 method to derive point charges comparable with those used in Amber (see methods for detail). Charges were symmetrised around the freely rotateable bonds of methyls and methoxy groups and this resulted in a charge distribution shown in Figure 3-9. The out-of-plane negative charge of the chromophore is reproduced by negative charges on carbon atoms and positively charged protons. The charge around the chromophore is dominated by the polarity of the diquinone of ring C and the positive charge of the carbons. The parameterised nogalamycin molecule was then subjected to 120 ps of dynamics simulations *in vacuo*. The temperature was increased to 600 K over 30 ps coinciding with the introduction of drug-drug restraints obtained from the complex. The structure was then cooled to 300 K over 70 ps. The final nogalamycin structure was generated after further cooling and energy minimisation.

The starting structure of the complex was built using the LEaP module of Amber 4.1 (Pearlman et al. 1995) by separating a molecule of canonical B-DNA to create an approximate 10Å gap to accommodate the intercalator. The structure was energy minimised to optimise bond lengths. The final complex was modelled using an explicit solvation protocol (as described in methods) (Cheatham and Kollman, 1996; Duan et al. 1997). All dynamics simulations were ran using the particle mesh Ewald method within Amber to accurately model long-range electrostatic interactions. A simulation of ~ 1 ns at constant temperature with NOE restraints was sampled every 0.5 ps to analyse the conformational dynamics and hydration during the course of the trajectory.

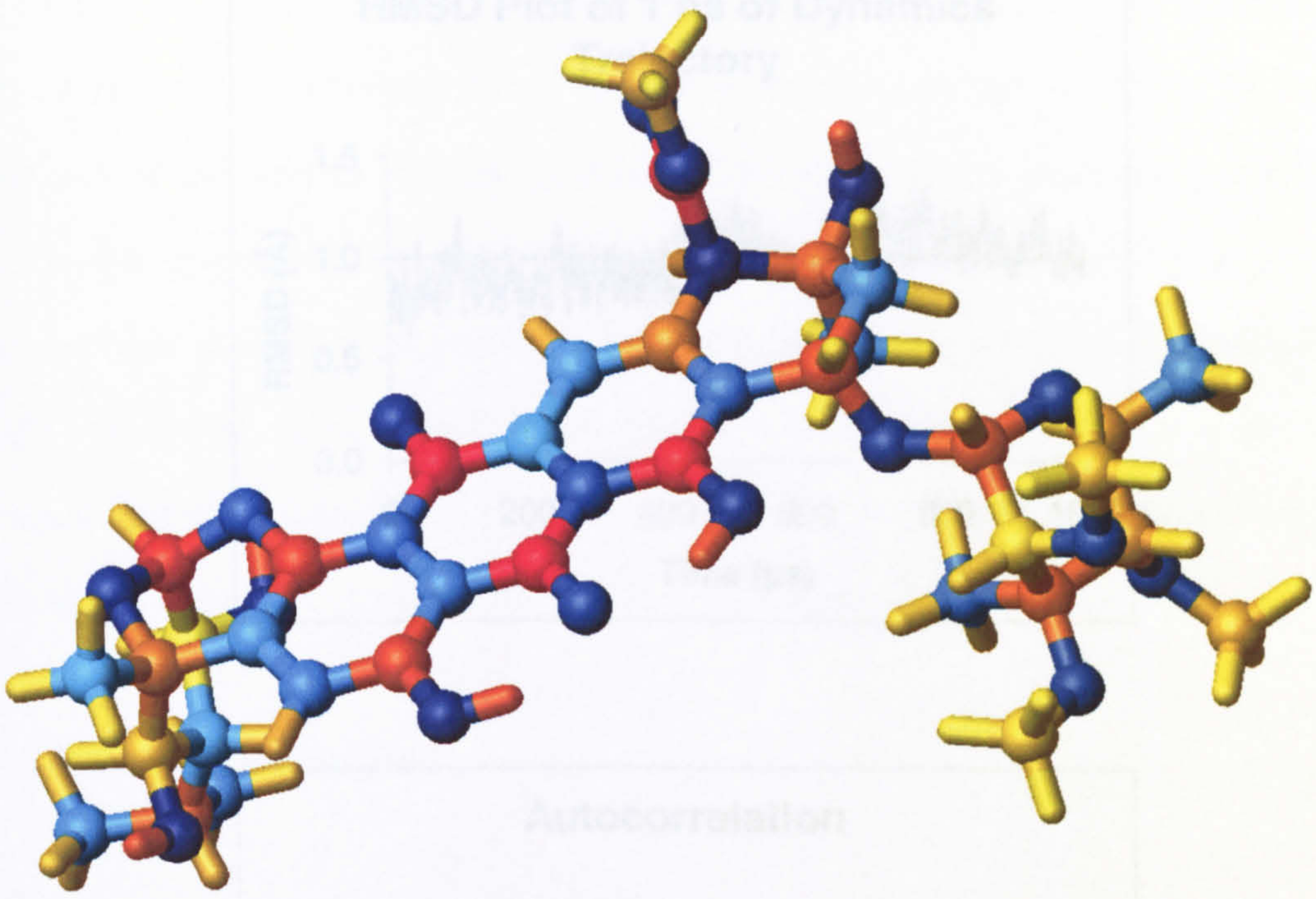
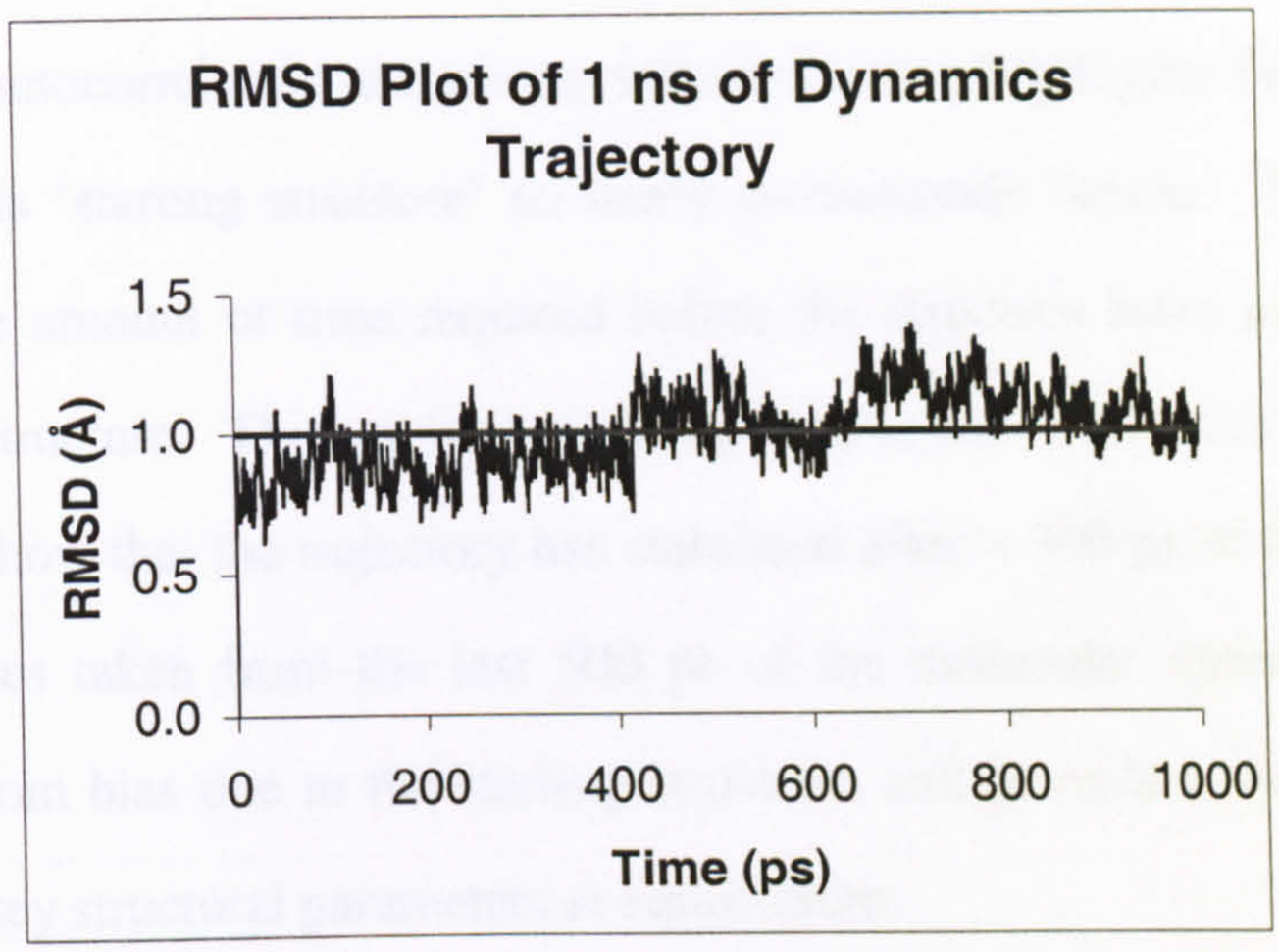


Figure 3-9: Calculated charge distribution on nogalamycin. Positive partial charges are represented by yellow to red (most positive) shading and negative charges are represented by light blue to dark blue (most negative) shading.

The complex is well defined by the NOE data and the molecular dynamics refinement protocol (for restraint violations see methods).

The sequence-dependent variation in key structural parameters has been analysed using CURVES (Lavery and Sklenar, 1996) over the final 500 ps of dynamics. A measure of convergence can be obtained by considering the autocorrelation. The autocorrelation method gives a measure of structural similarity, in the form of an averaged RMSD between all structures separated by a given number of picoseconds in a molecular dynamics trajectory.



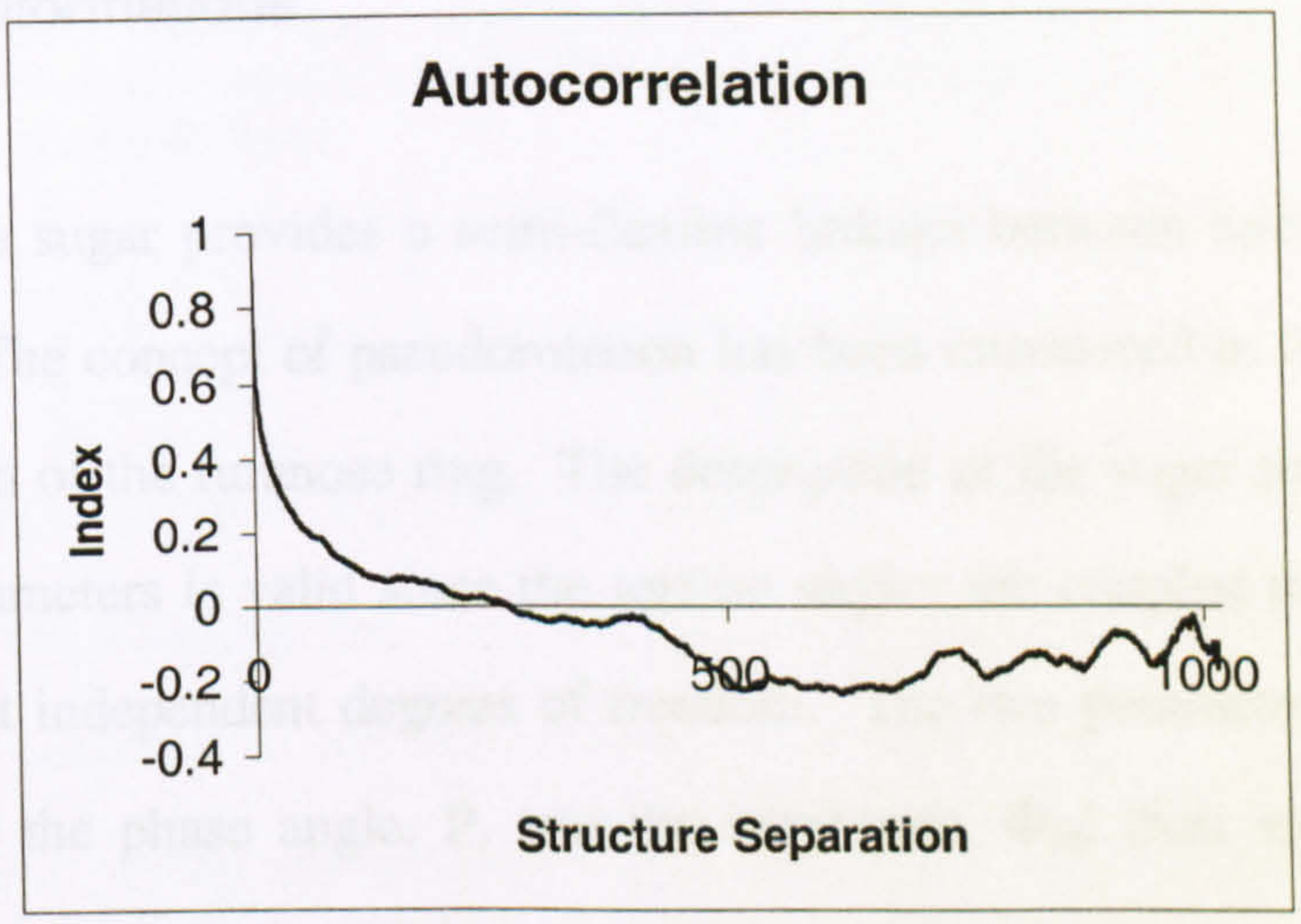


Figure 3-10: (top) RMSD of structures from the initial structure in the 1 ns dynamics calculation. (bottom) Autocorrelation plot of last 1ns of restrained molecular dynamics of drug DNA complex. Index plots correlation of structures that are separated by 0-1000ps.

With small separations, the structures would be expected to be similar since insufficient time has elapsed to cause significant change in the conformation of the molecule. Conversely as the structure separation increased, hence, the similarity would decrease.

The plot of the autocorrelation at each separation distance highlights the dependence of a structure on its 'starting structure' so many picoseconds before. This provides an indication of the amount of time required before the structure loses any potential bias due to starting structure. This analysis when applied to the d(ATGCAT)$_2$-nogalamycin dynamics data show that the trajectory has stabilised after ~ 300 ps of dynamics (Figure 3-10). Structures taken from the last 500 ps of the molecular dynamics run can be assumed free from bias due to the starting structure, and provide a dynamic picture of fluctuations in key structural parameters at equilibrium.

3.2.6 Sugar conformations

The deoxyribose sugar provides a semi-flexible linkage between nucleotide base and the backbone. The concept of pseudorotation has been introduced in the description of the conformation of the furanose ring. The description of the sugar ring torsion angles by just two parameters is valid since the torsion angles are coupled to each other, and therefore are not independent degrees of freedom. The two parameters that depict the ring pucker are the phase angle, P, and the amplitude, Φ_m; their relationship to the deoxyribose torsion angles is well known (Altona and Sundaralingam, 1972). Although an infinite number of ring conformations are available, two stable ring conformations have been identified. The North or N conformation possesses a pseudorotational phase angle (P) of 0° and adopts a C3'-endo C2'-exo arrangement. The South or S conformation is 180° on the pseudorotation cycle and is C3'-exo C2'-endo (Figure 3-11). Studies of nucleic acids reveal clustering of the available phase angles into two relatively narrow ranges that each occupy less than 10% of the available pseudorotational pathway (Rinkel and Altona, 1987). The 'North-like' conformation phase angles range from 3° to 34° but there is a preference for the lower range with a

mean of 10°. 'South-like' phase angles ranges from 139° to 175° with a mean of 158°. For all values of the phase angle, P, the amplitude, Φ, is relatively constant, ranging from 35° to 45° with a mean of 39°.

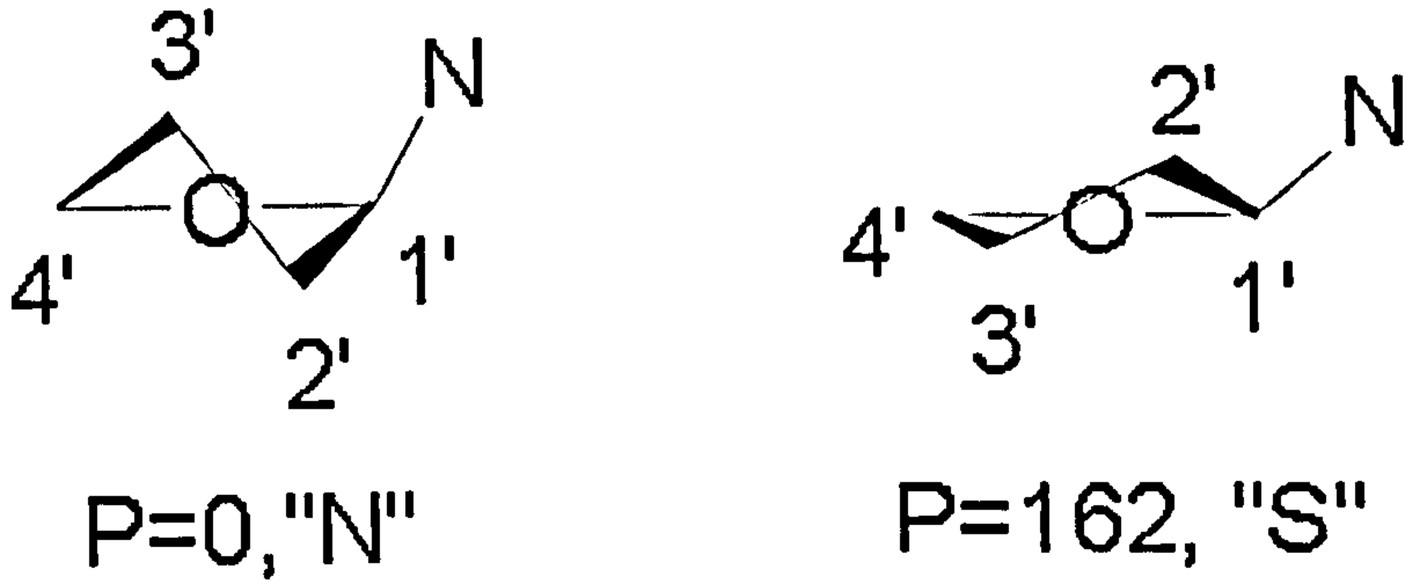


Figure 3-11: Stable conformations adopted by the deoxyribose sugars showing the northern (C3' endo) and southern (C2' endo) conformations, where N represents the nitrogen base.

The structure calculations on the complex show that the DNA duplex is right-handed, with all base pairs adopting the Watson-Crick configuration. Excluding the terminal nucleotides, the majority of the deoxyribose rings adopt the "S" conformation of B-DNA (Chazin et al. 1986), and are typically in the range P = 120°–160°.

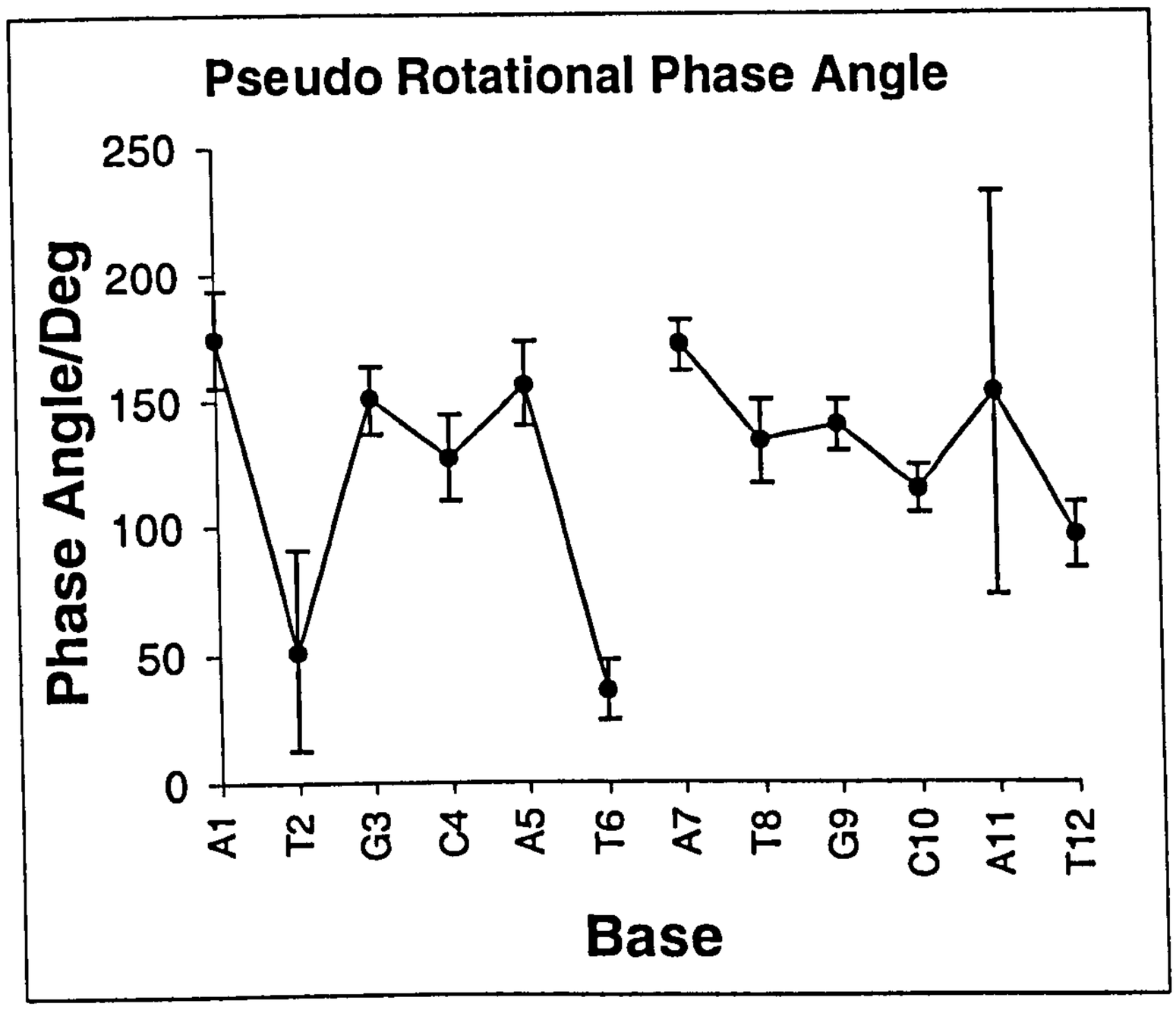


Figure 3-12: Pseudo rotational phase angle calculated for each base. Values are the mean average calculated from the last 200ps of a 1ns trajectory of rMD calculations. Error bars represent one standard deviation from the mean average.

Pseudorotational phase angles were calculated from the last 200 ps of the dynamics calculation. The calculated mean and error bars representing one standard deviation are in Figure 3-12. The 5'-terminal thymines (T6 and T12) have smaller values of between 30°–80°, in part reflecting their locations in the terminal positions. In general the standard deviations from mean values are relatively small. The sugar puckers of T2 and A11, located at the intercalation site, are the notable exceptions. These nucleotides have much larger standard deviations indicating that many conformations of similar energy are accessible over the course of the dynamics simulation, reflecting a change in steric interactions with neighbouring bases as the helix unwinds to accommodate the drug.

Table 3-5: Determination of sugar pucker and conformational purity based on coupling constants for H1' to H2'/2''.

Sugar H1'	ΣH1' (Hz)[a]	P (°)[b]	Apparent Multiplet[c]	Sugar Pucker[d]
G3	15.83	162	q (1:8:0.8)[e]	"S"
T6	13.37	36-54	t (1:1.5:0.8)	50%-60% "S"
T8	15.33	180	t (1:3.4:1)	"S"
G9	14.35	54-72	t (1:2.4:1)	"S" 80%
C10	15.63	171	q (1:1.8:0.9:1.2)	"S"
A11	11.26	18	t (1:1.6:1)	"N" 30% "S"

[a] Sum of all J couplings contributing to the H1' $\Sigma H1' = {}^3J_{H1',H2'} + {}^3J_{H1',H2''}$ Notes: ΣH1'>13.3 Hz Predominance of South (pS > 60%),. ΣH1'<12.5 Hz predominance of north (pS < 40%)

[b] Pseudo rotational phase angle describing the sugar pucker. Values of P based on a single conformational mode. Values were extrapolated between points with a 35° amplitude.

[c] The apparent multiplet structure of the H1' signal t=triplet, q=quartet. The ratio of peaks determined from Gaussian curve fitting is indicated in brackets.

[d] The predominant sugar conformation and % 'purity' of the conformer based on a two state model

[e] The resonance for G3 is a broad quartet. The ratio for the centre peak is the sum of the two innermost peaks.

Analysis of the coupling constants and sum of the $^3J_{H1'H2'}$ and $^3J_{H1'H2''}$ constants from the 1D spectra allows an estimation of the phase angle adopted by the sugar from experimental information (Table 3-5). The approach assumes that the rings adopt a single conformation, however, interconversion between conformations is likely, occurring on the nanosecond time scale (Porschke, 1978). The observed coupling constants are a time-averaged property. Since all the conformations sampled during the course of the NMR experiment contribute to the coupling constant, the values can be

interpreted in terms of population of 'S' and 'N' type conformations. It is experimentally difficult to determine all coupling constants accurately due to line width and overlap of resonance signals. Calculations of coupling constants versus sugar pucker shows that approximations can be made with only the sum of the H1'-H2' and H1'-H2'' coupling constants. The experimental data indicate that the sum of coupling constants (ΣH1') for T2 is relatively small (<12 Hz) indicating a significant deviation from a predominantly "S" conformer.

$$^{3}J_{HH} = P_1 \cdot \cos^2\phi - P_2 \cdot \cos\phi + P_3 + \sum_{i=1}^{3} \Delta\chi_i \left[P_4 + P_5 \cdot \cos^2\left(\xi_i\phi + P_6\left|\Delta\chi_i\right|\right)\right]$$

Equation 9: Modified Karplus equation for deoxyribose sugar rings to determine the 3 bond coupling constant for an angle ϕ. Where the exact values for P_1-P_6 are dependent upon the number of substituents. For the C1' and C2' torsion with 3 non-hydrogen substituents the values of P are: P_1 = 13.22, P_2 = -0.99, P_3 = 0, P_4 = 0.87, P_5 = -2.46, P_6 = 19.9°. $\Delta\chi$ is an electronegativity correction for each nucleus the values are $\Delta\chi C$ = 0.4, $\Delta\chi N$ = 0.85, $\Delta\chi O$ = 1.30. ξ takes into account the orientation of the non-hydrogen atom relative to the hydrogen atom. ξ is 1 for an angle of 120° and –1 for an angle of -120°.

Coupling constants were determined for $^{3}J_{H1'H2'}$ and $^{3}J_{H1'H2''}$ for structures in the dynamics trajectory using the modified Karplus equation (Equation 9). This allows an average value for ΣH1', to be calculated based on the rMD simulations, and allows comparison with experimental. The average value for ΣH1' calculated for T2 over the course of the trajectory is 12.1 Hz, in good agreement with experiment (Table 3-6) and reflects the predominantly northern geometry populated during the rMD. In contrast, the sugars of the base pair G3–C10, which also stacks with the bound intercalator, have much higher proportions of the C2'-endo conformation, as judged by larger ΣH1' values (> 15 Hz), and show a very small standard deviation from the value for the

average structure during the dynamics simulation. Again, ΣH1' values calculated from the dynamics simulation are in good agreement with experiment (differences < 1 Hz).

The coupling constant, calculated from the molecular dynamics simulation, for the A11 is higher than the experimentally measured value. Dynamics calculations show a high degree of variability but confine the sugar to an average southern conformation. Experimentally, the sugar pucker is weighted more towards the northern geometry (ΣH1' = 11.6 Hz). Although the populations are not accurately represented, the large degree of variability of sugar pucker suggests the S conformation possesses a shallower energy minimum compared to normal undistorted B-DNA. The effects of drug intercalation on deoxyribose conformations are consistent with those seen previously by NMR for binding at a TpG step (Searle and Wakelin, 1990), while X-ray data have generally been fitted to a single conformer.

Table 3-6: Calculated sum of coupling constants derived from the mean of 500ps of dynamics calculation versus experimentally measured values.

Base	Calculated ΣH1'[a]	Experimental ΣH1'[b]
A1	15.62 (0.65)	n/a[c]
T2	12.07 (1.26)	n/a[c]
G3	15.15 (0.86)	15.83
C4	15.41 (0.7)	n/a[c]
A5	15.29 (0.7)	n/a[c]
T6	11.26 (1.21)	13.37
A7	15.44 (0.7)	n/a[c]
T8	14.99 (0.75)	15.33
G9	14.94 (0.77)	14.35
C10	15.21 (0.64)	15.63
A11	14.43 (1.09)	11.26
T12	15.59 (0.72)	n/a[c]

a. Where ΣH1' is the sum of $^3J_{H1'-H2'}$ and $^3J_{H1'-H2''}$ in Hz.
b. Values determined by Gaussian curve fitting to multiplets in Hz.
c. Values that can not be determined experimentally due to resonance overlap

3.2.7 Conformation of the Phosphodiester Backbone

Some significant changes in backbone torsion angles are also associated with drug intercalation. The unwinding and distortion necessitated by drug binding introduces a specific change from the preferred B_I conformation to the less common B_{II}

conformation. B_I and B_{II} conformations are described by the torsion angles ε of residue *n* and ζ of residue *n+1* such that B_I = t (ε_n =±180°), g^- (ζ_{n+1}=-60°), and B_{II} = g^- (ε_n=-60°), t (ζ_{n+1}=±180°), using the nomenclature of Diekmann (Diekmann, 1989)

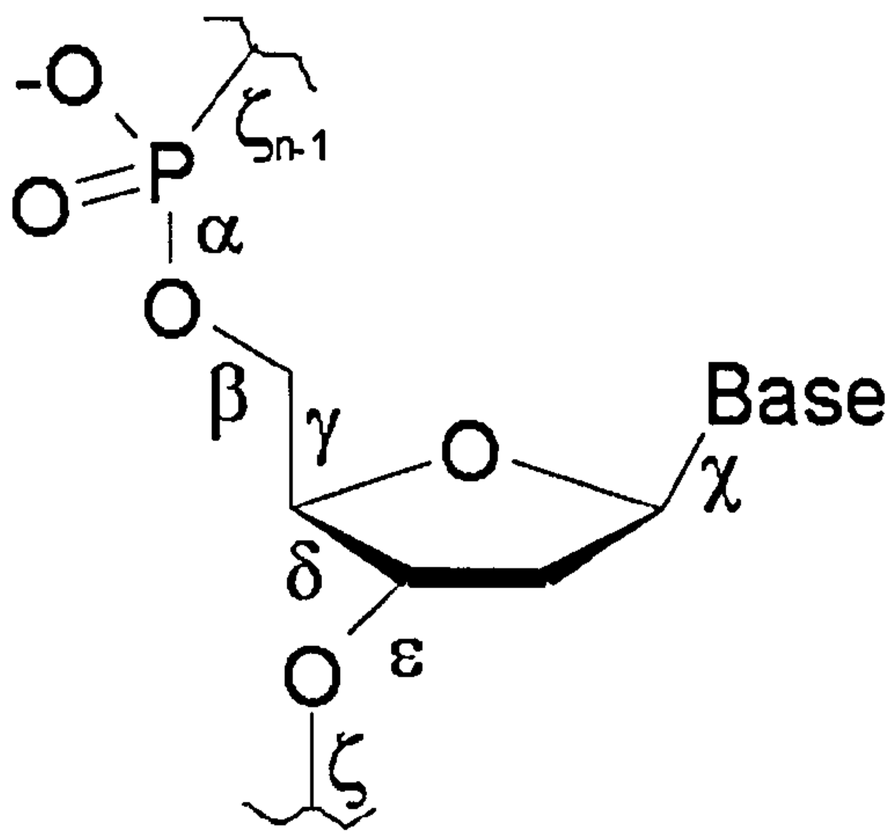


Figure 3-13: Torsion angles describing the phosphodiester backbone based on the Diekmann nomenclature.

This occurs specifically on one strand of the duplex at the CpA step. The dials plots (Figure 3-14) clearly illustrate the differences in backbone conformation along the sequence with changes in ε of C10 and ζ of A11 reflecting a flip from the B_I (t, g^-) to B_{II} (g^-, t) conformation. This localised asymmetric distortion of the backbone at the CpA step is consistent with ^{31}P chemical shift data that show a large downfield shift of > 1 ppm only for the CpA phosphate. This is in agreement with earlier studies of nogalamycin-DNA interactions (Lane and Searle, 1992). The B_I to B_{II} conformational transition at the CpA step in response to drug binding does not result in any increase in backbone dynamics compared with other base steps, as illustrated by the amplitude of fluctuations during the course of the dynamics trajectory (Figure 3-17).

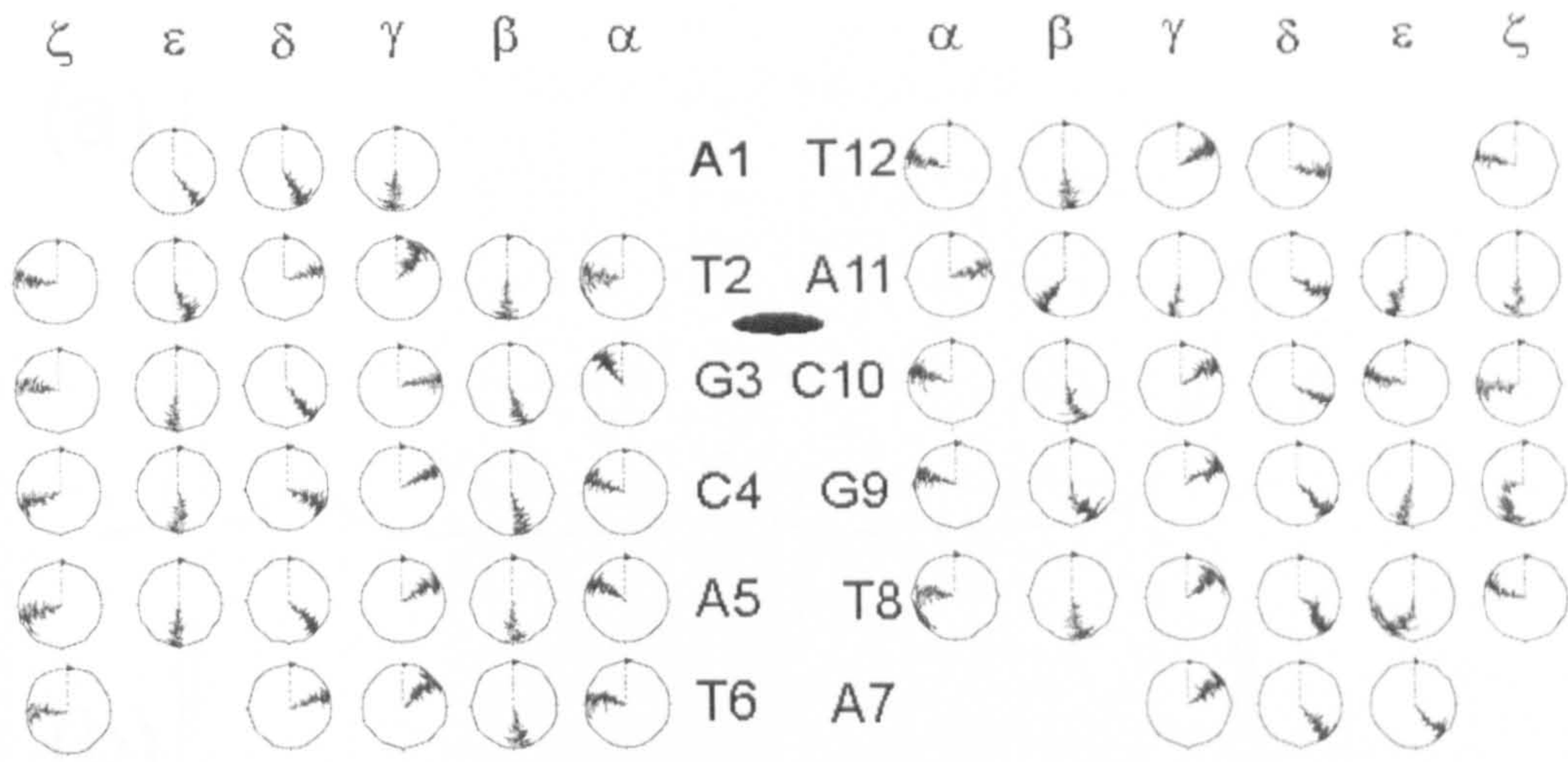


Figure 3-14: Dials plots. Each dial represents the angle formed at a particular point during the dynamics with time increasing from the centre to the outside of the dial. The nogalamycin is represented as a filled ellipse and shown intercalating at the T^2pG^3 step.

The ^{31}P NMR shows two DNA phosphates with significant chemical shift perturbation in comparison to the bulk of the phosphate environments (Figure 3-15). This is indicative of an intercalative mode of binding. The spectra recorded in the presence of proton coupling gives an increase in peak width proportional to the unresolved H3' and H5' couplings. The difference in peak width and chemical shift between the two phosphates at the intercalation site reveals the asymmetry of binding. This is consistent with the observed DNA backbone conformations.

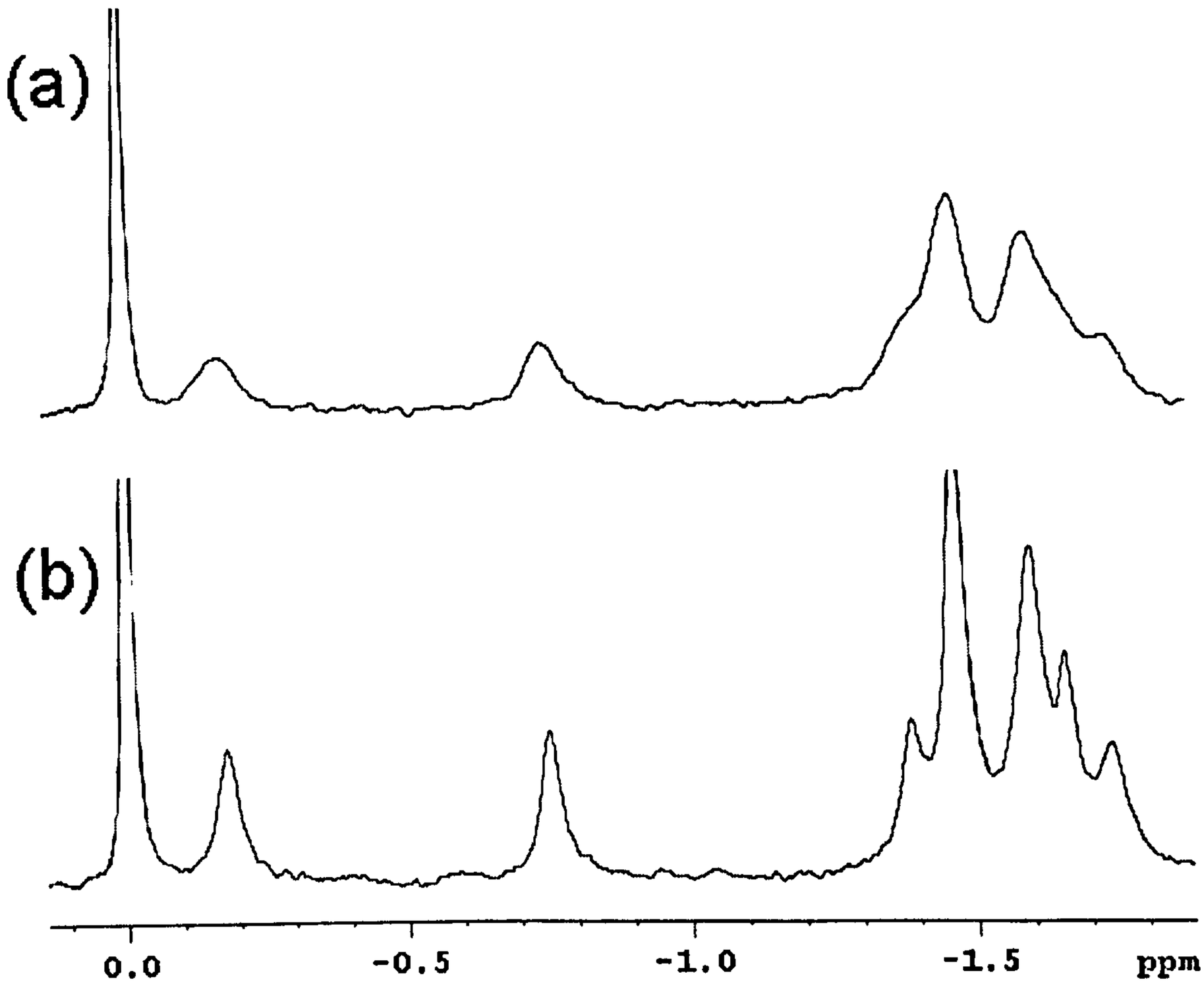


Figure 3-15: ^{31}P spectra of d(ATGCAT)$_2$-nogalamycin recorded at 288K (a), and ^{1}H decoupled (b). Calibrated relative to the phosphate signal from buffer

Earlier studies of ^{31}P chemical shifts, ^{31}P-^{1}H coupling constants, T_1 and T_2 relaxation measurements and estimates of chemical shift anisotropy, while consistent with an asymmetric distortion of the DNA backbone, similarly do not support large amplitude motions of the phosphate groups at the intercalation site (Lane and Searle, 1992).

3.2.8 Helical twist, rise and slide

The structural parameters for DNA base pairs and base steps give an insight into the structural deformations introduced upon binding of the antibiotic (Figure 3-16). The

parameters can also be collated over the course of the restrained molecular dynamics calculations. Slide, rise, helical twist, shift, buckle and sugar pucker have been analysed as shown in Figure 3-17. The mean values are indicated with error bars representing one standard deviation from the mean. This shows not only the mean value but allows the fluctuations experienced during the molecular dynamics simulations to be observed.

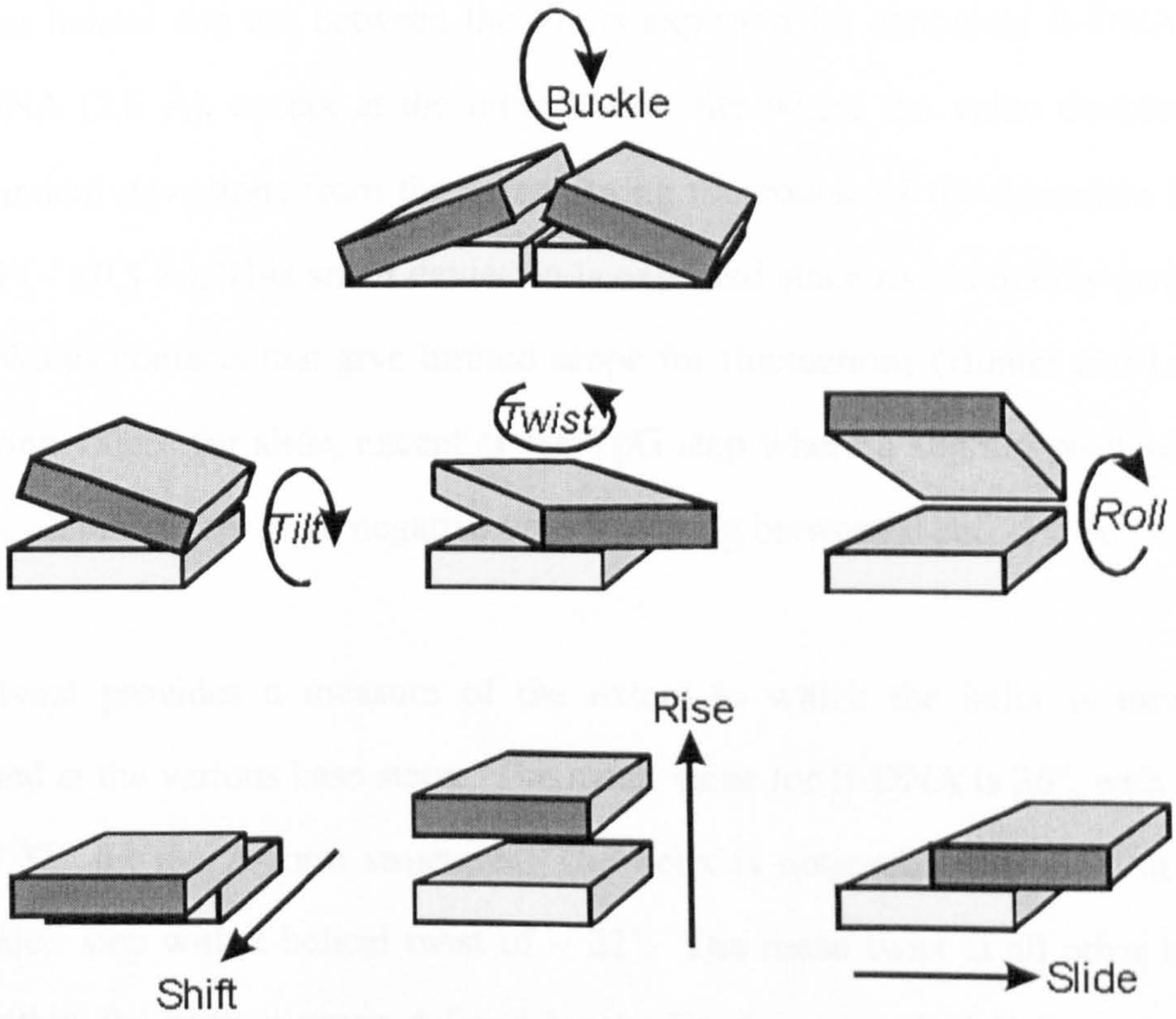


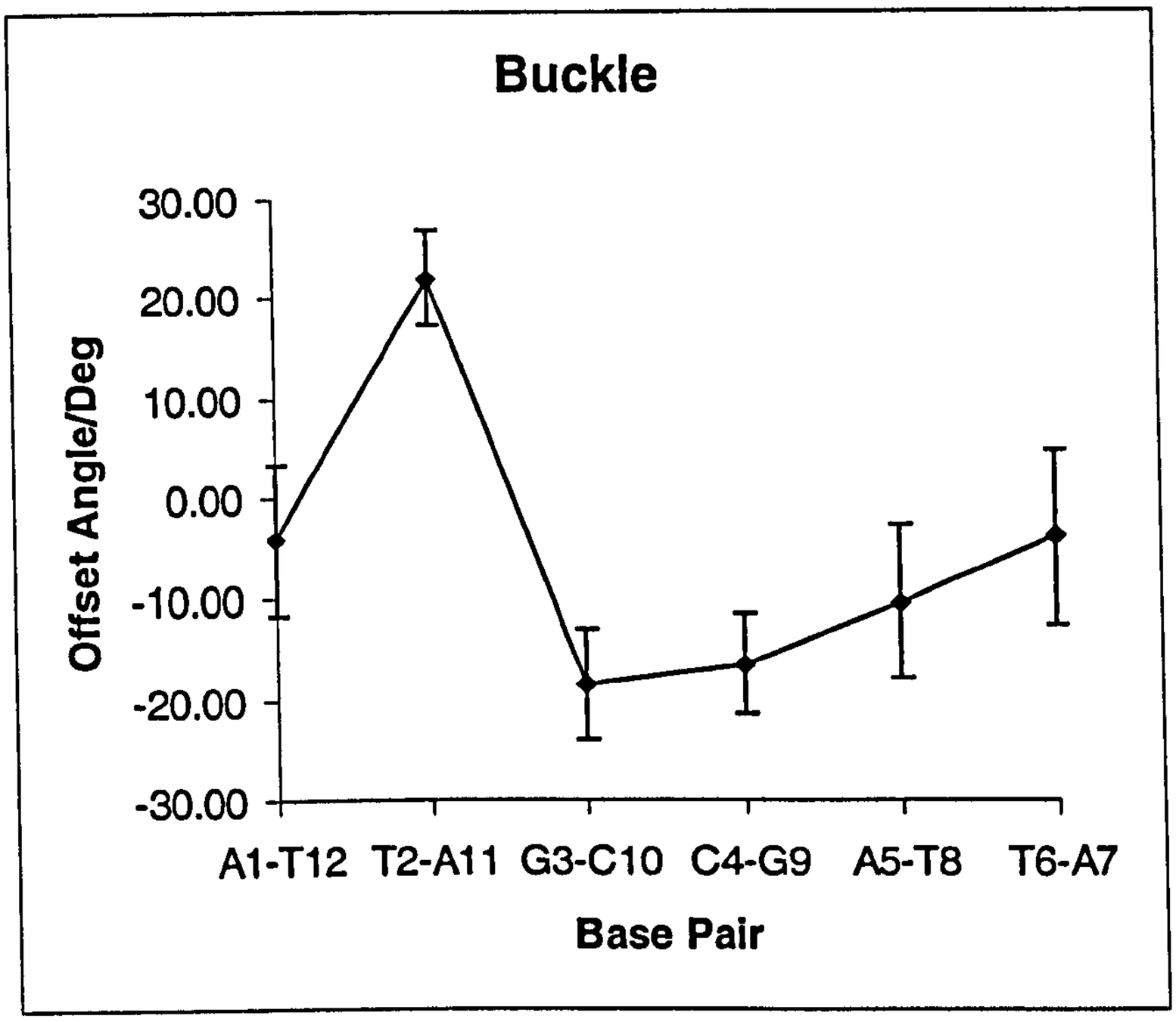
Figure 3-16: Translational and rotational parameters show degrees of freedom for a nucleotide base and base pair.

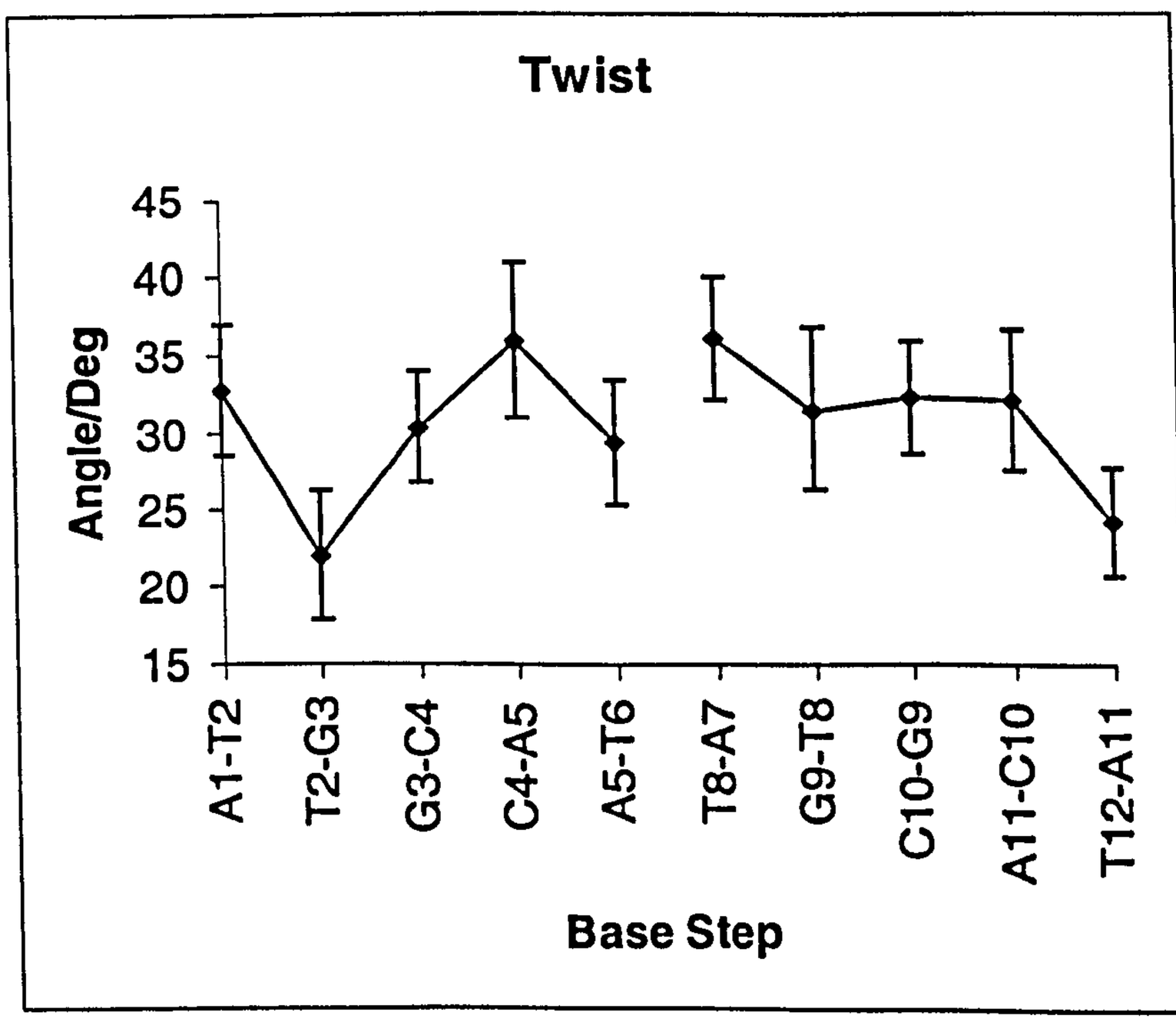
Two factors affect the accessible values for the structural parameters in free DNA. These factors are the stacking interactions between base pairs and the constraining force exerted by the sugar phosphate backbone. The interplay between these factors couples

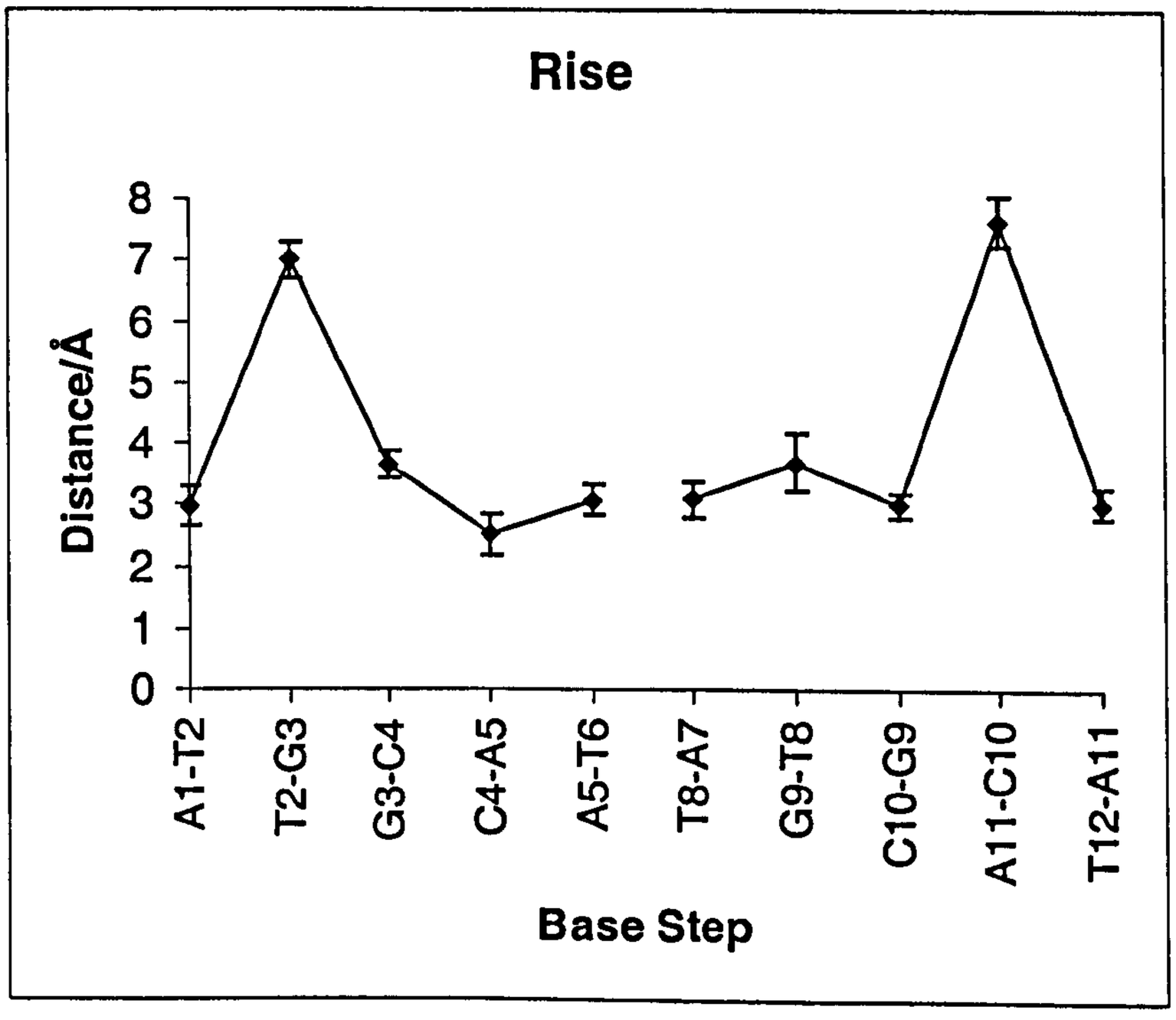
structural parameters and backbone angles reducing the number of degrees of freedom (Packer and Hunter, 1998). The backbone also serves to couple structural variation at a base step to neighbouring base steps. This has important consequences since this leads to a propagation of effects to near neighbours, especially when large distortions from ideal geometry in the structure are present.

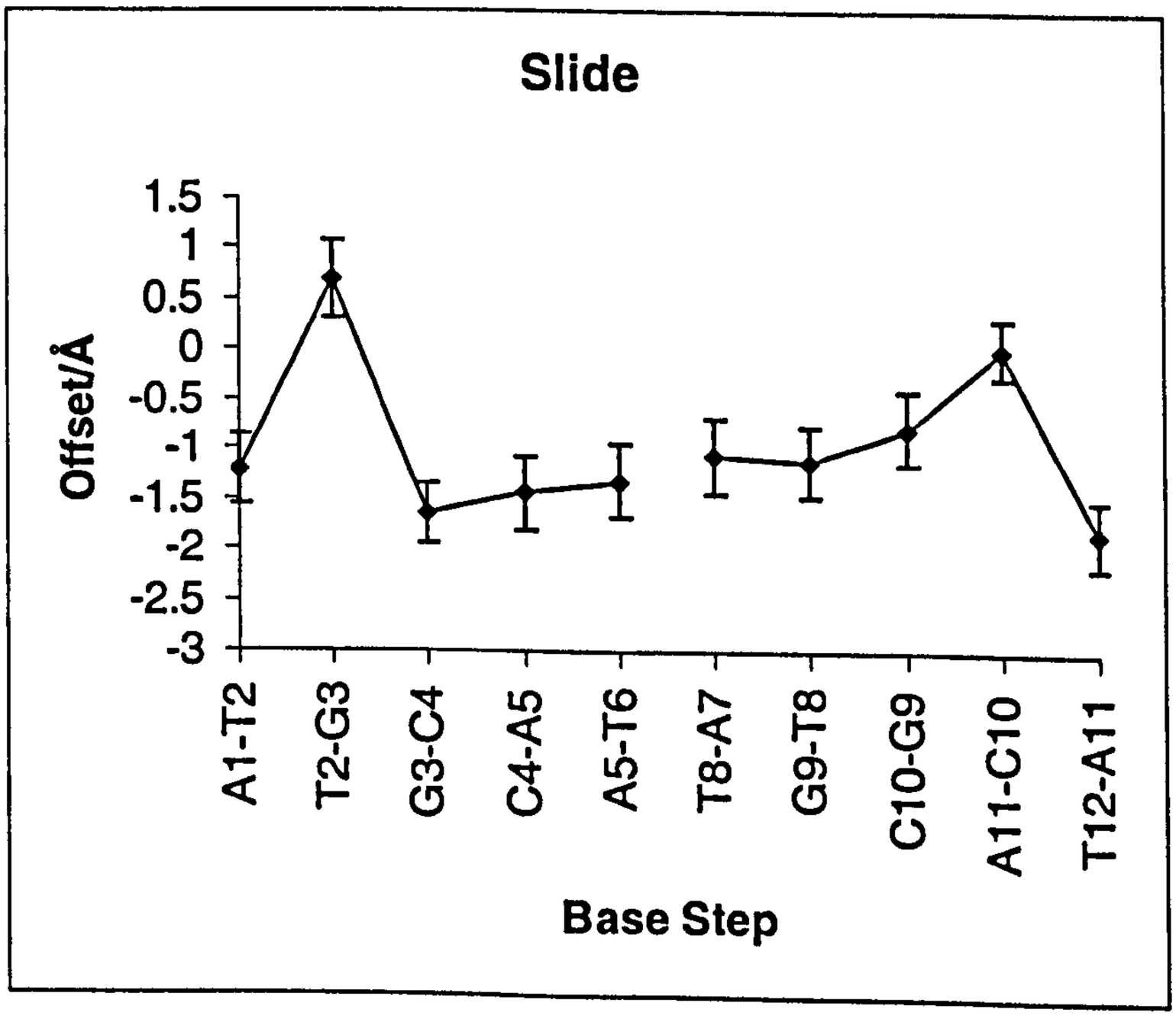
Values for helical rise are between the limits expected for canonical B-DNA (3.3 Å) and A-DNA (2.6 Å), except at the intercalation site where the value doubles. In all cases, standard deviations from the mean during the course of the dynamics trajectory are small (~ ±0.5 Å). This small deviation is expected since rise is mainly governed by van der Waals contacts that give limited scope for fluctuations (Hunter and Lu, 1997). Considering values for slide, except at the TpG step where a slightly positive value is observed, all other steps have negative values varying between 0 and -1.5 Å.

Helical twist provides a measure of the extent to which the helix is unwound or overwound at the various base steps. The mean value for B-DNA is 36°, with a smaller value of 32° for the A-form structure. The helix is noticeably unwound at the TpG intercalation step with a helical twist of ~ 22°. The mean twist at all other base steps varies within the narrow range defined by the limiting canonical B-form and A-form structures. However, when standard deviations from these values are considered (typically ±5°) then the structure is in constant flux between partially unwound and partially overwound conformations (Table 2). It is interesting to note that twist and slide variations are accommodated by an increase in the backbone length. Hence such unwinding may help alleviate the strain incurred by an increased rise.


Buckle
30.00
20.00
10.00
0.00
-10.00
-20.00
-30.00
Offset Angle/Deg
A1-T12
T2-A11
G3-C10
C4-G9
A5-T8
T6-A7
Base Pair


Twist
45
40
35
30
25
20
15
Angle/Deg
A1-T2
T2-G3
G3-C4
C4-A5
A5-T6
T8-A7
G9-T8
C10-G9
A11-C10
T12-A11
Base Step


Rise
Distance/Å
8
7
6
5
4
3
2
1
0
A1-T2
T2-G3
G3-C4
C4-A5
A5-T6
T8-A7
G9-T8
C10-G9
A11-C10
T12-A11
Base Step


Slide
Offset/Å
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
-3
A1-T2
T2-G3
G3-C4
C4-A5
A5-T6
T8-A7
G9-T8
C10-G9
A11-C10
T12-A11
Base Step

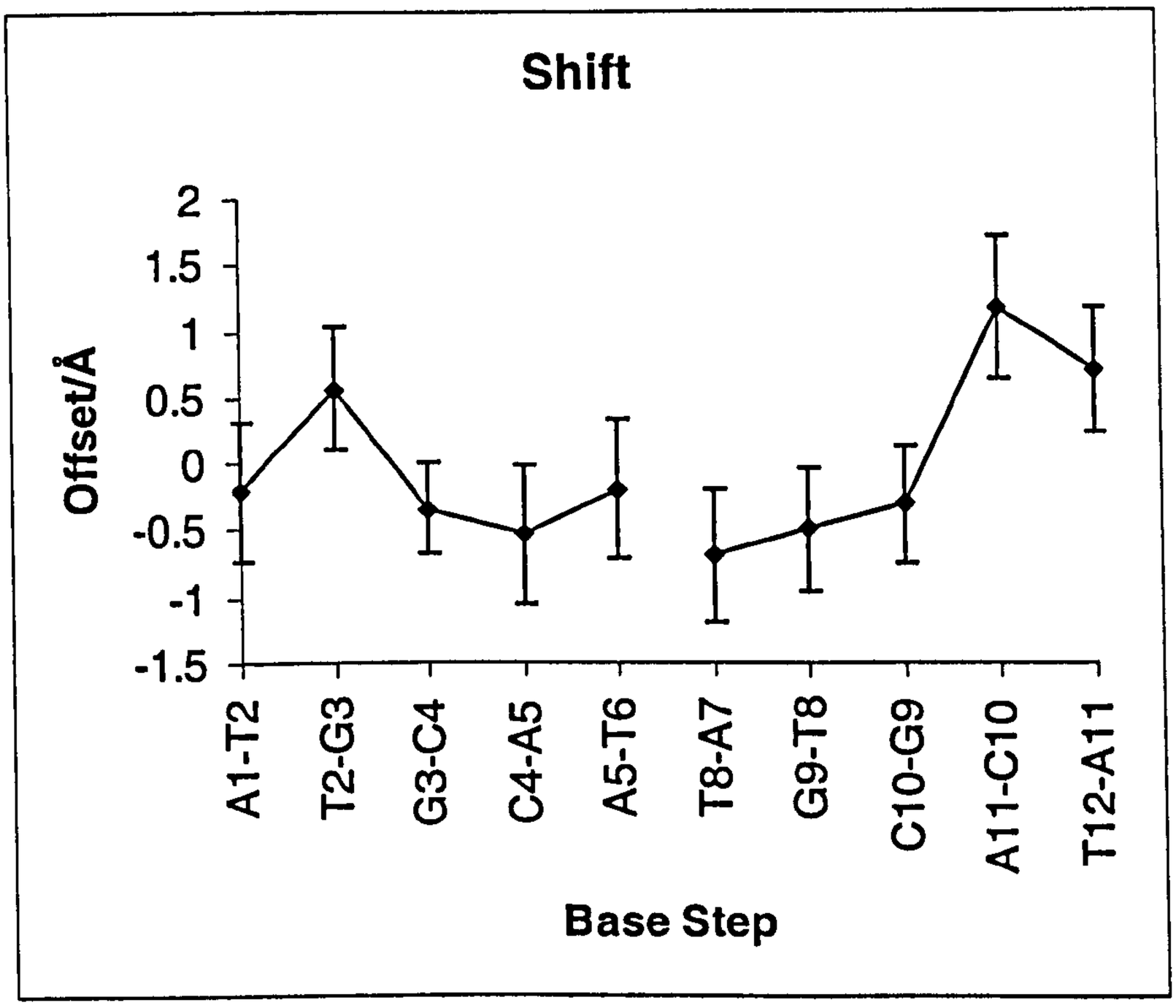


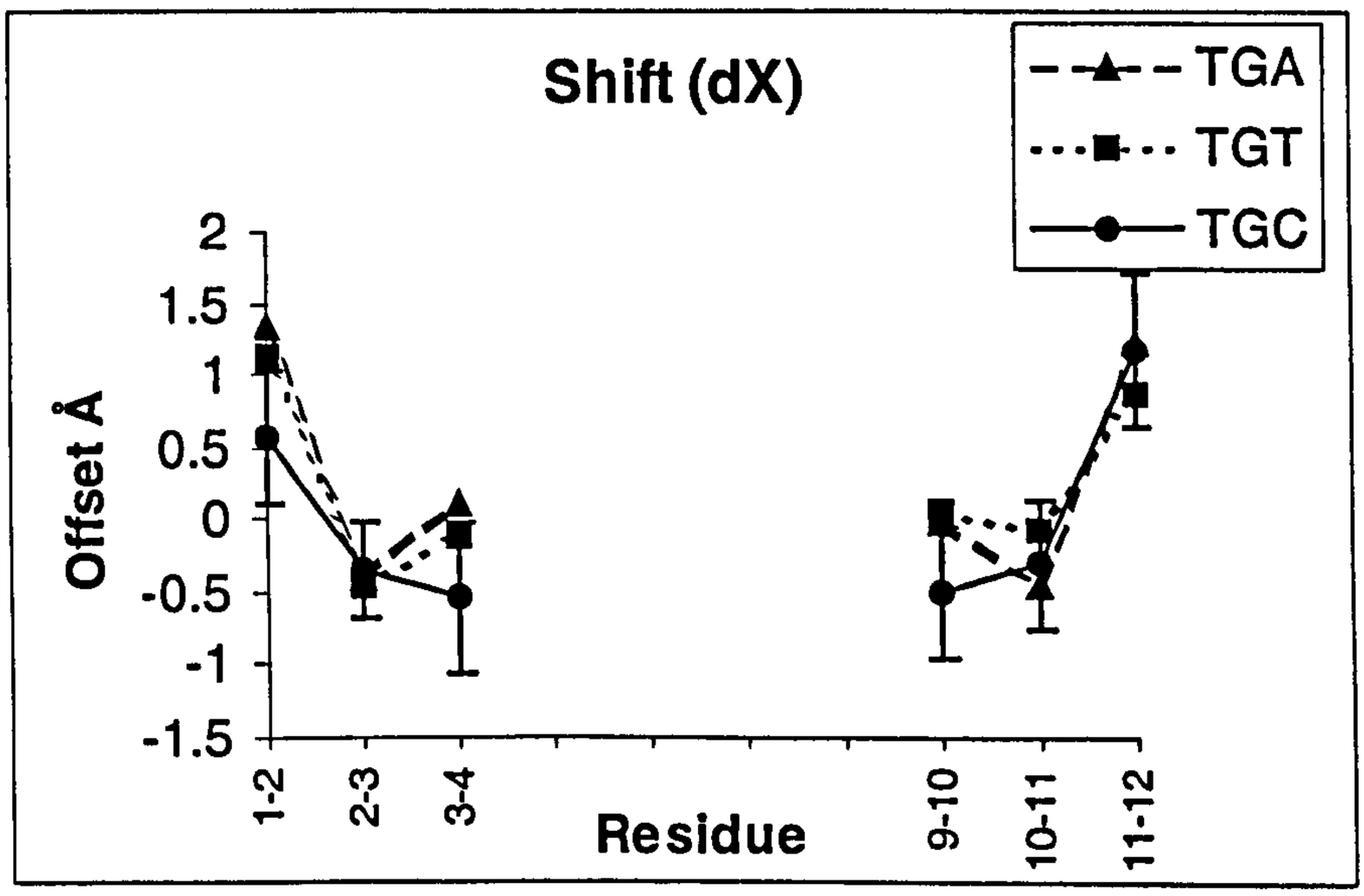
Figure 3-17: *Structural parameters. The values represent mean sampled over the last 500 ps of a 1.1 ns dynamics calculation with error bars representing 1 standard deviation from this mean. Slide, twist, rise, shift are for each base step. The value of buckle is recorded for each base pair.*

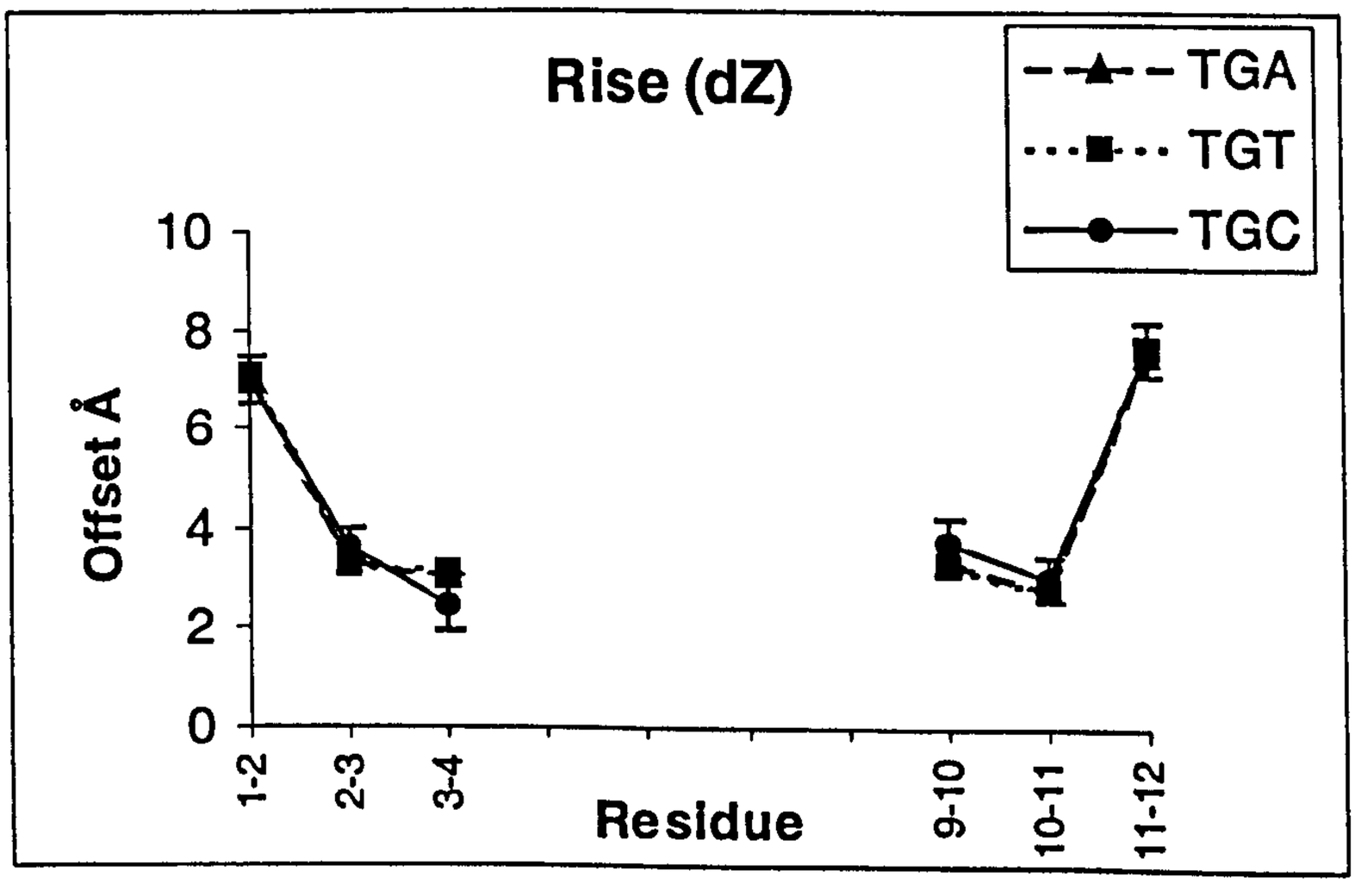
The values observed for shift are negative (~ -0.5Å) for all base steps except at the intercalation site. The results from the dynamics show positive shift values at the T2-G3 and A11-C10 base steps. There is also a marked difference in the shift value adopted by the CpA of 1.2 Å versus that adopted by the TpG step of 0.5 Å. The positive shift again affects the backbone length. Large positive shifts are coupled to the B_I to B_{II} transition notably the values adopted by the ε and ζ backbone torsion angles (Prive et al. 1987; Hunter and Lu, 1997). Such a correlation is consistent with the backbone information and ^{31}P NMR data.

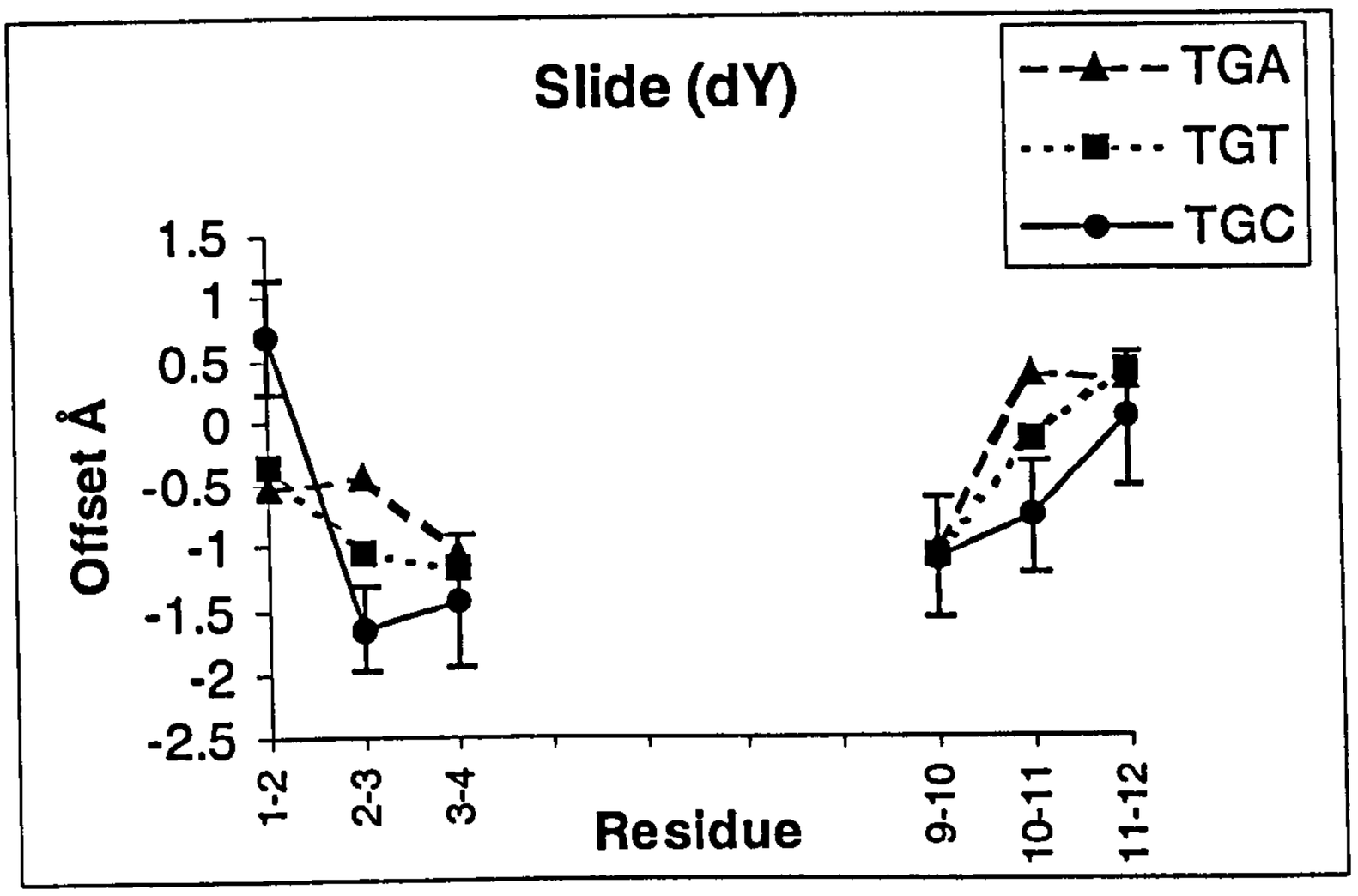
The major structural perturbations highlighted by all of these parameters are confined to the TpG and CpA intercalation site. There exists a large degree of asymmetry in the conformation of DNA at the intercalation site. As a consequence of maximising van der Waals interactions between drug and adjacent basepairs, the latter shows pronounced buckling. Slide, shift and helical twist show very similar dynamic fluctuations across the sequence, as reflected by the standard deviations shown in Figure 3-17. Since nogalamycin clamps bulky sugar substituents in both grooves of the helix, fluctuations in these parameters would certainly not be expected to be larger than at other base steps, on the basis that steric interactions with the drug restrict the amplitude of their motion.

3.2.9 Comparisons with 5'-TpG sites in X-ray structures

Figure 3-18 compares the structural parameters for the corresponding portions of the nogalamycin crystal structures (from deposited PDB co-ordinates), where the antibiotic is also bound at the 5'-TpG sites within the duplexes d(T•GTAC•A)$_2$ and d(T•GATC•A)$_2$ (Smith et al., 1995, 1996a). Superimposed are the analysis of rise, slide, twist and shift, calculated using CURVES. For the AT•GCAT site (this work) with that of the T•GTAC•A and T•GATC•A intercalation sites identified in the X-ray structures.


Shift (dX)
TGA
TGT
TGC
Offset Å
2
1.5
1
0.5
0
-0.5
-1
-1.5
1-2
2-3
3-4
9-10
10-11
11-12
Residue


Rise (dZ)
TGA
TGT
TGC
Offset Å
10
8
6
4
2
0
1-2
2-3
3-4
9-10
10-11
11-12
Residue



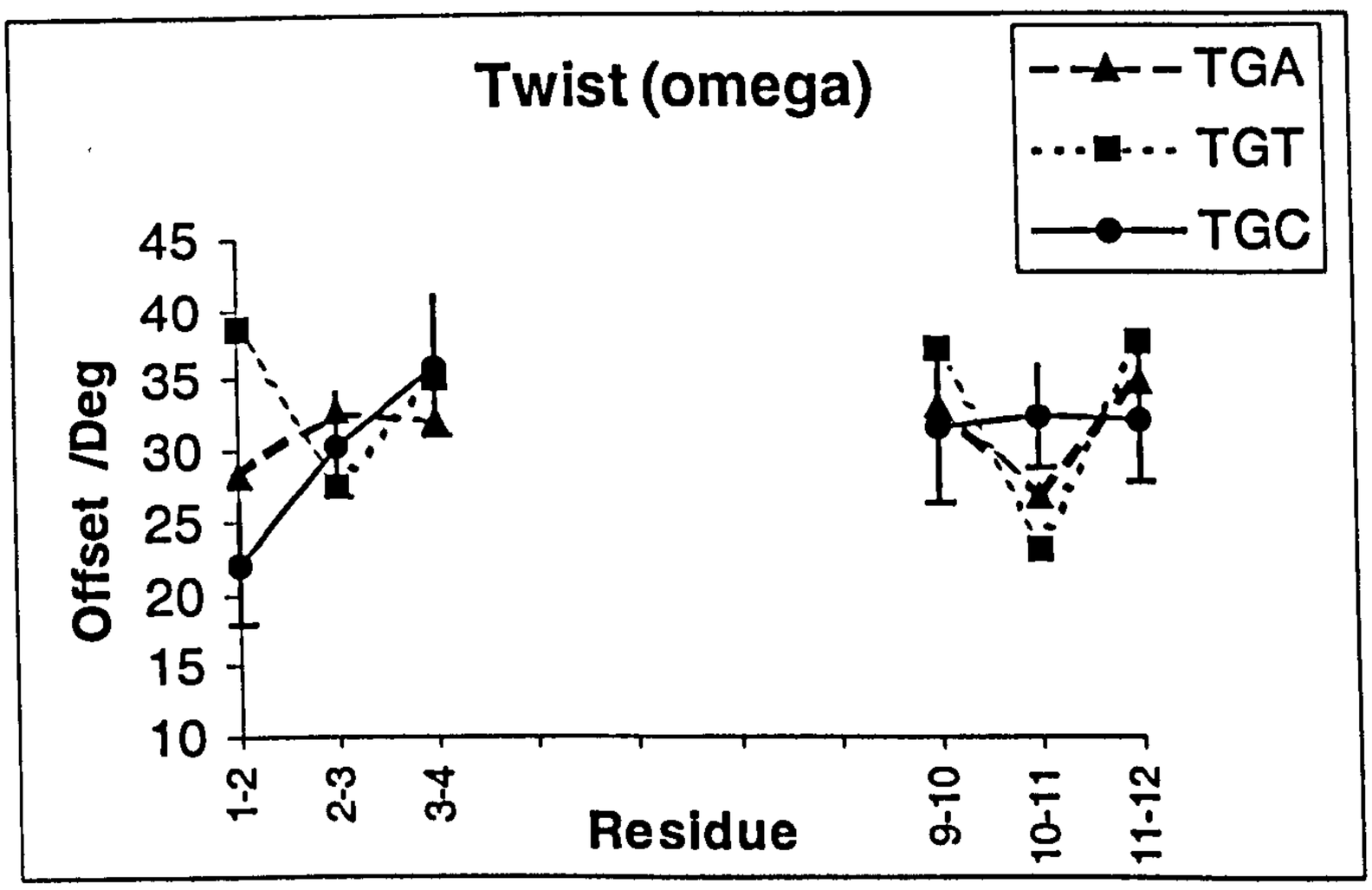


Figure 3-18: Comparison of structural parameters between TG binding sites from this study d(ATGCAT)$_2$-nogalamycin [d(-TGC-)] with crystal structures d(TGATCA)$_2$-nogalamycin$_2$ [d(TGA-)] and d(TGTACA)$_2$-nogalamycin$_2$ [d(TGT-)].

In previous work, (Smith et al. 1996) sequence-dependent differences in X-ray structures have been analysed in great detail. Included in Figure 3-18 are standard

deviations from the mean NMR structure of the AT•GCAT complex. Shift and rise are very similar in all three cases with standard deviations, in large part, encompassing any differences between structures. Helical twist is very similar for the AT•GCAT and T•GATCA structures (again within the range indicated by the standard deviations). However, the T•GTACA structure shows a large overwinding at the TpG step compensated by an underwinding at adjacent steps (see Figure 3-18). Slide values are generally slightly negative in the X-ray structures but are positive at the TpG intercalation site in the NMR structure. Subsequent steps show a more negative slide than the X-ray determined structures.

Analysis of the dynamics of the complex during the rMD simulation provides some insight into structural fluctuations in both the DNA conformation and that of the drug. The covalent architecture of nogalamycin makes it a fairly rigid structure, with the only significant degree of flexibility possible through rotations about the glycosidic bonds connecting the nogalose and ring A. Some flexing of the sugar pucker in the bicyclic ring system is also evident but the pairwise RMSD between drug structures over the final 500 ps of the dynamics is small (0.26 Å). The drug conformation is very similar in the two X-ray structures (RMSD 0.57 Å), which in turn are similar to the mean structure (RMSD 0.78 Å for TGT complex and 0.57 Å for TGA complex).

3.2.10 Drug-DNA Stacking Interactions and Intermolecular Hydrogen Bonding

Drug-base pair stacking interactions are illustrated in Figure 3-19, where only flanking base pairs (T2–A11 and G3–C10) in direct contact with the drug aglycone are included for clarity. The drug stacks more towards the TpG side of the intercalation site adding further to the intercalation site asymmetry. Van der Waals interactions with the

aglycone of the antibiotic appear to be maximised by the flanking base pairs buckling significantly, so as to more effectively wrap around the antibiotic (Figure 3-19).

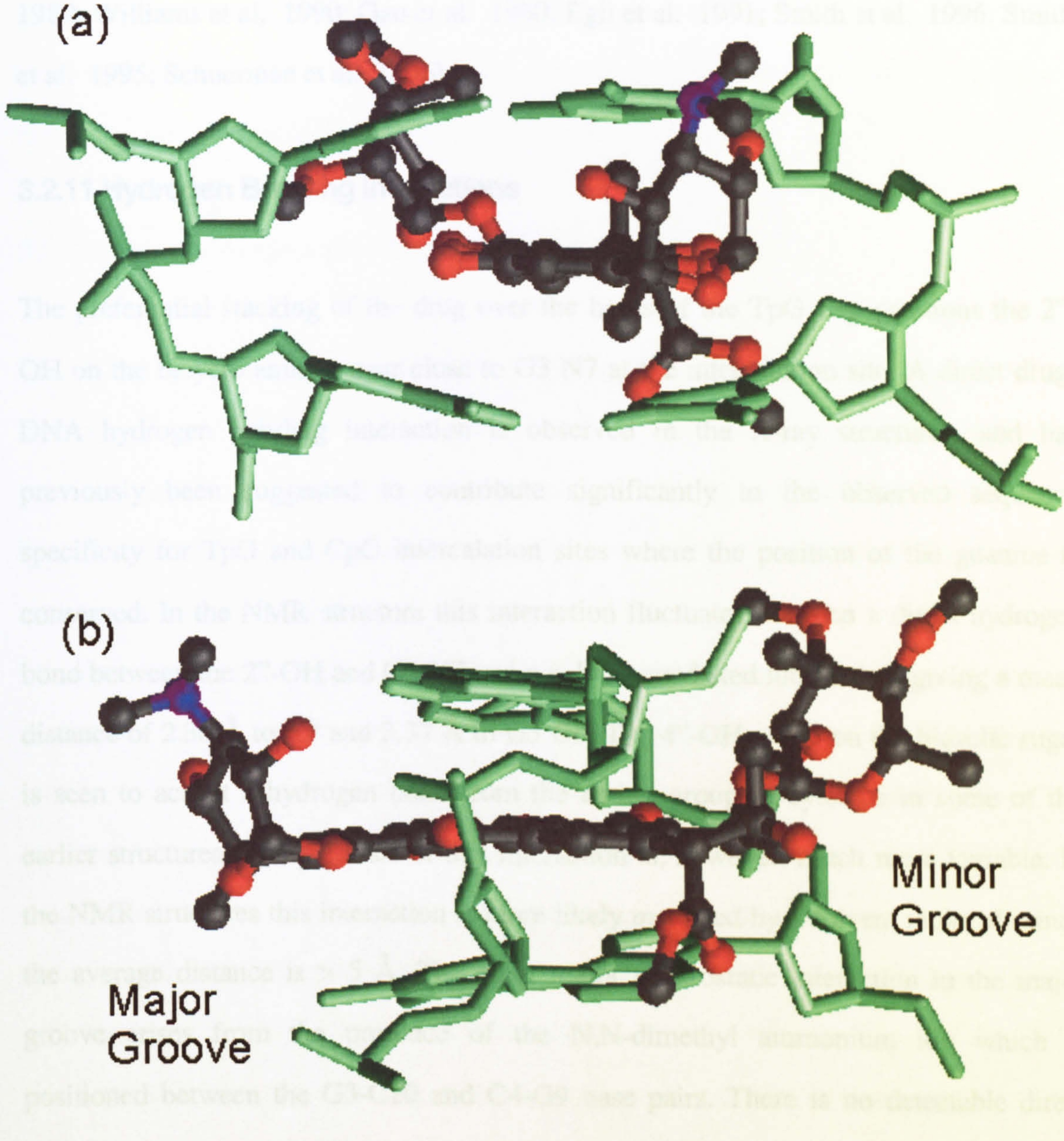


Figure 3-19: Stacking interactions at the TpG base step. Preferential stacking over the TpG can be seen (a). Nogalamycin threaded through the intercalation site with sugars interacting with both major and minor grooves (b).

This gives the intercalation site a wedge-shaped appearance, a feature commonly observed in X-ray structures of a variety of anthracycline-DNA complexes (Liaw et al. 1989; Williams et al. 1990; Gao et al. 1990; Egli et al. 1991; Smith et al. 1996; Smith et al. 1995; Schuerman et al. 1996).

3.2.11 Hydrogen Bonding Interactions

The preferential stacking of the drug over the bases of the TpG step positions the 2"-OH on the bicyclo amino sugar close to G3 N7 at the intercalation site. A direct drug-DNA hydrogen bonding interaction is observed in the X-ray structures, and has previously been suggested to contribute significantly to the observed sequence specificity for TpG and CpG intercalation sites where the position of the guanine is conserved. In the NMR structure this interaction fluctuates between a direct hydrogen bond between the 2"-OH and G3 N7 and a solvent-mediated interaction, giving a mean distance of 2.30 Å to N7 and 3.37 Å to G3 O6. The 4"-OH group on the bicyclic sugar is seen to accept a hydrogen bond from the amino group of cytosine in some of the earlier structures. The presence of this interaction is, however, much more variable. In the NMR structures this interaction is more likely mediated by a solvent molecule since the average distance is > 5 Å. The other major electrostatic interaction in the major groove arises from the presence of the N,N-dimethyl ammonium ion which is positioned between the G3-C10 and C4-G9 base pairs. There is no detectable direct hydrogen bonding with the base pairs, however, the positive charge sits in a region of high negative potential close to the N7 and O6 of G3, and between phosphates along the DNA backbone. This suggests that water-mediated electrostatic interactions may play an important part in stabilising the complex as discussed in the next two chapters. The carbonyl group of the methylester on ring A offers the only real potential for hydrogen

bonding in the minor groove but the TpG step, in contrast to the CpG step, offers no suitable direct hydrogen bond donor. The major contributions to binding in the minor groove appear to be the hydrophobic interactions with the many drug methyl and methoxy groups spread over the nogalose and ring A sugars which bury a significant non-polar surface area at the drug-DNA interface.

3.2.12 Molecular Basis for Nogalamycin-DNA Recognition

The structure of the nogalamycin-d(ATGCAT)$_2$ complex has been studied in detail by a combination of NMR and rMD simulations. The high-resolution structure (PDB file: 1qch) reveals molecular details of the interaction with the preferred TpG binding site. The analysis shows binding to this site in the thermodynamically preferred orientation, such as to exclude binding at the second TpG (CpA) step on steric grounds. Nogalamycin is able to thread through the DNA helix forming interactions in both the major and minor groove simultaneously. The interactions in the two grooves, that stabilise the complex, are fundamentally different in nature. In the minor groove, the hydrophobic nogalose sugar buries a large hydrophobic surface area, which must add considerably to the stability of the complex without imparting any particular sequence specificity. In contrast, interactions in the major groove are largely, but not exclusively, with hydrophilic groups that are able to interact directly with the edges of the DNA bases through hydrogen bonding. In particular, the drug 4'-OH is directed towards the N7 of Guanine at the TpG step, accounting for the occurrence of guanine at this position (XpG) in all high affinity binding sites. However, we have shown that this interaction is dynamic in nature, and for at least some part of the time may be solvent-mediated. As will be discussed in Chapter 4 and Chapter 5, solvent appears to play an important role in mediating interactions in the major groove. Water is a very versatile

"glue" which appears to be widely exploited in this context, as is also evident from X-ray analysis. Major groove interactions are not entirely electrostatic in origin. Hydrophobic interactions between the bridgehead sugar protons of the bicyclic sugar with the thymine methyl group at the TpG intercalation site are also evident. These latter interactions may in part explain the specificity for the TpG step.

3.3. Investigations into Nogalamycin Binding Specificities at d(TpG) Sites in Longer Sequences.

Studies were performed with two other sequences d(AGTGATCACT)$_2$ and d(GGTGATCACC)$_2$ to investigate binding at TG sites. Both sequences contain two TpG (and symmetry related CpA) high affinity nogalamycin binding sites. The sequences are symmetric, but occupancy of only one site will destroy this symmetry. Binding at the second site will restore the symmetry and this should give rise to an NMR signal pattern during drug titration that is easy to observe, as the spectrum should gradually 'simplify' as the end-point is reached for a 2:1 complex.

The initial titration (Figure 3-20) showed the appearance of the down field signal at 8.5 ppm comparable to that observed with the hexamer sequence. This resonance was attributed to the H8 of the adenine at the intercalation site. As further additions of nogalamycin were made, spectral complexity was observed to increase. Even at drug:duplex ratios of greater than 2:1 the expected simplification was not observed and a symmetrical 2:1 complex was not obtained.

The inability to form 2:1 drug:duplex complexes with the longer sequences is rather striking since X-ray structures have been reported of the core hexamer sequence $d(TGATCA)_2$. It was found for this sequence that two drug molecules are readily accommodated in symmetry related orientations at the TpG and CpA terminal sites. The initial conclusion was that when this sequence is embedded with a longer sequence, steric factors prevent more than one drug molecule occupying the two 'high affinity' sites. In the hexamer the greater flexibility appears to overcome this problem. The chemical shift pattern observed with the $d(AGTGATCACT)_2$ on binding with nogalamycin is consistent with one of the TpG sites being occupied.

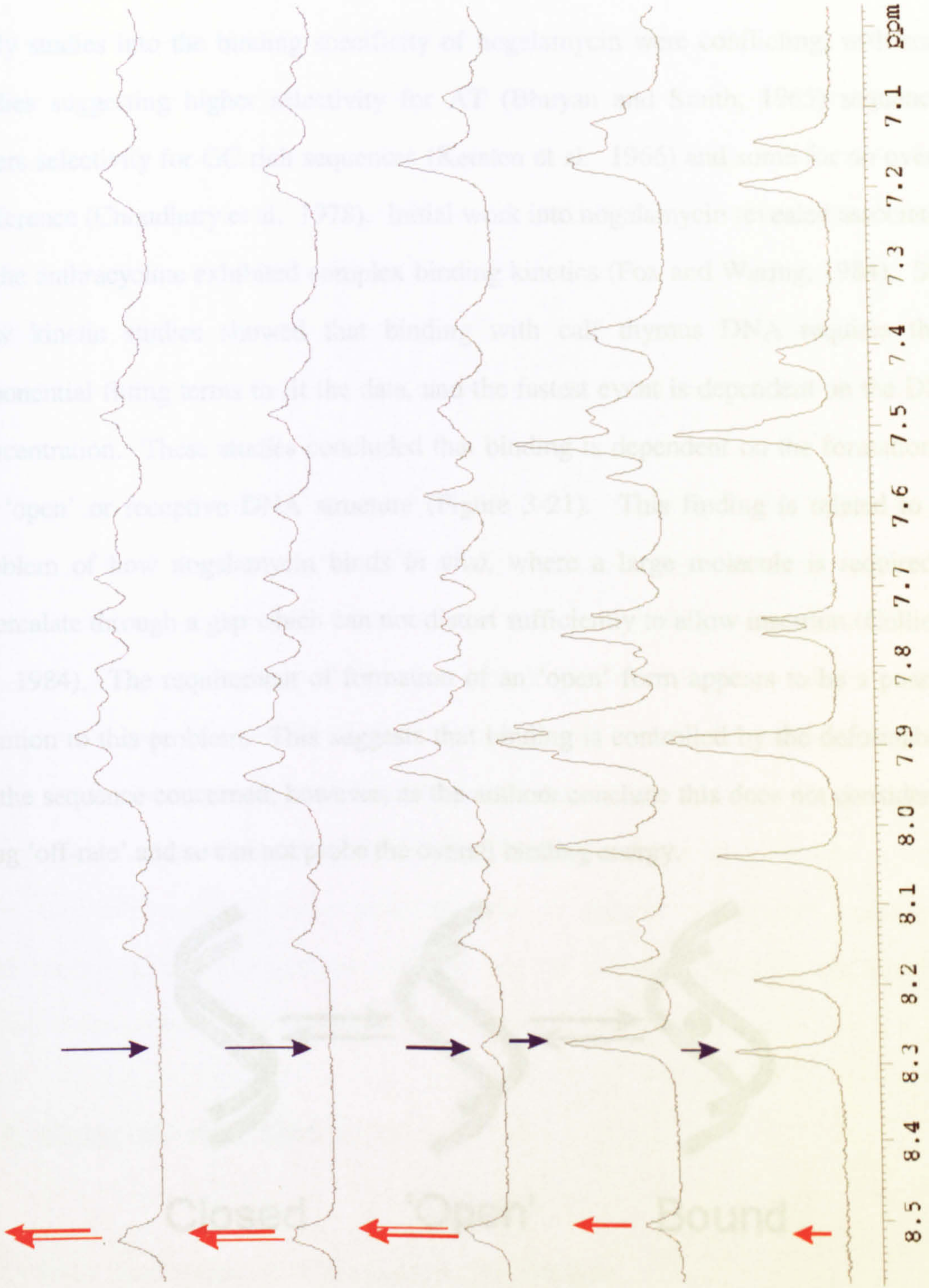


Figure 3-20: Titration of d(AGTGATCACT) with nogalamycin recorded at 298K. Increasing equivalence of nogalamycin versus duplex from 0:1) to 2:1. Arrows highlight a number of clearly defined signals due to free DNA that decrease (blue arrows) and increasing signals from the complex (red arrows). As the concentration of nogalamycin is increased the spectra fails to simplify indicating binding at asymmetric sites.

Early studies into the binding specificity of nogalamycin were conflicting, with some studies suggesting higher selectivity for AT (Bhuyan and Smith, 1965) sequences, others selectivity for GC rich sequences (Kersten et al. 1966) and some for no overall preference (Choudhury et al. 1978). Initial work into nogalamycin revealed association of the anthracycline exhibited complex binding kinetics (Fox and Waring, 1984). Stop flow kinetic studies showed that binding with calf thymus DNA requires three exponential fitting terms to fit the data, and the fastest event is dependent on the DNA concentration. These studies concluded that binding is dependent on the formation of an 'open' or receptive DNA structure (Figure 3-21). This finding is related to the problem of how nogalamycin binds *in vivo,* where a large molecule is required to intercalate through a gap which can not distort sufficiently to allow insertion (Collier et al. 1984). The requirement of formation of an 'open' form appears to be a possible solution to this problem. This suggests that binding is controlled by the deformability of the sequence concerned, however, as the authors conclude this does not consider the drug 'off-rate' and so can not probe the overall binding energy.

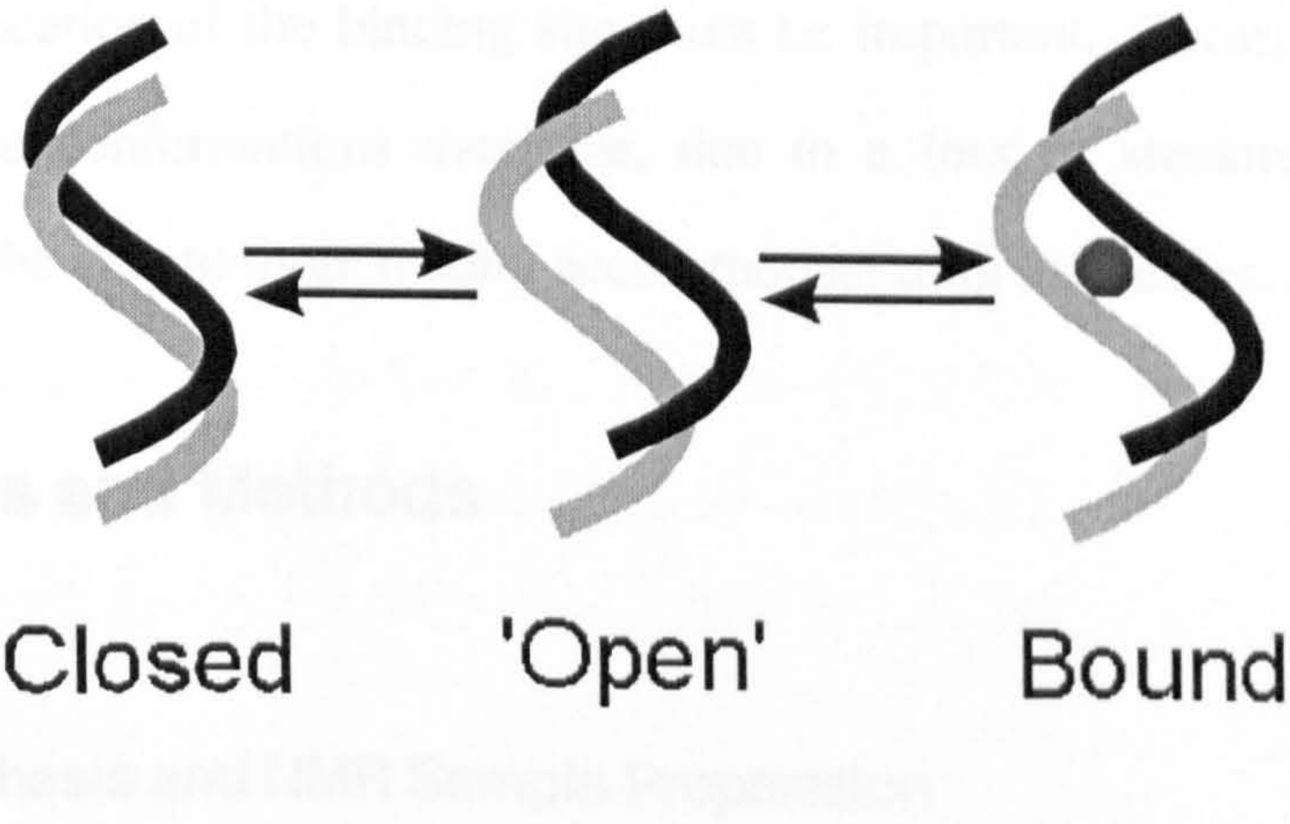


Figure 3-21: Schematic portraying sequence of events required for drug association. The rate determining step is thought to be the transformation to a receptive 'open' form.

Studies into the protection of sequences to cleavage by DNAse I showed an average of 4-5 bases were protected due to drug intercalation. The protection patterns were dependent on the concentration of the anthracycline. For protection to be observed concentrations of greater than 0.5 μM were required. Stable protection patterns were only observed in the 5-10 μM concentration range. The conclusion was that nogalamycin lacks absolute sequence specificity but is able to bind many sequences with different affinities. This finding is consistent with the kinetic model for binding.

The first binding event between the $d(AGTGATCACT)_2$ sequence and nogalamycin appears to be at the high affinity TG binding site, however, other sites are being occupied as the concentration is increased. This is not observed with the smaller hexameric sequences and is probably due to steric controls that inhibit further binding. Longer sequences pose problems, due to the complexity introduced by the occupancy of lower affinity sites, leading to complex, intractable, NMR spectra. If as suggested above the rate determining step for binding is the formation of a favourable 'open' form, then the location of the binding site must be important. Terminal binding sites with more viable conformations available, due to a loss of stacking and backbone interactions may be able to more readily accommodate drug molecules.

3.4. Materials and Methods

3.4.1 DNA Synthesis and NMR Sample Preparation

The DNA sequences d(ATGCAT), $d(AGTGATCACT)_2$ and $d(GGTGATCACC)_2$ were synthesised using standard solid-phase phosphoramidite chemistry and purified 'trityl-

on' by reverse-phase HPLC, using triethylammonium acetate (TEAA) buffer (pH 7.0) and an acetonitrile gradient. The dimethoxytrityl group was cleaved by treatment with 50% aqueous acetic acid for 1 h at 35°C. Acetic acid was extracted into ether and the oligonucleotides finally dialysed to remove residual TEAA and acetic acid, to introduce Na^+ as the counter ion. All sequences were shown to be >95% pure by ^1H NMR spectroscopy. The sample of nogalamycin was kindly provided by Upjohn, Kalamazoo, MI, USA and was used without further purification after checking by ^1H NMR. The drug was soluble in water at pH 4.0. Drug-DNA complexes were formed by adding aliquots of a stock solution to the oligonucleotide in 100mM NaCl, 10mM NaD_2PO_4. The end-point of the titration was determined by examination of signals from the free DNA.

For the hexamer sequence the titration was stopped at the 1:1 ratio of drug to duplex. At this point, signals from the free drug could not be identified suggesting that the drug may be in excess by only a few percent (ratio 1:<1.05).

The pH of the samples was adjusted to pH 7.0 at the end of the titration. Samples were lyophilised several times from D_2O. The sodium salt of trimethylsilylpropionate was used as an internal reference compound, sodium azide (0.6 mM) as an anti-bacterial agent and EDTA (0.6 mM) to complex any heavy metal ions.

3.4.2 NMR Analysis

NMR data were collected at 500MHz on a Bruker DRX500 spectrometer and processed on an R4600PC Silicon Graphics Indy workstation using XWINNMR software. Standard phase-sensitive 2D NMR pulse sequences were used throughout, including

NOESY (States et al. 1982), ROESY (Bax and Davis, 1985b; Bothner-by et al. 1984), DQF-COSY (Piantini et al. 1982; Marion and Wüthrich, 1983; Rance et al. 1983), TOCSY (Braunschweiler and Ernst, 1983; Bax and Davis, 1985a) and WATERGATE NOESY and ROESY (Piotto et al. 1992) for solvent suppression in 90% H_2O solutions. NOESY spectra were acquired at mixing times between 50 and 300 ms and typically 1024 complex data points were collected for each of 512 t_1 increments with 64 transients for each.

NOE restraints: A set of 205 NOE restraints were determined from NOE cross-peak integration in 500 MHz NOESY spectra collected at 288K. Volume integrals were normalised to several fixed reference distances within the DNA structure including deoxyribose H2'–H2" (1.85 Å), thymine CH_3–H6 (3.00 Å), cytosine H5–H6 (2.45 Å) and drug H8 geminal pair (1.84 Å) according to the NOEs being calibrated (Embrey et al. 1993). Distances were estimated from data at a number of mixing times using linear regression to extrapolate to 0 ms mixing time (Baleja et al. 1990). An error bound of 20% was applied to D_2O-derived restraints, and an error of 50% on distances estimated from a single H_2O NOESY spectrum.

3.4.3 Molecular Dynamics Simulations and Structural Analysis:

Structure calculations were carried out on an R10000QC Silicon Graphics server using AMBER 4.1 (Pearlman et al. 1995). The drug was parameterised and point charges determined by a semi-empirical approach within Spartan 3.1 (Wavefunction, 1995) using the AM1 method. Point charges were determined from an electrostatic potential surface. Charges were then symmeterised for freely rotateable systems in accordance with the procedures defined for the AMBER forcefield (Bayly et al. 1993).

The starting structure for the drug-complex was modelled using the LEaP module within AMBER (Pearlman et al. 1995) by sandwiching the drug between two fragments of canonical B-DNA, generated using the NUCGEN module and hydrogens were added with the LEaP module. Explicit counterions were added using a coulombic grid potential to neutralise the negative charge (11 sodium ions and one drug charge per complex) and the system was solvated to a minimum distance of 5 Å around the solute using boxes of 216 TIP3P water molecules. This gave a periodic box size of approximately 60 Å x 60 Å x 41 Å for the starting structure, respectively. All simulations were run using the SANDER module of AMBER 4.1 using a 1 fs time step, a temperature of 300 K with Berendsen temperature coupling (Berendsen et al. 1984) and a time constant of 0.2 ps. A 9Å cut-off was applied to the Lennard-Jones interactions (Berendsen et al. 1984), and constant pressure with isotropic position scaling was employed with a time constant of 0.2 ps. The list of non-bonded atoms was updated every 25 steps. Equilibration was performed by first holding the positions of the DNA, drug and counterions fixed and running 5000 steps of minimisation (500 steps steepest decent minimisation 4500 steps conjugate gradient method), followed by a further 5000 steps on the whole system. Then, with the co-ordinates of the DNA, drug and counterions frozen again using a harmonic potential constraint, the water was subjected to 10 ps molecular dynamics at 100K. After this initial equilibration, all subsequent simulations were run using the particle mesh Ewald method within AMBER 4.1. In successive 10 ps steps, the positional constraint on the counterions were released, then the drug, before the system was heated to 300 K over 5 ps and held there for a further 5 ps, giving a total of 40 ps equilibration. The NOE restraints were then introduced gradually over 2.5 ps at 300 K by increasing the restraint force constant linearly from 0 to 32 kcal mol^{-1} $Å^{-2}$, followed by 1 ns of constant temperature dynamics. At the end of which the system was cooled to 1K over 2 ps, and held at 1K

for 5 ps, to generate a low energy structure that was subsequently energy minimised. All structures were displayed using MOLMOL (Koradi et al. 1996). The co-ordinates of the energy minimised structure of the nogalamycin-d(ATGCAT)$_2$ complex are deposited with the Protein Data Bank under accession code 1QCH.

Structural statistics: Total number of NOE restraints = 205; DNA-DNA restraints = 131; DNA-drug restraints = 23; drug-drug restraints = 51. RMS deviation from input restraints averaged over dynamics simulation = 0.18 Å (final 100 ps); number of restraint violations between 0.3–0.6 Å = 6; number of restraint violations > 0.5 Å = 1. RMSD from the mean structure over final 100 ps (± standard deviation) = 0.66 Å (± 0.12 Å). RMSD of energy minimised structure from ideal covalent geometry: bond lengths = 0.024 Å; bond angles = 2.74°; impropers: no torsion angles were identified with energy penalties > 5 kcal mol^{-1}.

3.5. Bibliography

Altona, C. and Sundaralingam, M. (1972) Conformational Analysis of the Sugar Ring in Nucleosides and Nucleotides. A New Description Using the Concept of Pseudorotation. *J. Am. Chem. Soc.* **94**, 8205-8212.

Baleja, J.D., Moult, J. and Sykes, B.D. (1990) Distance measurement and structural refinement with NOE data. *J. Mag. Res.* **87**, 375-384.

Bax, A. and Davis, D.G. (1985a) MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy. *J. Mag. Res.* **65**, 355-360.

Bax, A. and Davis, D.G. (1985b) Practical aspects of two-dimensional transverse NOE spectroscopy. *J. Mag. Res.* **63**, 207-213.

Bayly, C.I., Cieplak, P., Cornell, W.D. and Kollman, P.A. (1993) A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints For Determining Atom-Centered Charges: The RESP model. *J. Phys. Chem.* **97**, 10269

Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F. and Dinola, A. (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684-3690.

Bhuyan, B.K. and Dietz, A. (1965) Fermentation, Taxonomic, and Biological Studies of Nogalamycin. *Anti. Agents and Chem.* 836-844.

Bhuyan, B.K. and Reusser, F. (1970) Comparative biological activity of nogalamycin and its analogs. *Cancer Research* **30**, 984-989.

Bhuyan, B.K. and Smith, C.G. (1965) Differential Interaction of Nogalamycin with DNA of Varying Base Composition. *Proc. Nat. Acad. Sci. USA* **54**, 566-572.

Bothner-by, A.A., Stephens, R.L., Lee, J., Warren, C.D. and Jeanloz, R.W. (1984) Structure determination of a tetrasaccharide - transient nuclear Overhauser effects in the rotating frame. *J. Am. Chem. Soc.* **106**, 811-812.

Braunschweiler, L. and Ernst, R.R. (1983) Coherence transfer by isotropic mixing: application to proton correlation spectroscopy. *J. Mag. Res.* **53**, 521-528.

Chazin, W.J., Wüthrich, K., Rance, M., Hyberts, S., Denny, W.A. and Leupin, W. (1986) ^{1}H Nuclear Magnetic Resonance Assignments for d(GCATTAATGC)$_2$ Using Experimental Refinements of Established Procedures. *J. Mol. Biol.* **190**, 439-453.

Cheatham, T.E. and Kollman, P.A. (1996) Observation of the A-DNA to B-DNA transition during unrestrained molecular dynamics in aqueous solution. *J. Mol. Biol.* **259**, 434-444.

Cheatham, T.E., Miller, J.L., Fox, T., Darden, T.A. and 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.

Choudhury, K., Choudhury, I., Biswas, N. and Neogy, R.K. (1978) Interaction of nogalamycin with nucleic acids. *Indian J. Biochem. Biophys.* **15**, 373-376.

Collier, D.A., Neidle, S. and Brown, J.R. (1984) Mollecular Models for the Interaction of the Anti-Tumour Drug Nogalamycin with DNA. *Biochemical Pharmacology* **33**, 2877-2880.

Darden, T.A., York, D.M. and Pedersen, L.G. (1993) Particle Mesh Ewald - An N.LOG(N) Method for Ewald Sums In Large Systems. *J. Chem. Phys.* **98**, 10089-10092.

Diekmann, S. (1989) Definitions and nomenclature of nucleic acid structure parameters. *J. Mol. Biol.* **205**, 787-791.

Duan, Y., Wilkosz, P., Crowley, M. and Rosenberg, J.M. (1997) Molecular Dynamics simulation study of DNA Dodecamer d(CGCGAATTCGCG) in solution: Conformation and hydration. *J. Mol. Biol.* **272**, 553-572.

Egli, M., Williams, L.D., Frederick, C.A. and Rich, A. (1991) DNA Nogalamycin Interactions. *Biochem.* **30**, 1364-1372.

Embrey, K.J., Searle, M.S. and Craik, D.J. (1993) Interaction of Hoechst 33258 with the minor groove of the A+T-rich DNA duplex d(GGTAATTACC)2 studied in solution by NMR spectroscopy. *Eur. J. Biochem.* **211**, 437-447.

Ennis, H.L. (1981) Nogalamycin Inhibits Ribonucleic-Acid Synthesis in Growing and Developing Cells of the Slime-Mold Dictyostelium-Discoideum. *Anti. Agents and Chem.* **19**, 657-665.

Fisher, J.F. and Aristoff , P.A. (1988) *Prog. Drug Res.* **32**, 411

Fox, K.R., Brasset, C. and Waring, M.J. (1985) Kinetics of Dissociation of Noglamycin from DNA - Comparison With Other Anthracycline Antibiotics. *Biochim. Biophys. Acta* **840**, 383-392.

Fox, K.R. and Waring, M.J. (1984) Evidence of different binding sites for nogalamycin in DNA revealed by association kinetics. *Biochim. Biophys. Acta* **802**, 162-168.

Fox, K.R. and Waring, M.J. (1986a) Footprinting Reveals That Nogalamycin and Actinomycin Shuffle Between DNA-Binding Sites. *Nuc. Acids. Res.* **14**, 2001-2014.

Fox, K.R. and Waring, M.J. (1986b) Nucleotide-Sequence Binding Preferences of Nogalamycin Investigated by DNAse-I Footprinting. *Biochem.* **25**, 4349-4356.

Gao, Y.G., Liaw, Y.C., Robinson, H. and Wang, A.H.J. (1990) Binding of the Antitumor Drug Nogalamycin and its Derivatives to DNA - Structural Comparison. *Biochem.* **29**, 10307-10316.

Hunter, C.A. and Lu, X.-L. (1997) DNA Base-stacking Interactions: A comparison of theoretical calculations with oligonucleotide X-ray crystal structures. *J. Mol. Biol.* **265**, 603-619.

Kersten, W., Kersten, H. and Szybalski, W. (1966) Physicochemical Properties of Complexes Between Deoxyribonucleic Acid and Antibiotics Which Affect Ribonucleic Acid Synthesis (Actinomycin, Daunomycin, Cinerubin, Nogalamycin, Chromomycin, Mithramycin, Olivomycin). *Biochem.* **5**, 263-244.

Koradi, R., Billeter, M. and Wüthrich, K. (1996) MOLMOL. *J. Mol. Graph.* **14**, 51-55.

Lane, A.N. and Searle, M.S. (1992) ^{31}P NMR investigation of the backbone conformation and dynamics of the hexamer duplex d(5'-GCATGC)2 in its complex with the antibiotic nogalamycin. *FEBS Letters* **297**, 292-296.

Curves 51-helical analysis of irregular nucleic acids. Lavery, R. and Sklenar, H. (1996) Laboratoire de Biochimie Theorique, CNRS URA 77, Institur de Brologie Physico-Chimique, Paris.:

Lee, H., Darden, T.A. and Pedersen, L.G. (1995) Accurate crystal-molecular dynamics simulations using particle mesh Ewald - RNA dinucleotides ApU and GpC. *J. Chem. Phys.* **243**, 229-235.

Li, L.H., Kuentzel, S.L., Murch, L.L., Pschigoda, L.M. and Krueger, W.C. (1979) Comparative biological and biochemical effects of nogalamycin and its analogs on L1210 leukemia. *Cancer Research* **39**, 4816-4822.

Liaw, Y.C., Gao, Y.G., Robinson, H., van der Marel, G.A., van Boom, J.H. and Wang, A.H.J. (1989) Antitumor Drug Nogalamycin Binds DNA in Both Grooves Simultaneously - Molecular-Structure of Nogalamycin DNA Complex. *Biochem.* **28**, 9913-9919.

Marion, D. and Wüthrich, K. (1983) Application of phase sensitive two-dimensional correlated spectroscopy (COSY) for measurements of H-1-H-1 spin-spin coupling- constants in proteins. *Biochem. Biophys. Res. Comm.* **113**, 967-974.

Packer, M. and Hunter, C.A. (1998) Sequence-dependent DNA Structure: The Role of the Sugar-phosphate Backbone. *J. Mol. Biol.* **280**, 407-420.

Amber 4.1. Pearlman, D.A., Case, D.A., Caldwell, J.W., Ross, W.S., Cheatham, T.E., Fergusson, D.M., Seibel, G.L., Singh, U.C., Weiner, P.K. and Kollman, P.A. (1995) 4.1. University of California, San Francisco:

Piantini, U., Sorensen, O.W. and Ernst, R.R. (1982) Multiple quantum filters for elucidating NMR coupling networks. *J. Am. Chem. Soc.* **104**, 6800-6801.

Piotto, M., Saudek, V. and Sklenar, V. (1992) Gradient-tailored excitation for single quantum NMR spectroscopy of aqueous solutions. *J. Bio. Mol. NMR* **2**, 661-665.

Porschke, D. (1978) *Biopolymers* **17**, 315-325.

Prive, G.G., Heinemann, U., Chandrasegaran, S., Kan, L.S., Kopka, M.L. and Dickerson, R.E. (1987) Helix Geometry, Hydration, and G.A Mismatch in a B-DNA Decamer. *Science* **238**, 498-504.

Rance, M., Sorensen, O.W., Bodenhausen, G., Wangner, G., Ernst, R.R. and Wüthrich, K. (1983) Improved spectral resolution in COSY H-1-NMR spectra of proteins via double quantum filtering. *Biochem. Biophys. Res. Comm.* **117**, 479-485.

Rinkel, L.J. and Altona, C. (1987) Conformational Analysis of the Deoxyribofuranose Ring in DNA by Means of Sums of Proton-proton Coupling Constants: A Graphical Method. *J. Biomol. Struct. Dyna.* **4**, 621-649.

Robinson, H., Liaw, Y.C., van der Marel, G.A., van Boom, J.H. and Wang, A.H.J. (1990) NMR-Studies on the Binding of Antitumor Drug Nogalamycin to DNA Hexamer d(CGTACG). *Nuc. Acids. Res.* **18**, 4851-4858.

Schuerman, G.S., Smith, C.K., Turkenburg, J.P., Dettmar, A.N. and Vanmeervelt, L. (1996) DNA-Drug Refinement - A Comparison of the Programs NUCLSQ, PROLSQ, SHELXL93 and X-PLOR, Using the Low-Temperature d(TGATCA)-Nogalamycin Structure. *ACTA Crysta. D* **52**, 299-314.

Searle, M.S. and Bicknell, W. (1992) Interaction of the anthracycline antibiotic nogalamycin with the hexamer duplex d(5'-GACGTC)$_2$ - An NMR and molecular modelling study. *Eur. J. Biochem.* **205**, 45-58.

Searle, M.S., Hall, J.G, Denny, W.A. and Wakelin, L.P.G (1988) NMR-Studies of the Interaction of the Antibiotic Nogalamycin With the Hexadeoxyribonucleotide Duplex d(5'-GCATGC)$_2$. *Biochem.* **27**, 4340-4349.

Searle, M.S. and Wakelin, L.P.G (1990) Conformation and dynamics of the deoxyribose rings of a (nogalamycin)$_2$-d(5'-GCATGC)$_2$ complex studied in solution by 1H-n.m.r. spectroscopy. *Biochem. J.* **269**, 341-346.

Smith, C.K., Brannigan, J.A. and Moore, M.H. (1996) Factors affecting DNA sequence selectivity of nogalamycin intercalation: the crystal structure of d(TGTACA)$_2$-Nogalamycin$_2$. *J. Mol. Biol.* **263**, 237-258.

Smith, C.K., Davies, G.J., Dodson, E.J. and Moore, M.H. (1995) DNA-nogalamycin interaction: The crystal structure of d(TGATCA) complexed with nogalamycin. *Biochem.* **34**, 415-425.

States, D.J., Haberkorn, R.A. and Ruben, D.J. (1982) A two-dimensional nuclear Overhauser experiment with pure absorption phase in 4 quadrants. *J. Mag. Res.* **48**, 286-292.

van Houte, L.P.A., van Garderen, C.J. and Patel, D.J. (1993) The antitumour drug nogalamycin forms two different intercalation complexes with d(GCGT)d(ACGC). *Biochem.* **32**, 1667-1674.

SPARTAN 3.1. Wavefunction, I. (1995) 3.1. Waveform Inc., USA.

Williams, H.E. and Searle, M.S. (1998) Structure of the nogalamycin-d(ATGCAT)$_2$ complex in solution: DNA recognition at an isolated TpG site. *J. Chem. Soc. Perkin. Trans. 1* 3-5.

Williams, L.D., Egli, M., Gao, Q., Bash, P., van der Marel, G.A., van Boom, J.H., Rich, A. and Frederick, C.A. (1990) Structure of Nogalamycin Bound to a DNA Hexamer. *Proc. Nat. Acad. Sci. USA* **87**, 2225-2229.

York, D.M., Darden, T.A. and Pedersen, L.G. (1993) The Effect of Long-Range Electrostatic Interactions in Simulations of Macromolecular Crystals - a Comparison of the Ewald and Truncated List Methods. *J. Chem. Phys.* **99**, 8345-8348.

York, D.M., Yang, W.T., Lee, H., Darden, T.A. and Pedersen, L.G. (1995) Toward the Accurate Modeling of DNA - The Importance of Long-Range Electrostatics. *J. Am. Chem. Soc.* **117**, 5001-5002.

Young, M.A., Ravishanker, G. and Beveridge, D.L. (1997) A 5-nanosecond molecular dynamics trajectory for B-DNA: Analysis of structure, motions, and solvation. *Biophys. J.* **73**, 2313-2336.

Zhang, X.L. and Patel, D.J. (1990) Solution Structure of the Nogalamycin-DNA Complex. *Biochem.* **29**, 9451-9466.

Chapter 4. Hydration of the d(ATGCAT)$_2$-Nogalamycin Complex Determination by Restrained Molecular Dynamics

4.1. Introduction: The role of Hydration in Macromolecular Structure

A variety of experimental techniques have shown that nucleic acids are extensively hydrated in both the crystal and in solution, and that solvation plays an important part in determining the overall conformation (Berman, 1994; Liepinsh et al. 1994; Eisenstein and Shakked, 1995; Kochoyan and Leroy, 1995; Egli et al. 1996; Johannesson and Halle, 1998). Low water activity is known to induce a B-DNA to A-DNA transition that has been rationalised in terms of optimising interactions with water molecules (Drew and Dickerson, 1981; Prive et al. 1987; Cheatham and Kollman, 1996). The concentration of water in aqueous solution is sufficiently high that all solvent accessible sites in DNA are close to a water molecule for most of the time. Based on estimates of diffusion-limited association and dissociation rate constants, site-occupancy at the surface of a solute particle in the absence of any interaction between solvent and solute has been calculated to be > 96%, with an average residence time at 10°C of ~0.5 ns (Conte et al. 1996). Thus, water can be considered to be 'bound' if the residence time of a water molecule is significantly longer than 0.5 ns (Otting et *al.*, 1991; (Kubinec and Wemmer, 1992; Otting and Liepinsh, 1995; Jacobson et al. 1996; Conte et al. 1996). Reports of DNA and RNA hydration studies by NMR have identified water molecules with residency times >0.5 ns that have been proposed to play a thermodynamic role in

structure stabilisation (Kubinec and Wemmer, 1992; Liepinsh et al. 1992; Liepinsh et al. 1994; Fawthrop et al. 1993) and in mediating drug-DNA interactions in the minor groove (Jenkins and Lane, 1997; Lane et al. 1997; Lane et al. 1997).

Two X-ray structures have described the binding of nogalamycin at high affinity 5'-TpG sites at the ends of the hexamer duplexes d(T•GTAC•A)$_2$, and d(T•GATC•A)$_2$ (• represents the position of the bound antibiotic) (Smith et al. 1996; Smith et al. 1995; Schuerman et al. 1996). Both of these structures are extensively hydrated, with evidence for many solvent-mediated interactions between drug and DNA primarily in the major groove. In particular, the N,N-dimethylammonium ion exhibits solvent mediated interactions with the edges of several DNA bases, rather than with the negatively charged sugar-phosphate backbone, suggesting water-mediated sequence recognition. The DNA major and minor grooves are also extensively hydrated with evidence of chains of water molecules lying along the grooves. This static picture of the degree of hydration of the complex suggests an important role for the solvent, however, dynamic information is required to discriminate between different types of water molecules with different residence times at the solute-solvent interface.

Water molecule interactions with solute can broadly be divided into three main categories. Bulk water is defined as the solvent that is not influenced by the solute, this water makes up the matrix in which the molecules are solvated. This passive description of water is perhaps the more traditional way of thinking of solvent. The remaining types of water are clasified as strongly and weakly associated hydration (Figure 4-1).

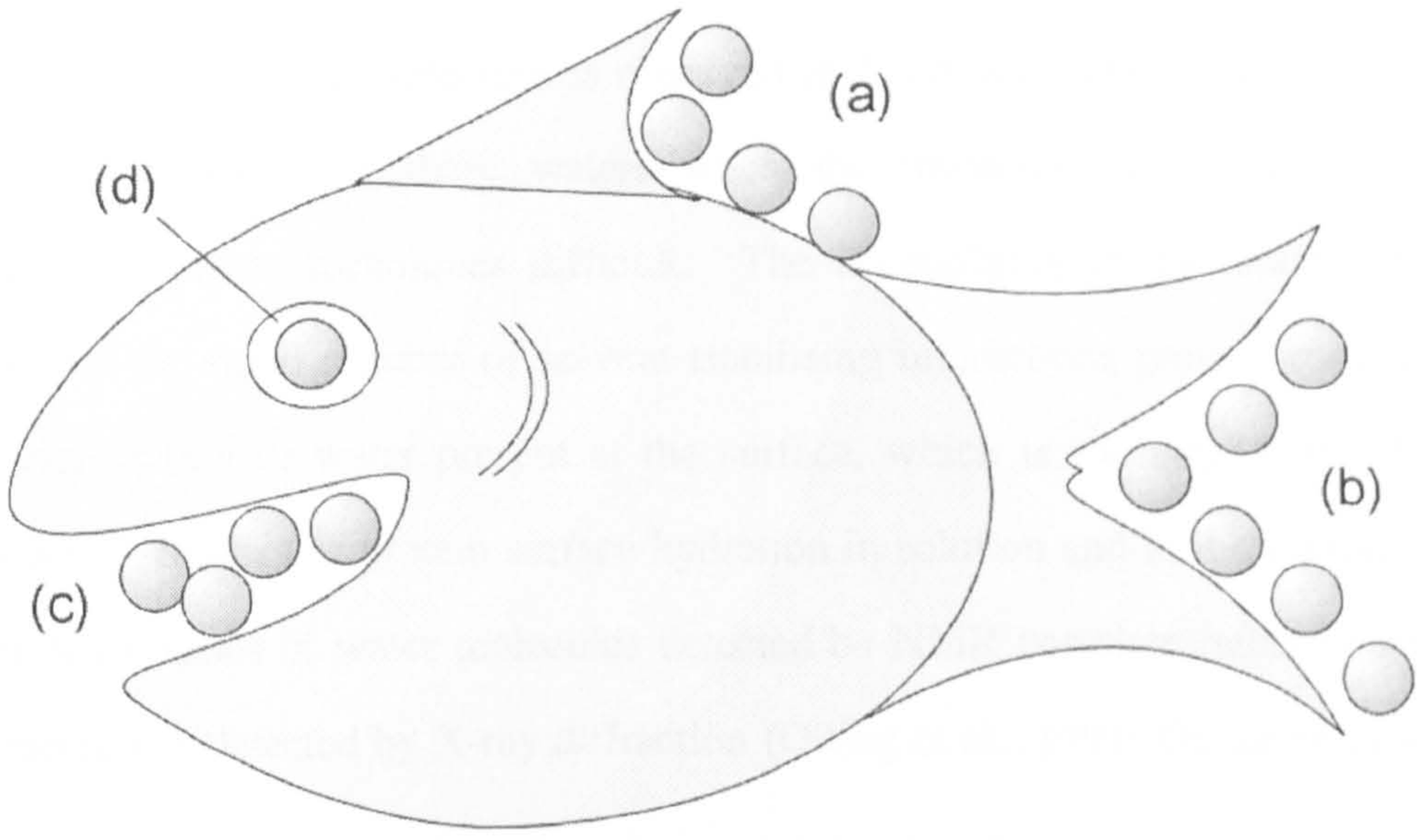


Figure 4-1: Types of water environments encountered. Surface water (a) and water associated with shallow grooves (b) are able to exchange with bulk solvent freely unless interactions restrict this exchange. Extremely narrow grooves hinder exchange with bulk solvent and increase residence time (c). Internal water molecules are isolated from bulk solvent and remain internalised while the solute remains folded (d)

A schematic representing possible water environments is shown in Figure 4-1. The accessibility of the solvent to bulk gives surface water molecules short residence times (Figure 4-1 (a)). Such a situation is also observed in wide grooves if no stabilising interactions are present (Figure 4-1 (b)). Long residence time water is often internalised within proteins (Figure 4-1 (d)). These water molecules are isolated from bulk solvent and are usually observed in the core of proteins. Long residence time waters that have access to bulk solvent have been observed in structures with very narrow grooves such as the dihydrofolatereductase-methotrexate-NADPH complexes (Gerothanassis et al. 1992). Such a situation is depicted (Figure 4-1 (c)) where residence time is increased due to the restriction on solvent exchange with bulk.

Surface hydration of two molecules is observed in X-ray diffraction structures, but the short residence times of these waters when the molecule is in solution makes observation by other techniques difficult. The accessibility of the surface to bulk solvent, and the small number of solvent-stabilising interactions present results in the high concentration of water present at the surface, which is exchanging rapidly with bulk solvent. Studies of protein surface hydration in solution and in the crystal suggest that residence times of water molecules detected by NMR correlate rather poorly with water molecules detected by X-ray diffraction (Otting et al. 1991; Otting and Liepinsh, 1995).

In contrast to surface hydration, the detection of a long- lived spine of hydration within DNA A-tract structures appears to be in good agreement with X-ray studies (Prive et al. 1987; Schneider et al. 1993; Kubinec and Wemmer, 1992; Liepinsh et al. 1994; Liepinsh et al. 1992; Fawthrop et al. 1993; Lane et al. 1997). The experiments conducted on the Dickerson dodecamer, d(CGCGAATTCGCG)$_2$, (Liepinsh et al. 1992) show the hydration environment of the major groove around the 5-CH_3 of the thymines to posses a short residence time. The short residence time hydration environment is also observed to exist around base H6 and H8 protons, which also project into the major groove. In contrast, the solvent environment around the AH2s, that project into the minor groove, is significantly more stable with residence times > 1 ns being reported. Although the groove is accessible to solvent the 'ordered' water structures appear to persist. Although these molecules can be classed as 'surface' water molecules, the residence times are atypical of hydration in such an environment. Surface water molecules present in other systems such as proteins only possess very short residence times. This contrast supports the theory that water molecules within the AT tract perform some structural role that in turn stabilises the hydration shell. This

NMR evidence is consistent with hydration spines that are observed in these A-tract sequences in X-ray structures.

Recent calculations with DNA duplexes and triplexes have shown that integration of water populations over the duration of the dynamics simulation can lead to the identification of water density in the grooves which is clearly greater than in bulk water. This is indicative of a negative free energy of transfer from bulk solvent to the sites of DNA hydration (Duan et al. 1997; Young et al. 1997; Shields et al. 1998; Shields et al. 1997). Relatively few simulations of solvent structure around DNA have been reported, and even fewer involve drug complexed DNA fragments.

Within this and the following chapter we show that X-ray analysis, NMR methods and dynamics simulations are complementary techniques that provide different perspectives on the important role that water molecules have in mediating the nogalamycin-DNA interaction.

4.2. Investigation of Hydration by Restrained Molecular Dynamics

The use of explicit solvent models allows us to probe elements of hydration structure that are difficult to detect in solution. All solvent molecules present in the dynamics simulation use the TIP3P parameterisation within Amber 4.1. The 3-point water model developed at Purdue University has been shown to reproduce experimentally observed parameters of water such as density and co-ordination (Jorgensen et al. 1983). TIP3P solvent molecules are parameterised with a three centred electrostatic potential localised

on the atomic co-ordinates and an approximated Leonard-Jones potential localised on the oxygen atoms. Analysis of the radial distribution functions of solvent alone shows TIP3P reproduces the first solvation shell accurately up to distances of 3.6 Å. This validates the use of the TIP3P model for solvent interactions. Dynamics simulations allow solvent molecules free motion around the periodic box with only minimal approximations to the solvent force field, and these are mainly restricted to the internal stretch and bending motions of the water molecule. Thus hydration sites might be revealed if water molecules occupy a particular site near a solute for a period of time. This is a vague concept that needs refinement since with two thousand water molecules present in a periodic box, each able to move around within the system, how can sites be identified and what is the definition of significant occupancy?

4.3. Calculation of Water Occupancy

Although molecular motions are simulated on the femtosecond time scale co-ordinate information about the system is only sampled at half picosecond intervals for reasons of data volume. It is possible to use these snap-shots of the molecular dynamics to spot trends in water occupancy. Each snap shot must be normalised to remove rotational and transitional motion thus allowing all solute molecules and associated solvent molecules to be superimposed. Control studies looking at errors in processing give about 0.5Å^3 misplacement of atoms due to this normalising procedure. Sections of space surrounding the solute can then be divided up into a grid of cubes typically with a volume set to 1Å^3. Water density around the solute can be calculated for each cube position by summing the water molecules present in each cube. This can be repcated

for each snap shot and a cumulative total for each spatial position can thus be calculated (Figure 4-2).

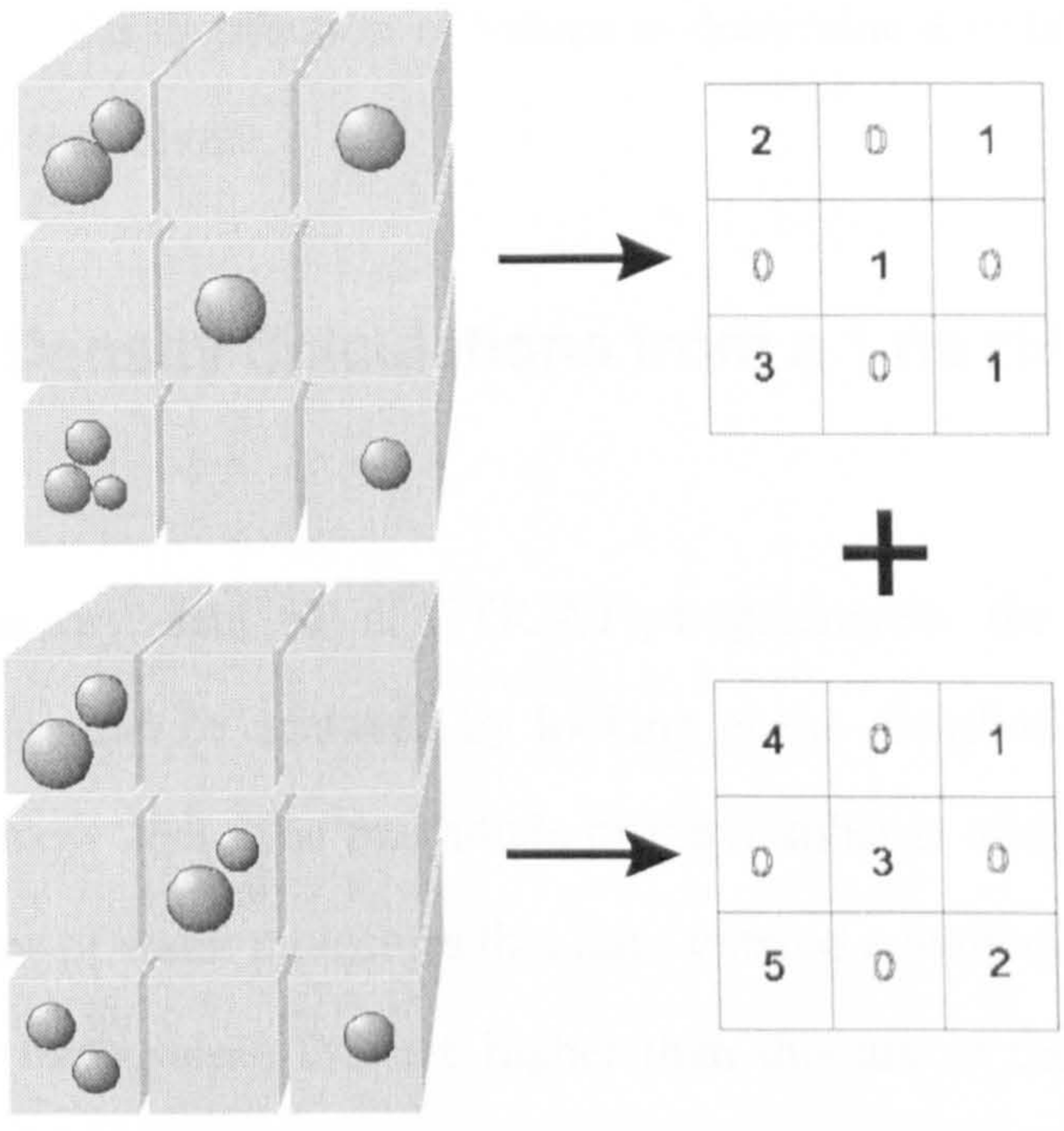


Figure 4-2: Schematic of the method employed for determining water occupancy from dynamics simulations: The entire system is divided into a number of cubes of equal volume. The number of water molecules in each cube is summed for each snap-shot giving a cumulative total for each cube. The final total is a measure of occupancy for each region of space within the dynamics simulation.

We therefore have a numerical measure of site occupancy but we need to define significant occupancy. A graphical method is to plot a histogram showing the number of sites with specific occupancies. The number of cubes with a given cumulative total is plotted in Figure 4-3. With a box dimension of 60x45x45 Å, there would be ca.

120000 individual cubes. With the motion of approximately 2500 water molecules, it would be expected that each volume of space would have a water molecule present for some period of a 1ns dynamics. The discrimination of background occupancy can be achieved by looking at the distribution of values to determine a suitable threshold that disregards the bulk of the solvent.

4.4. Results of Density Calculations from a 1 ns rMD Trajectory

From the 1ns trajectory data of d(ATGCAT)$_2$-nogalamycin the removal of this background occupancy can be achieved by looking at the distribution and discarding cubes with values below 286. The magnitude of the number is arbitrary being just the total number of atoms of water molecules that have entered a particular region of space, however cubes that have values that are higher than this are in the minority and are therefore different from the majority of bulk solvent sites. The threshold is arbitrary but a value that selects only 2% of the total number of cells gives sufficient occupancy information but avoids general background clutter.

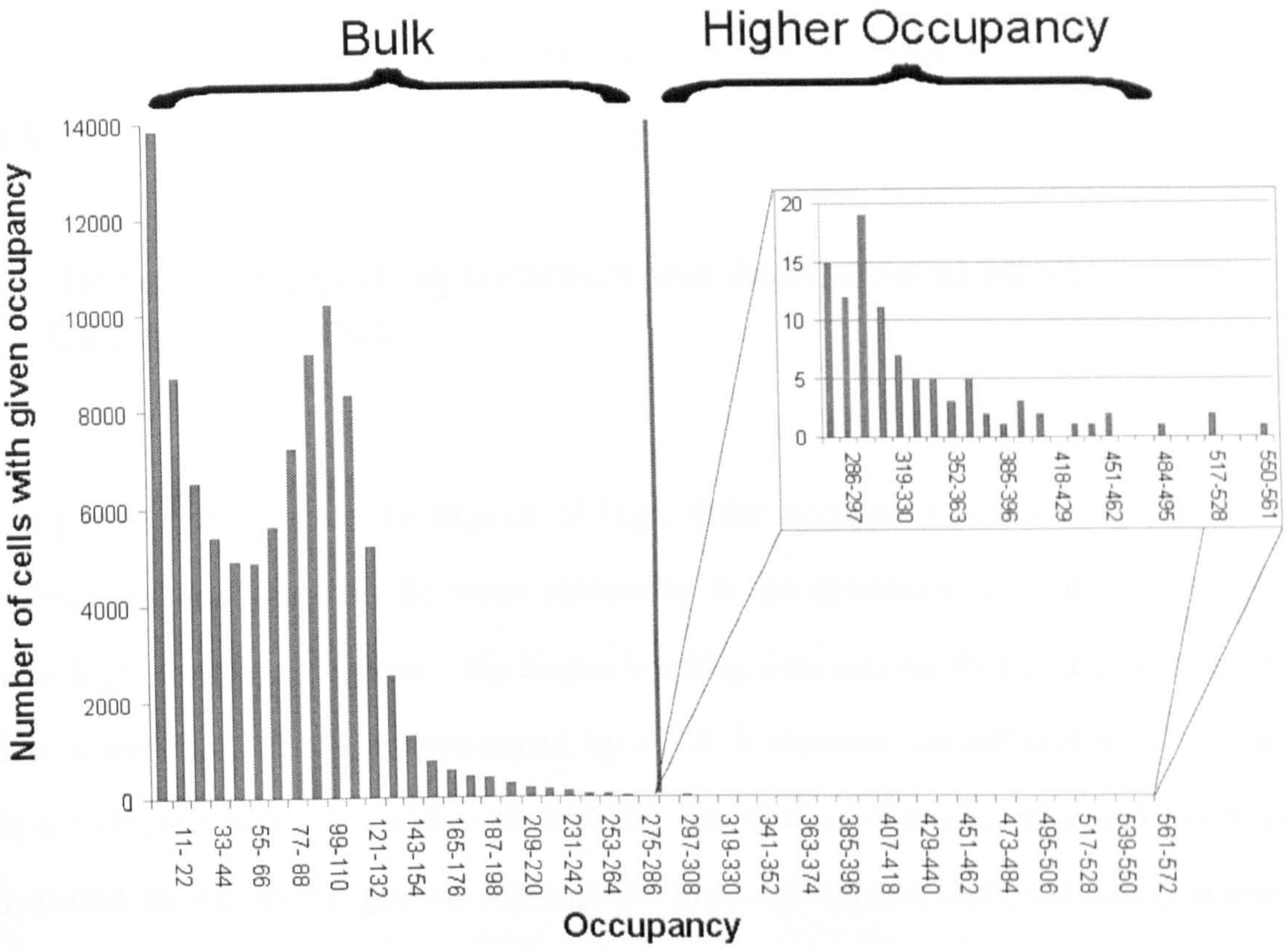


Figure 4-3: Number of cubes from the occupancy analysis with a cumulative total. The majority, 98%, of the cubes have occupancies less than 286 units. Cells below horizontal line account for 98% of all observations. This graph defines significant occupancy.

One ns of MD simulation data at 300K was analysed by the method described to identify hydration sites in the nogalamycin-d(ATGCAT)$_2$ complex. The analysis highlights 64 cells (inset in Figure 4-3) with high occupancy and these can be displayed overlaid with the average DNA structure. High occupancy hydration sites are observed mainly in the major groove, particularly in close proximity to the charged N,N-dimethylammonium group of the nogalamycin ligand. Further hydration is observed within the minor groove located around the 4'-OCH_3 of the nogalose sugar and around

the C10 methyl ester on the A ring. Hydration is also observed around the phosphate backbone of the duplex.

4.5. Specific H-bonding Interactions Available at High Occupancy Sites

It is possible to examine the regions of high water occupancy revealed by the above analysis and highlight specific water molecules in the dynamics calculation that are in these high occupancy regions. Hydrogen bonding interactions that could be made by these water molecules was determined by a 3.8 Å distance cut-off and a 34° angular cut-off (Figure 4-4). A number of bridging interactions between drug and DNA are identified in the major groove binding the α-D-3,6-dideoxy-3(N,N-dimethylamino)-glucose. The 2''-OH and the quinone on ring C provide examples of solvent mediated interactions. Further bridging roles are also identified between the N7 of guanine-3 and the drug 2''-OH, and the cytosine-10 N4 amino group and the drug 4''-OH. Dynamics simulation suggests that these bridging interactions are not static but transiently formed. Indeed we also see evidence for direct interactions between the 2''-OH and the N7 of guanine. A fluctuating behaviour between direct and solvent mediated interactions is suggested over the time course of the dynamics simulation. The solvent environment allows a greater flexibility in the nature of the interactions stabilising the complex either through direct interactions or through solvent mediated effects.

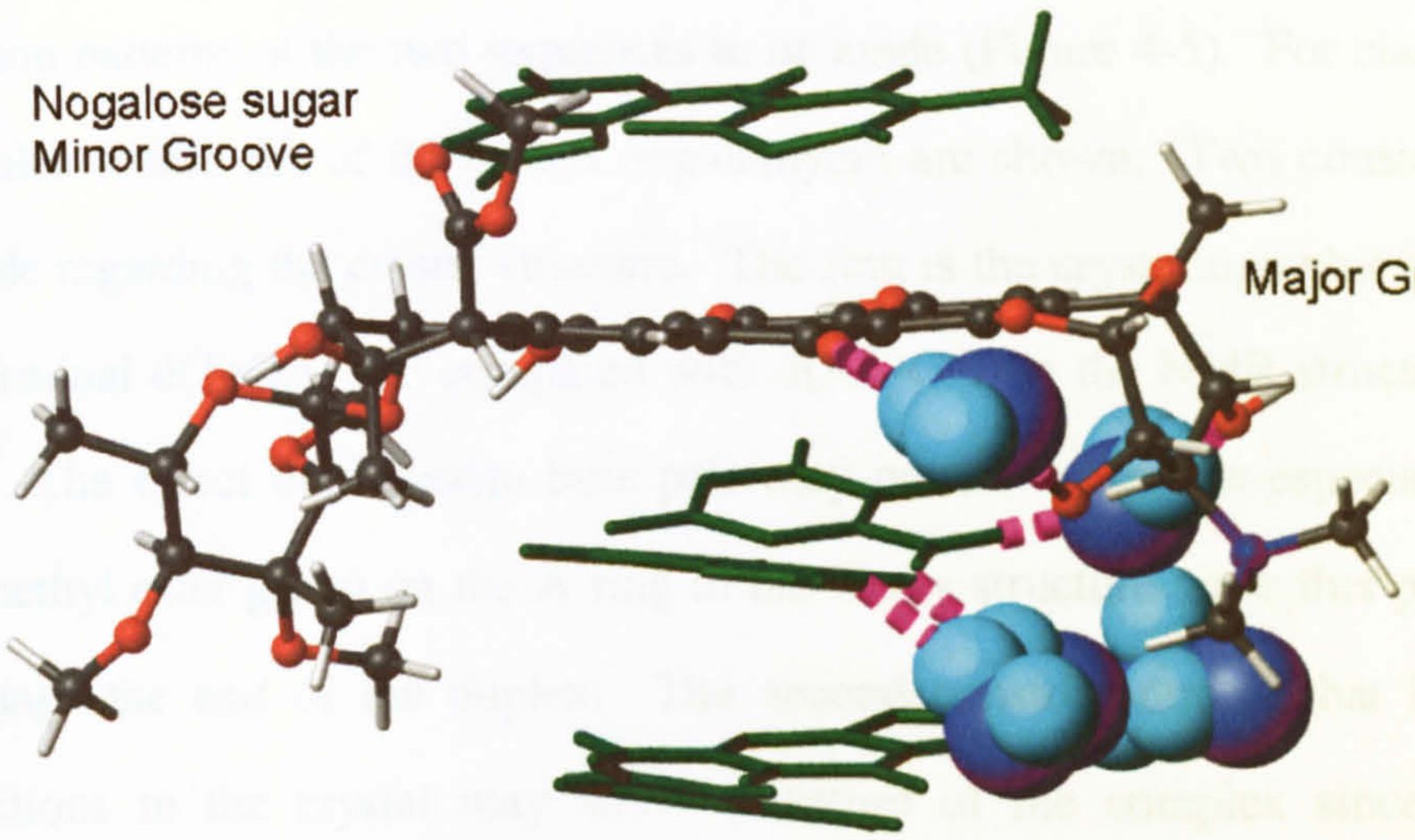


Figure 4-4: Water interactions that are present around the α-D-3,6-dideoxy-3(N,N-dimethylamino)-glucose of nogalamycin from one snap-shot of dynamics. Nogalamycin shown with atomic colouring, is intercalated through a dTpG step with the major groove binding α-D-3,6-dideoxy-3(N,N-dimethylamino)-glucose sugar oriented to the right. Three DNA basepair steps are shown in green, with the sugars and phosphate backbone removed for clarity. Water molecules, represented by blue spacefilled spheres, show a number of mediating intermolecular and intramolecular interactions.

4.6. Comparisons of rMD with X-ray Data

X-ray crystal structures deposited with the Protein data bank for the d(TGATCA)$_2$-Nogalamycin$_2$ complex were studied to assess the nature of hydration in the crystal structure with nogalamycin bound at a TpG site. Alignment of the heavy atoms of nogalamycin molecules allowed superimposition of the TG-bound nogalamycin in the X-ray structure with the corresponding portion of the d(ATGCAT)$_2$

complex determined here by NMR. This allows a graphical comparison of the hydration patterns of the two sequences to be made (Figure 4-5). For clarity only water molecules within 5Å of the bound nogalamycin are shown. Two considerations must be made regarding the crystal structure. The first is the crystallographic nogalamycin is in a terminal d(T•GA) site compared with d(AT•GC) in the NMR structure (• = bound drug). The effect of the extra base pair may perturb hydration especially around the C10 methyl ester group on the A ring in the X-ray structure since this part of the drug overhangs the end of the duplex. The second consideration is that inter-molecular interactions in the crystal may affect solvation of the complex since a number of duplexes are present in the unit cell and crystal packing of duplexes excludes hydration. This second effect should be minimal within the 5Å cut-off used around the drug binding site.

Both structures show similarity between the major groove hydration with extensive solvent occupancy surrounding the charged N,N-dimethylammonium group of the α-D-3,6-dideoxy-3(N,N-dimethylamino)-glucose sugar. Hydration in the minor groove also shows a number of similarities particularly around the 4'-OCH_3 of the nogalose sugar in the minor groove. The structures only differ in their hydration pattern around the C10-methyl ester group. Within the X-ray structure the nogalose sugar occupies a site with increased solvent access, however in the NMR structure the C10-Methyl ester is located further along the minor groove and is partially buried within it, resulting in a different solvent pattern.

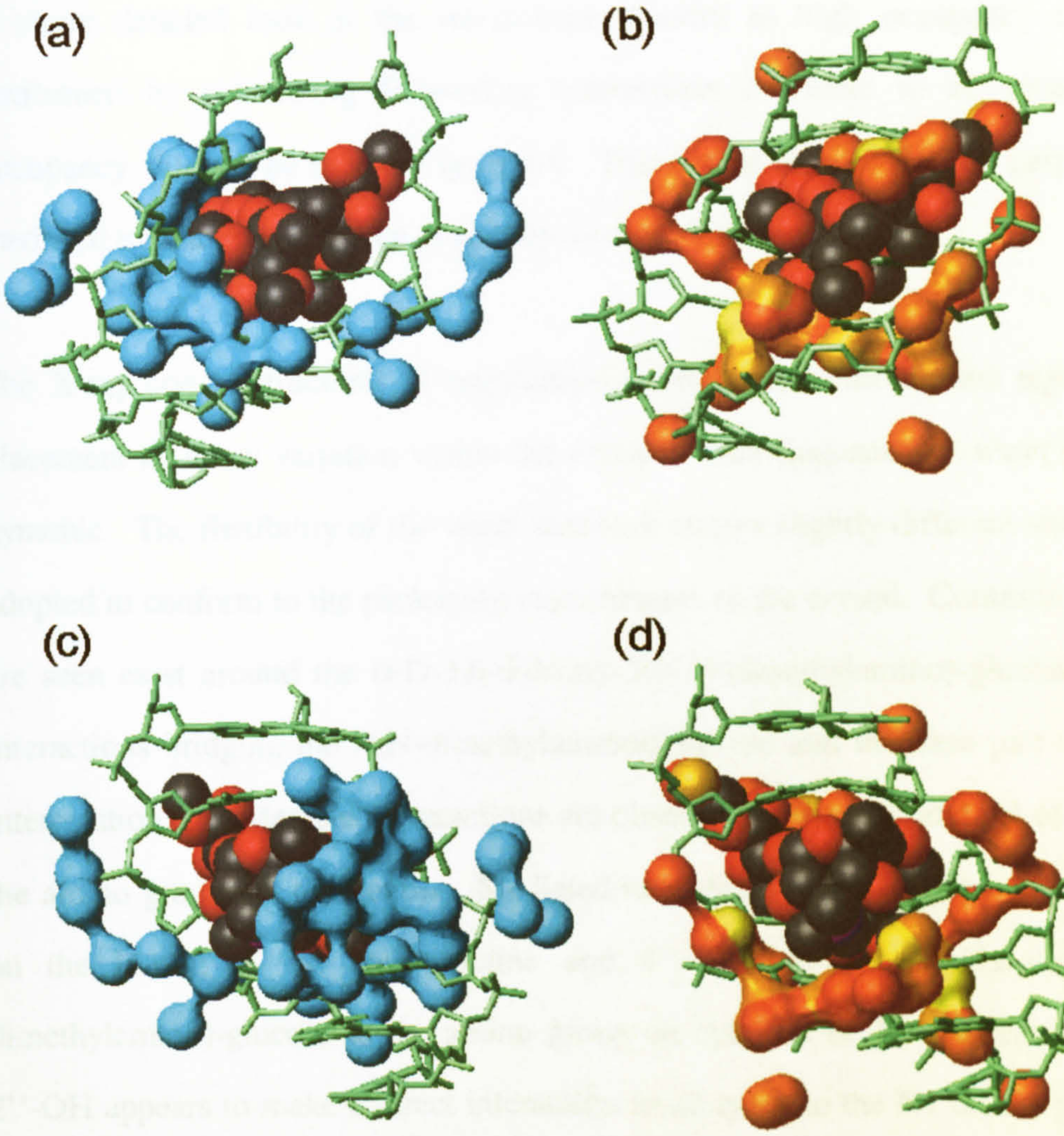
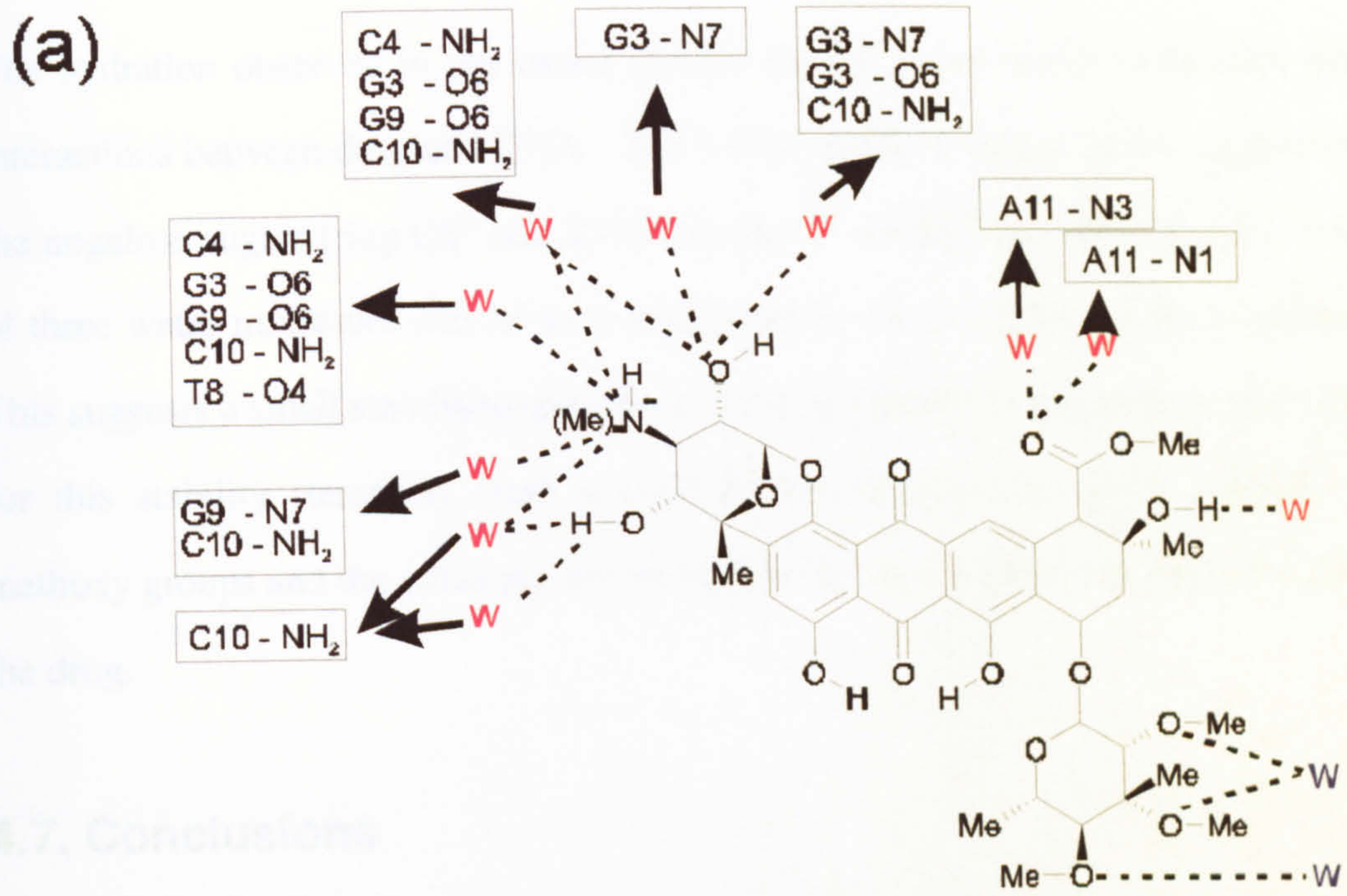


Figure 4-5: Comparison of MD calculated and X-ray determined water for the d(ATGCAT)$_2$-nogalamycin complex. Models shown are energy minimised models of the d(ATGCAT)$_2$-nogalamycin complex with the nogalamycin intercalator shown in spacefill. Views are into the minor groove (a) and (b), and the major groove (c) and (d). The positions of high occupancy water calculated from rMD calculations are shown in (b) and (d) with a graduated scale yellow representing highest occupancy. The hydration taken from an X-ray structure of nogalamycin$_2$-(TGATCA)$_2$ that was overlaid with the d(ATGCAT)$_2$ complex is shown in (a) and (c). Similar patterns of hydration are seen in both complexes with extensive hydration present in the major groove around the protonated N,N-dimethylammonium group. Similar hydration patterns are also observed around the phosphate backbone and in the minor groove around the 4'-OCH$_3$ of the nogalose sugar.

A more detailed look at the interactions present at high occupancy sites can be performed by calculating H-bonding interactions available to all centres of high occupancy, this can be seen in Figure 4-6. This makes for a useful comparison of water mediated interactions between nogalamycin complexes.

The X-ray crystal structures of nogalamycin that contain information regarding water placement all show variation within the crystal. This suggests that water placement is dynamic. The flexibility of the water structure allows slightly different structures to be adopted to conform to the packaging requirements of the crystal. Common features that are seen exist around the α-D-3,6-dideoxy-3(N,N-dimethylamino)-glucose with water interactions bridging the N,N-dimethylammonium ion and the base pair step after the intercalation site. Mediated interactions are observed to the carbonyl O4 of thymine and the amino group N7 of adenine. Mediated interactions are also seen around the 4-OH on the D-ring of the anthracycline and 4''-OH on the α-D-3,6-dideoxy-3(N,N-dimethylamino)-glucose to the amino group on cytosine at the intercalation site. The 2''-OH appears to make a direct interaction in all cases to the N7 on the guanine at the intercalation site within the X-ray structures although in some instances there is a co-ordinated water molecule at this site. The 6-OH on ring B makes no solvent contact in any of the crystal structures and this is confirmed by the present NMR analysis of the exchange rates (see section 5.3).



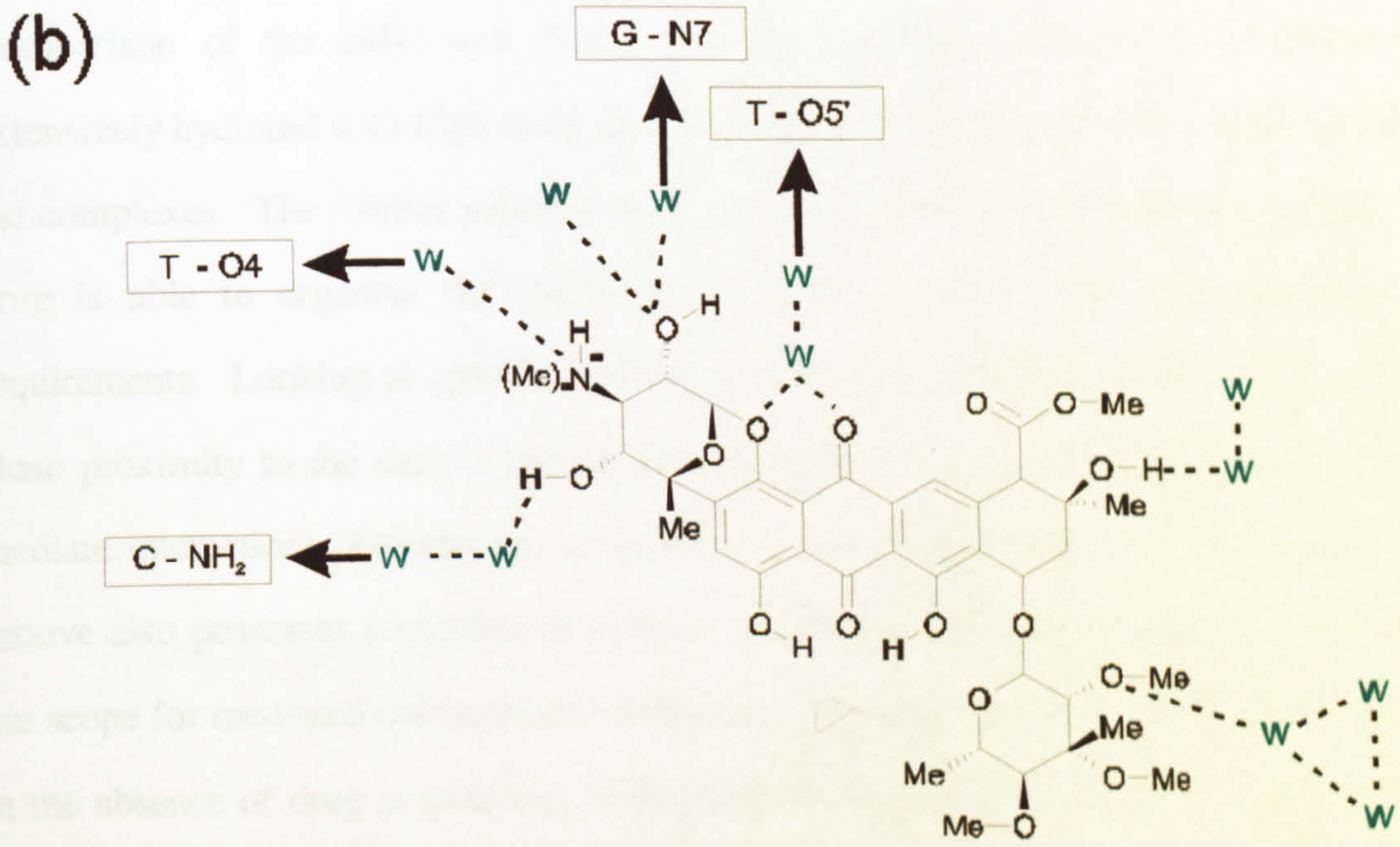


Figure 4-6: Water and water mediated interactions as observed in the X-ray structure (nogalamycin)$_2$-d(TGTACA)$_2$ (b) and the rMD calculated structure nogalamycin-d(ATGCAT)2 (a). water molecules are depicted by the letter 'W'. Electron density attributed to water molecules found in the X-ray structure are shown with mediated interactions to DNA protons represented by a box. Water molecules for the rMD structure were determined by detecting interactions available to water molecules if they were centred at the high occupancy positions.

The hydration observed in the minor groove shows fewer water molecules mediating interactions between drug and DNA. The 9-OH on the A-ring and the oxygen atoms on the nogalose sugars (ring O1' and 2'–OCH_3 and 3'–OCH_3) are all hydrated. A network of three water molecules can be seen associated to the 2'-OCH_3 of the nogalose sugar. This suggests a small stabilising pocket of water is formed in this region, with the origin for this stability stemming from a combination of the lone pairs available on the methoxy groups and the small pocket formed in the minor groove by the intercalation of the drug.

4.7. Conclusions

Comparison of the rMD and X-ray structures revealed that both structures are extensively hydrated with high occupancy hydration being located to discreet regions of the complexes. The similar patterns observed between different sequences suggest the drug is able to organise the hydration in order to satisfy the hydrogen bonding requirements. Looking at specific interactions that are available to water molecules in close proximity to the drug it can be seen that the major groove hydration is able to mediate interactions between key drug protons and nearby DNA protons. The minor groove also possesses a number of protons associated with water molecules, however, the scope for mediated interactions is reduced. The water structure in the minor groove in the absence of drug is generally considered transient in non-AT tract sequences due to a lack of stabilising interactions. The occurrence of water located around the nogalose sugar could be stabilised by pockets formed at the drug-DNA interface as well as through interactions with the lone pairs of the nogalose -OCH_3 groups.

4.8. Bibliography

Berman, H.M. (1994) Hydration of DNA: Take 2. *Cur. Op. Struct. Biol.* **4**, 345-350.

Cheatham, T.E. and Kollman, P.A. (1996) Observation of the A-DNA to B-DNA transition during unrestrained molecular dynamics in aqueous solution. *J. Mol. Biol.* **259**, 434-444.

Conte, M.R., Conn, G.L., Brown, T. and Lane, A.N. (1996) Hydration of the RNA duplex r(CGCAAAUUUGCG)$_2$ determined by NMR. *Nuc. Acids. Res.* **24**, 3693-3699.

Drew, H.R. and Dickerson, R.E. (1981) Structure of a B-DNA Dodecamer .3. Geometry of Hydration. *J. Mol. Biol.* **151**, 535-556.

Duan, Y., Wilkosz, P., Crowley, M. and Rosenberg, J.M. (1997) Molecular Dynamics simulation study of DNA Dodecamer d(CGCGAATTCGCG) in solution: Conformation and hydration. *J. Mol. Biol.* **272**, 553-572.

Egli, M., Portman, S. and Usman, N. (1996) RNA hydration: a detailed look. *Biochem.* **35**, 8489-8494.

Eisenstein, M. and Shakked, Z. (1995) Hydration patterns and intermolecular interactions in A-DNA crystal structures. Implications for DNA recognition. *J. Mol. Biol.* **248**, 662-678.

Fawthrop, S.A., Yang, J.C. and Fisher, J. (1993) Structural and dynamic studies of a non-self-complementary dodecamer DNA duplex. *Nuc. Acids. Res.* **21**, 4860-4866.

Gerothanassis, I.P., Birdsall, B., Bauer, C.J., Frenkiel, T.A. and Feeney, J. (1992) Nuclear Magnetic Resonance Detection of Bound Water Molecules in the Active Site of *Lactobacillus casei* Dihydrofolate Reductase in Aqueous Solution. *J. Mol. Biol.* **226**, 549-554.

Jacobson, A., Leupin, W., Liepinsh, E. and Otting, G. (1996) Minor groove hydration of DNA in aqueous solution: sequence-dependent next neighbor effect of the hydration lifetimes in d(TTAA)$_2$ segments measured by NMR spectroscopy. *Nuc. Acids. Res.* **24**, 2911-2918.

Jenkins, T.C. and Lane, A.N. (1997) AT selectivity and DNA minor groove binding: Modelling, NMR and structural studies of the interactions of propamidine and pentamidine with d(CGCGAATTCGCG)$_2$. *Biochim. Biophys. Acta* **1350**, 189-204.

Johannesson, H. and Halle, B. (1998) Minor Groove hydration of DNA in solution: Dependence on base composition and sequence. *J. Am. Chem. Soc.* **120**, 6859-6870.

Jorgensen, W.L., Chanderasekhar, J., Madura, J.D., Impey, R.W. and Klein, M.L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926-935.

Kochoyan, M. and Leroy, J.L. (1995) Hydration and Solution Structure of Nucleic Acids. *Cur. Op. Struct. Biol.* **5**, 329-333.

Kubinec, M.G. and Wemmer, D.E. (1992) NMR Evidence for DNA Bound Water in Solution. *J. Am. Chem. Soc.* **114**, 8739-8740.

Lane, A.N., Jenkins, T.C. and Frenkiel, T.A. (1997) Hydration and solution structure of d(CGCAAATTTGCG)$_2$ and its complex with propamidine from NMR and molecular modelling. *Biochim. Biophys. Acta* **1350**, 205-220.

Liepinsh, E., Leupin, W. and Otting, G. (1994) Hydration of DNA in Aqueous-Solution - NMR Evidence for a Kinetic Destabilization of the Minor-Groove Hydration of d-(TTAA)$_2$ Versus d-(AATT)$_2$ Segments. *Nuc. Acids. Res.* **22**, 2249-2254.

Liepinsh, E., Otting, G. and Wüthrich, K. (1992) NMR-Spectroscopy of Hydroxyl Protons in Aqueous-Solutions of Peptides and Proteins. *J. Bio. Mol. NMR* **2**, 447-465.

Otting, G. and Liepinsh, E. (1995) Selective Excitation of Intense Solvent Signals in the Presence of Radiation Damping. *J. Bio. Mol. NMR* **5**, 420-426.

Otting, G., Liepinsh, E. and Wüthrich, K. (1991) Protein Hydration in Aqueous Solution. *Science* **254**, 974-980.

Prive, G.G., Heinemann, U., Chandrasegaran, S., Kan, L.S., Kopka, M.L. and Dickerson, R.E. (1987) Helix Geometry, Hydration, and G.A Mismatch in a B-DNA Decamer. *Science* **238**, 498-504.

Schneider, B., Cohen, D.M., Schleifer, L., Svrinivasan, A.R. and Olson, W.K. (1993) A systematic method for studying the spatial distribution of water molecules around nucleic acid bases. *Biophys. J.* **65**, 2291-2303.

Schuerman, G.S., Smith, C.K., Turkenburg, J.P., Dettmar, A.N. and Vanmeervelt, L. (1996) DNA-Drug Refinement - A Comparison of the Programs NUCLSQ, PROLSQ, SHELXL93 and X-PLOR, Using the Low-Temperature d(TGATCA)-Nogalamycin Structure. *ACTA Crysta. D* **52**, 299-314.

Shields, G.C., Laughton, C.A. and Orozco, M. (1997) Molecular dynamics simulations of the d(TAT) triple helix. *J. Am. Chem. Soc.* **119**, 7463-7469.

Shields, G.C., Laughton, C.A. and Orozco, M. (1998) Molecular dynamics simulation of a PNADNAPNA triple helix in aqueous solution. *J. Am. Chem. Soc.* **120**, 5895-5904.

Smith, C.K., Brannigan, J.A. and Moore, M.H. (1996) Factors affecting DNA sequence selectivity of nogalamycin intercalation: the crystal structure of d(TGTACA)$_2$-Nogalamycin$_2$. *J. Mol. Biol.* **263**, 237-258.

Smith, C.K., Davies, G.J., Dodson, E.J. and Moore, M.H. (1995) DNA-nogalamycin interaction: The crystal structure of d(TGATCA) complexed with nogalamycin. *Biochem.* **34**, 415-425.

Young, M.A., Ravishanker, G. and Beveridge, D.L. (1997) A 5-nanosecond molecular dynamics trajectory for B-DNA: Analysis of structure, motions, and solvation. *Biophys. J.* **73**, 2313-2336.

Chapter 5. Water Residency as Revealed by NMR

5.1. Introduction

NMR offers a unique insight into the nature of the interaction between a solute and its solvent environment. It allows the experimenter to detect solute protons close to solvent molecules as well as to distinguish between short and long residence time waters based on the signs and magnitude of the observed NOE and ROE.

Both the NOESY and ROESY 2D experiments are in widespread use for structure determination. The principle reason behind this is that these experiments give through space information regarding proton separation. The intensity of the cross peaks observed is proportional to the inter-proton separation and this is governed by the cross relaxation rate constant for the NOESY, σ^{NOE}, and ROESY, σ^{ROE}, respectively (Equation 10). Where $J(\omega)$ represents the spectral density function for a frequency ω.

$$\sigma^{NOE} = 6J(2\omega_0) - J(0)$$
$$\sigma^{ROE} = 3J(\omega_0) + 2J(0)$$

Equation 10: Relaxation rate constant for laboratory (NOE) and rotating (ROE) reference frames. Both values are dependant upon the spectral density function J().

For studies into hydration it can be safely assumed that all solvent exposed sites will be close to a water molecule. This is a statistical probability when a typical H_2O

concentration of 55M is used (pure water as solvent). This means the distance dependent term holds little information. A more detailed look at the factors governing relaxation of ROEs and NOEs shows a dependence upon molecular motions. The spectral density function for a given frequency is dependent upon a correlation time, τ_c. Correlation time is an intrinsic property of a molecule and based upon the timescale of motions that a molecule undergoes. Long correlation times correspond to slow motions and rapid motions are described by short correlation times.

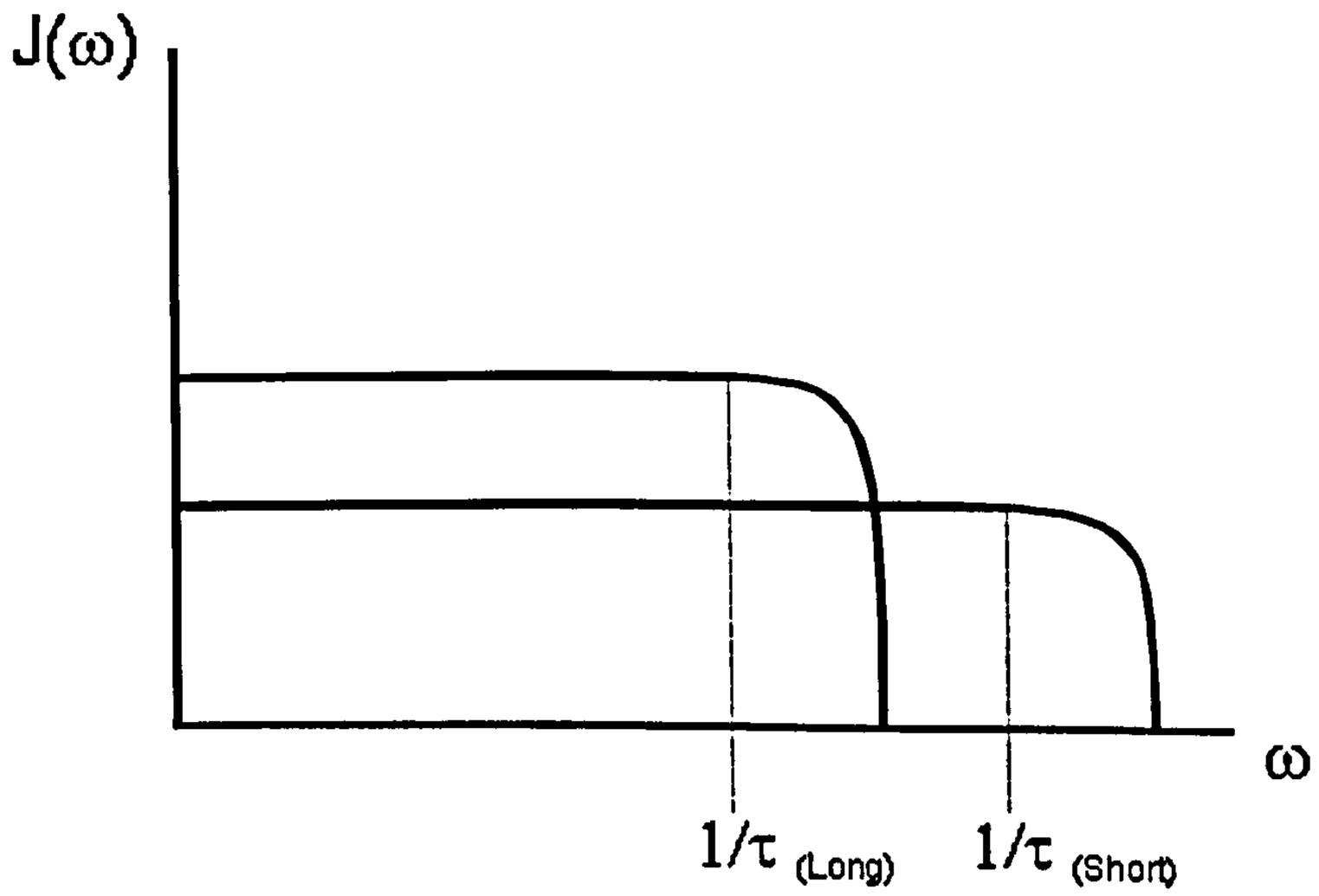


Figure 5-1: Spectral Density vs frequency. As the correlation time becomes longer the spectral density increases.

Figure 5-1 shows a general case for an oscillating frequency (J(ω)) for an arbitrarily short and long τ_c. The area under the graph is fixed therefore the effect of correlation time is to vary the upper limit. This has an effect on the overall relaxation rate T1, since only fluctuations of the correct frequency can stimulate relaxation. As the spectral density for a molecule with a relaxation frequency ω decreases, the relaxation rate

decreases, since fewer oscillations of the correct frequency are available. The effect for large biological molecules is that as the correlation time gets longer the molecules orientate more slowly so the nuclei relax faster. This has implications for relaxation pathways for ROESY and NOESY experiments. The two relaxation pathways available to NOE and ROE are the W_2 and W_0. The higher frequency W_2 pathway generates an NOE enhancement, if W_2 dominates this would manifest as a negative peak in a NOESY spectrum. The W_0 pathway by comparison produces a negative NOE enhancement. The relative proportion of these two pathways affects the overall observed NOE. The increase in correlation time causes the contribution from the higher frequency relaxation pathway to reduce relative to the W_0 pathway. This explains the sign change from positive NOE enhancement to negative NOE as the size of the molecule increases and hence correlation time becomes longer.

The dependence of the spectral density (J) on the correlation time, and the effect on the observable NOE, allows measurements to be made that reveal the subtle differences in molecular correlation times. The precise values of $J(\omega)$ depends on how correlation times are mapped to specific molecular motions. The general case that large molecules such as DNA tumble more slowly than small molecules such as water, holds, assuming that there is no interaction between the two. If a water molecule becomes 'attached' to the macromolecule what would the outcome be? Such a situation would be of interest since some stabilising interaction between solute and solvent would be occurring, and the structural role could be speculated.

$$J(\omega) = \tau/(1+\omega^2\tau^2)$$

Equation 11: Dependence of the spectral density function $J(\omega)$ upon effective correlation time, τ.

Three models of explaining molecular motion in terms of correlation time and thus spectral density have been suggested. The simplest model assumes water molecules in the hydration shell are part of the biological molecule. The correlation time, τ refers to the overall correlation time of the solute and its hydration shell (Equation 11). A graphical representation of this can be seen in Figure 5-2a. Such a model suggests one overall correlation time for both solute and solvent atoms. This model fails to correlate with experimental data where there is variation in the observed NOESY and ROESY spectra, with different nuclei showing different relaxation rates.

The "wobble in a cone" model (Fujiwara and Nagayama, 1985) allows for variability of correlation time by introducing an internal correlation time for solvent molecules. This model ignores any effect due to translational motion and assumes there is a tight binding between solvent and solute and this generates only positive peaks for σ^{NOE}. The final model assumes independent translation and rotation of a freely diffuseable solvent. A process that describes the behaviour of bulk solvent and a weakly associated solvent. The value of $J(\omega)$ becomes dependent upon the diffusion co-efficients of the molecules concerned.

$$\tau = \overline{x^2}/D$$

Equation 12: Einstein-Smoluchowski equation relating correlation time τ with rate of diffusion. Where D is the diffusion constant and X is a RMS displacement that defines the distance travelled to constitute exchange.

Positive values for σ^{NOE} can be obtained with diffusion coefficients of greater than $3x10^{-6}$ cm^2/s. If exchange between bulk solvent and a solute hydration site is defined as an RMS displacement of 4Å from the hydration site, positive σ^{NOE} values are obtained

for lifetimes less than 500ps based on the Einstein-Smoluchowski equation (Equation 12). Hence, the presence of a positive σ^{NOE} indicates a rapid exchange between bulk water and a solute hydration site.

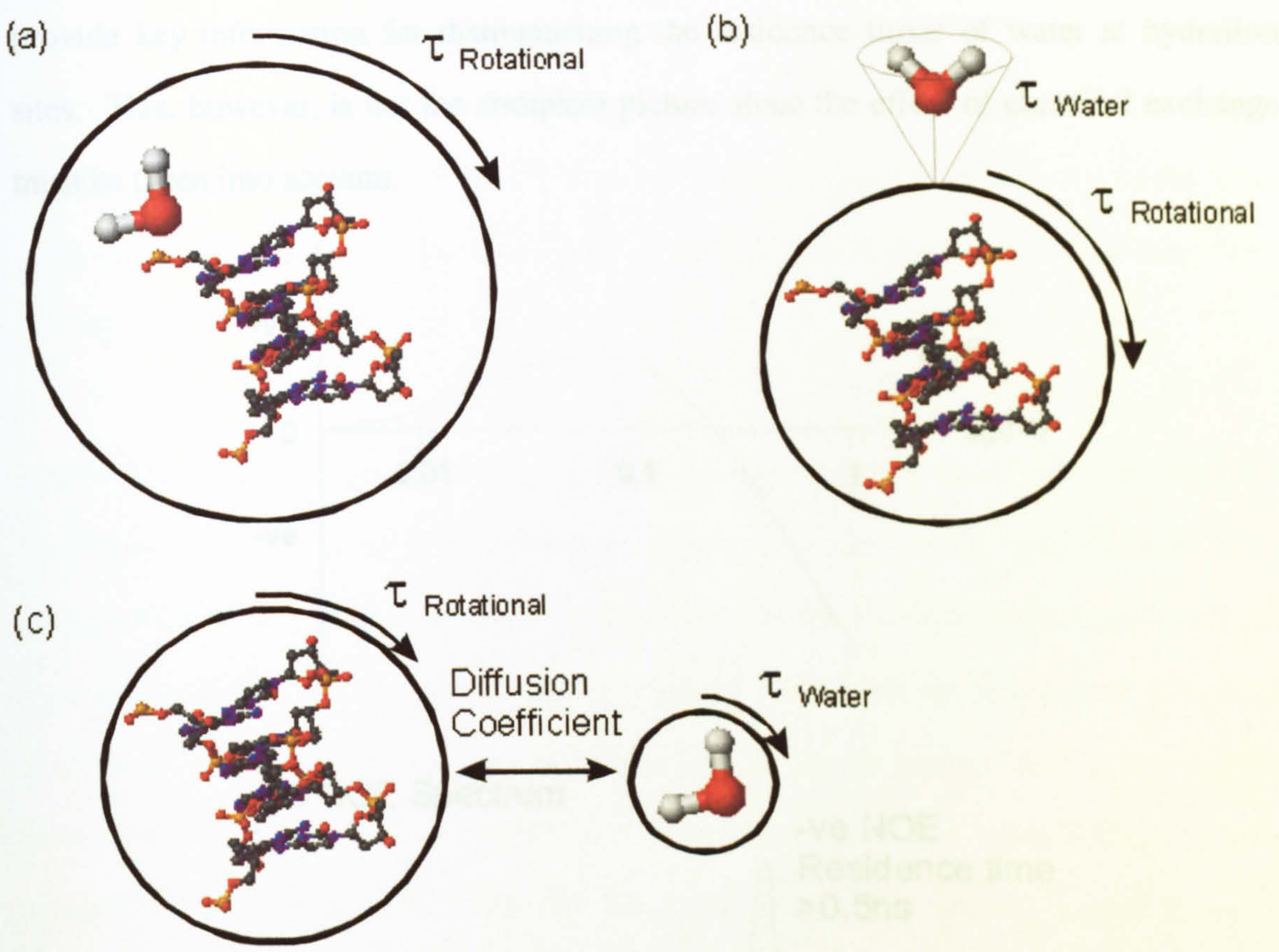


Figure 5-2: Models describing the relationship between correlation time and spectral density function. Isotropic rotational diffusion model, water molecules and solute molecules are assumed to rotate with an overall correlation time (a). Wobbling in cone model where water molecules are treated as associated but able to move within the limits of a cone giving rise to two separate correlation times (b). Random independent transitional and rotation model where molecules are free to re-orientate with independent correlation times and translational movements are summarised by the diffusion coefficient (c).

The two models of rigidly bound solvent with high internal correlation time and free diffusion define the two limiting cases. Both of these models always give positive

values for the cross-relaxation rate constant in the rotating reference frame (σ^{ROE}). However the laboratory reference frame experiment undergoes a sign change from positive at short correlation times to negative at longer correlation times and thus provide key information for distinguishing the residence times of water at hydration sites. This, however, is not the complete picture since the effect of chemical exchange must be taken into account.

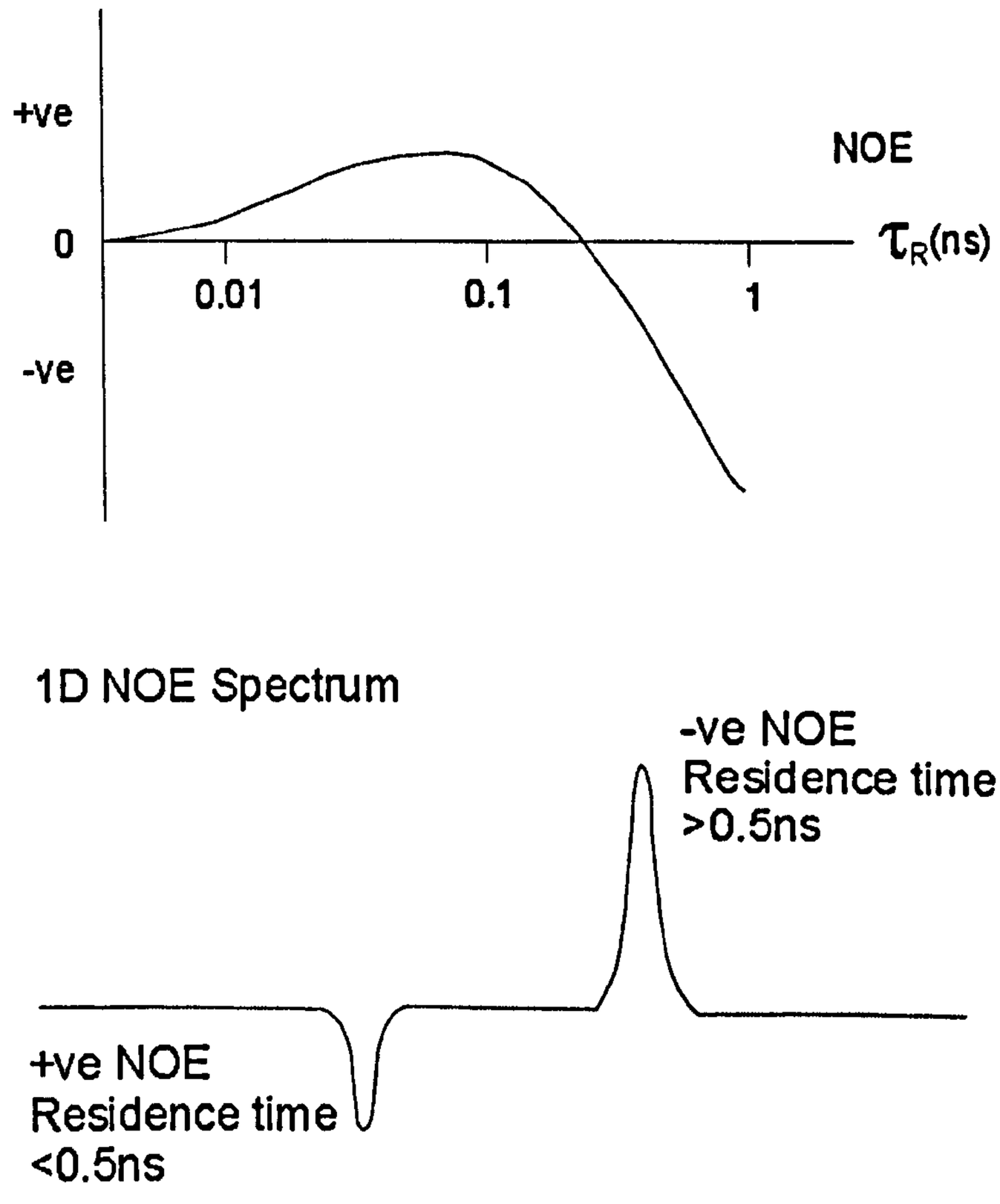


Figure 5-3: Graph of NOE dependence on correlation time. With the corresponding cross peak generated in a NOESY spectrum shown below. By convention positive NOE give rise to negative peaks.

5.2. Complication from Chemical Exchange Processes

The general cases can be considered by taking three protons representing non-exchangeable (H_N), exchangeable (H_E) and water (H_W) protons. The N-E-W model is a theoretical model that has been validated by experimental techniques. The cross-sections through ROESY and NOESY at the H_2O frequency reveal both interactions between water and exchangeable protons (H_W-H_E), and water and non-exchangeable (H_W-H_N).

Figure 5-4 shows the possible interactions a proton can experience and the predicted NMR peaks generated by the interaction. Figure 5-4(a,b) are the general cases for a non-exchangeable proton interacting with solvent. In the first case a NOE generated by a trapped water molecule causes a strong positive NOE peak to be observed (H_N-H_W). The other effect is between a non-exchangeable solute proton and a less rigidly associated water proton. Due to the short correlation time of a water molecule direct interactions between a freely diffusable water molecule and a non-exchangeable proton would be expected to give rise to negative NOE cross peak and a negative ROE cross peak (Figure 5-4 (b)). The presence of a positive NOE peak would also be expected to when chemical exchange is taking place (Figure 5-4 (d)). It is possible to distinguish the two scenarios by a positive sign ROE cross peak.

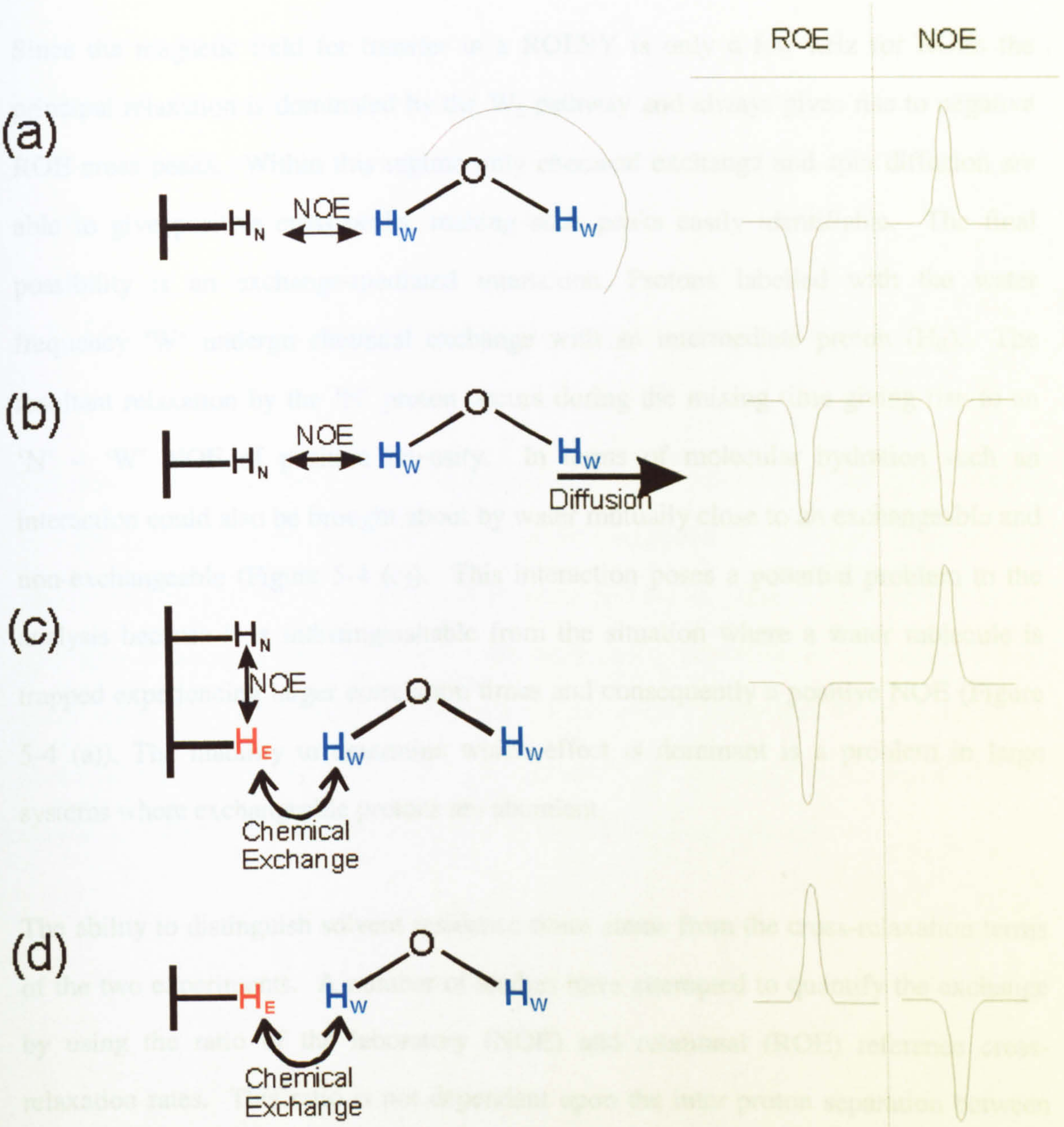


Figure 5-4:Diagram of the N-E-W model showing potential mechanisms for the generation of solute proton – water proton NOEs and ROEs. Where H_N and H_E represents non-labile and exchangeable solute protons, respectively, and H_W represents a water proton. The last two columns show the resultant ROE and NOE. Four potential mechanisms are shown. A non-labile proton interacting with a trapped or restricted water molecule (a), a non-labile proton interacting with a freely diffusable water proton (b). A non-labile proton interacting with an exchangeable proton which aquires a water frequency label via exchange during the mixing time (c). An exchangeable proton undergoing chemical exchange with a water proton (d). With the exception of (a) and (c) all other mechanisms give rise to a unique pattern of NOE and ROE.

Since the magnetic field for transfer in a ROESY is only a few kHz for ROEs the principal relaxation is dominated by the W_2 pathway and always gives rise to negative ROE cross peaks. Within this regime only chemical exchange and spin diffusion are able to give positive cross-peaks, making such peaks easily identifiable. The final possibility is an exchange-mediated interaction. Protons labelled with the water frequency 'W' undergo chemical exchange with an intermediate proton (H_E). The resultant relaxation by the 'N' proton occurs during the mixing time giving rise to an 'N' – 'W' NOE of positive intensity. In terms of molecular hydration such an interaction could also be brought about by water mutually close to an exchangeable and non-exchangeable (Figure 5-4 (c)). This interaction poses a potential problem to the analysis because it is indistinguishable from the situation where a water molecule is trapped experiencing larger correlation times and consequently a positive NOE (Figure 5-4 (a)). The inability to determine which effect is dominant is a problem in large systems where exchangeable protons are abundant.

The ability to distinguish solvent residence times stems from the cross-relaxation terms of the two experiments. A number of studies have attempted to quantify the exchange by using the ratio of the laboratory (NOE) and rotational (ROE) reference cross-relaxation rates. This ratio is not dependant upon the inter proton separation between solvent and solute and thus defines the residence times of the water molecules around the solute hydration sites. This is simpler to evaluate when little overlap is present, however, the difference in peak shapes makes an accurate determination of the ratio difficult. Further, the ROESY experiment utilises a sustained radio frequency pulse (a spin-lock) to create an environment in which nuclei interact in a different way to that of the NOESY. The use of such a spin locking pulse in the ROESY experiment results in distortions to integrals that vary across the spectrum. This also leads to further errors in

evaluating ROESY cross-peak intensity and volume. Due to the inherent problems with the accurate determination of intensities a more qualitative approach has been adopted based solely on the sign change and a semi-quantitative evaluation of the intensity of the σ^{ROE} and σ^{NOE} as strong, medium or weak.

5.3. Effect of drug intercalation on exchange rates

The relative exchange rates for the base pairs within the complex can be approximated by examination of a one dimensional experiment recorded in 90% H_2O (Figure 5-5). The imino protons present on the N3 of thymine and N1 of guanine are involved in the hydrogen bonding of the DNA duplex. Exchange of a proton on an intermediate time scale results in a broadening of the peak. Line widths therefore reveal information about the relative stability of the base pairings since the imino protons are shielded from solvent and therefore exchange when in a Watson-Crick base pairing.

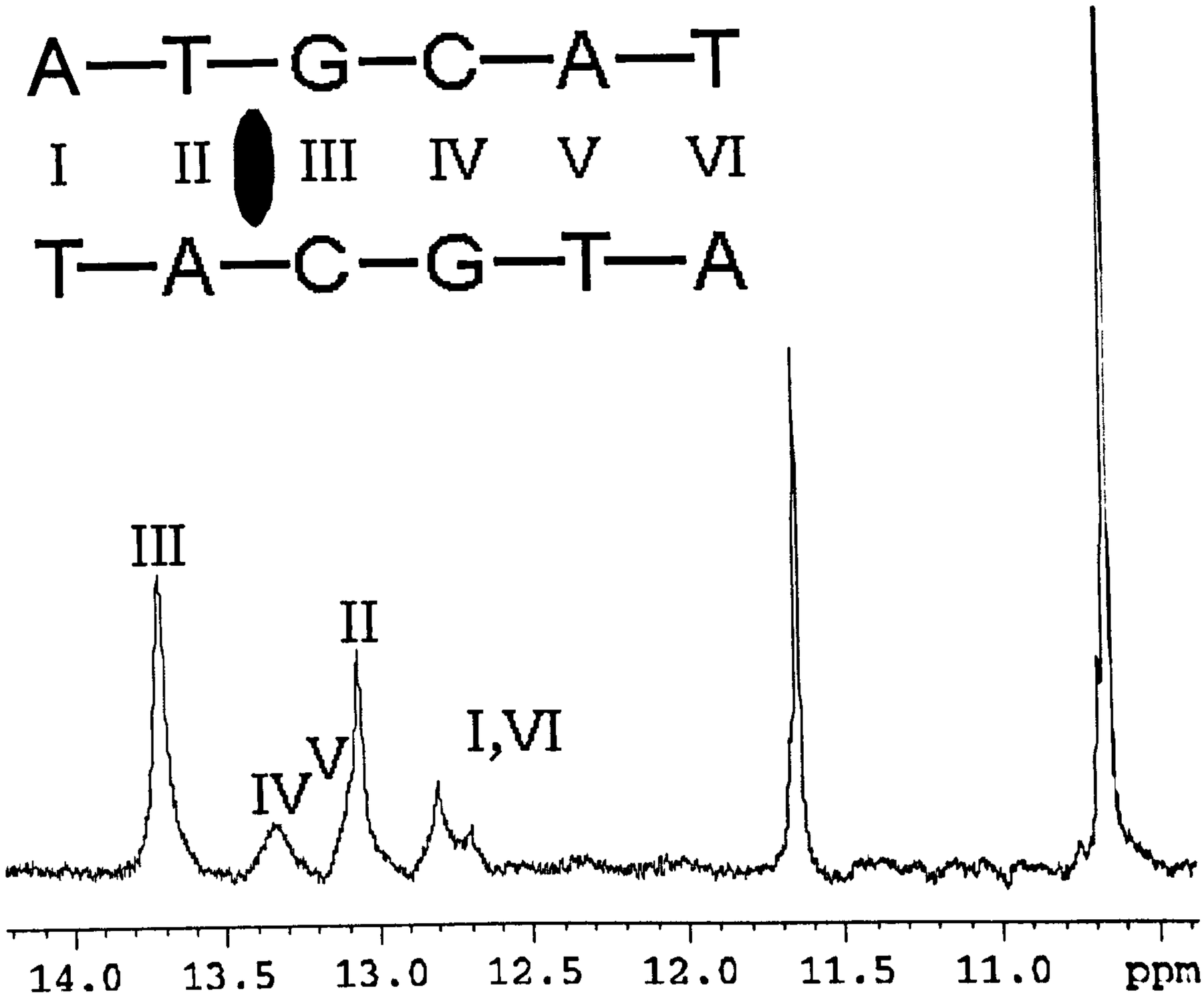


Figure 5-5: One dimensional ^1H NMR recorded at 288K showing the imino protons and phenolic resonances from the $d(ATGCAT)_2$ – nogalamycin complex. Resonances assigned to the thymine and guanine basepairs are labelled I – VI. The two signals at 11.7 and 10.7 ppm are assigned to the 4-OH and 6-OH respectively.

The two terminal dA-dT basepair steps (Figure 5-5, peaks I and IV) have comparatively large line widths consistent with the bases unpairing. This is to be expected in short oligonucleotides where fraying of the terminal base pairs occurs. The downfield signal at 13.7 ppm is attributed to the dG_3-dC_{10} base pair. The peak strongly suggests that this site is protected from exchange by the intercalating drug, this is consistent with the model. A similar picture is portrayed by the dT_2-dA_{11} base pair at 13.1ppm. Despite disruption to this base pair by the intercalator significant protection from exchange is afforded by the anthracycline. The final assessment of the exchange rates have been

estimated in the order of $G_3 < T_2 < G_9 < T_8 < T_6 \sim T_{12}$ consistent with a stabilisation of base pairs at the drug intercalation site. The two exchangeable phenolic protons (6-OH and 4-OH) on the aglycone ring system of nogalamycin show a difference in exchange characteristics. The upfield peak of 6-OH at 10.7ppm shows a sharper more intense peak than the 4-OH. ROESY cross-section data between the phenolic protons and water (data not shown) also shows no evidence for exchange for the 6-OH. This is consistent with our structural model that suggests this proton is shielded from solvent access in the minor groove by the hydrophobic nogalose sugar.

5.4. Hydration analysis in the nogalamycin – $d(ATGCAT)_2$ complex

Data for σ^{NOE} and σ^{ROE} can be obtained by a projection from the 2D data at the H_20 frequency (Figure 5-6). This shows the observed σ^{NOE} following the convention that negative values of $\sigma^{NOE/ROE}$ are represented by positive peaks and positive values are represented by negative peaks. Figure 5-7 shows a number of signals generated by positive and negative σ^{NOE} giving rise to the differing signs.

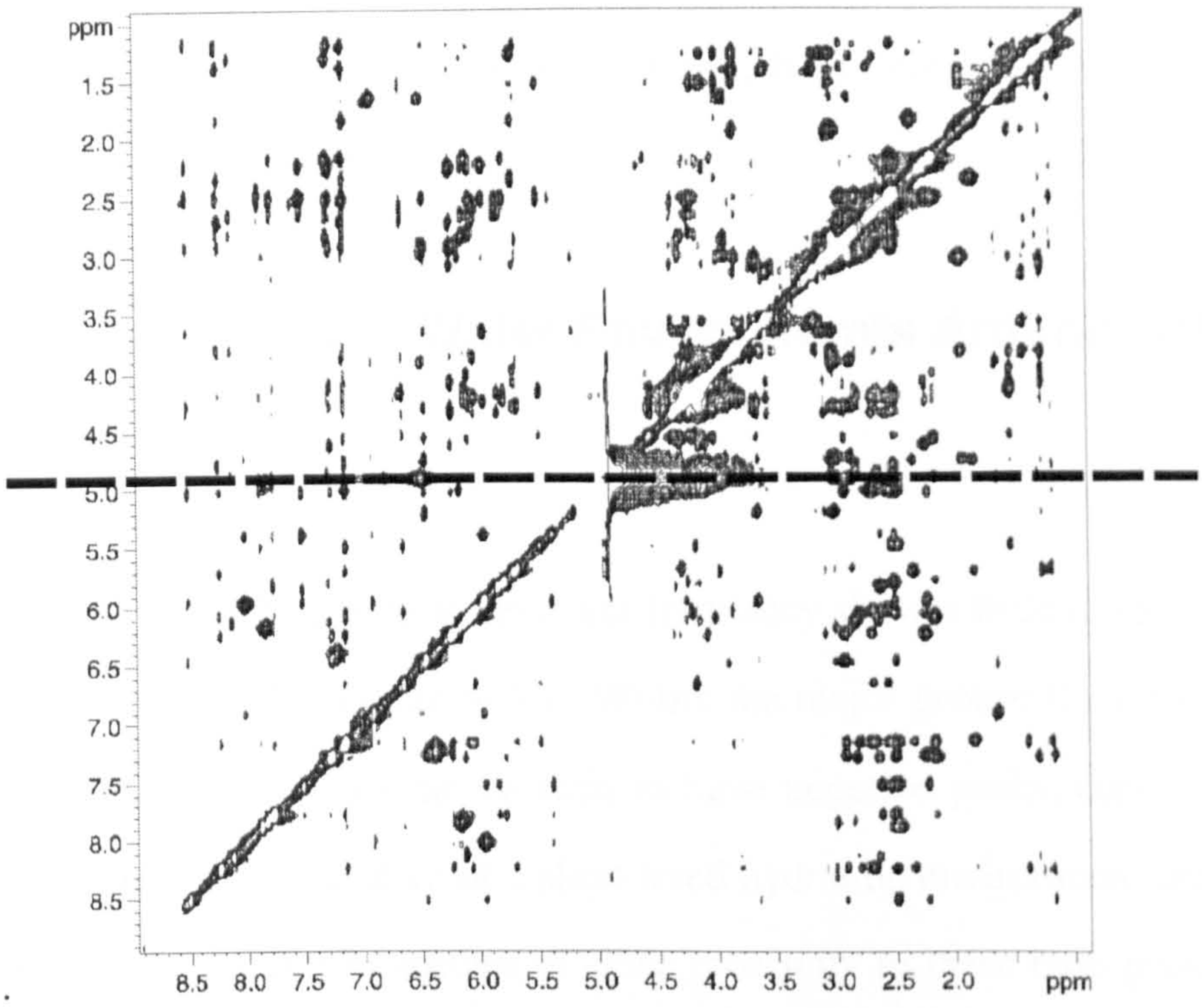


Figure 5-6: Cross-section are taken through NOESY and ROESY experiments at the frequency of water resonance.

NOESY cross-peaks can arise from exchange mediated transfer and not just dipolar coupling and this gives rise to strong positive peaks in the NOESY spectra. However, the ROESY cross-section allows us discriminate between these two cases based on the apparent sign of the ROE. Since as was discussed previously only positive σ^{ROEs} are possible, only negative peaks would be observed for all dipolar couplings. The presence of a positive peak can only arise through the different pathway of exchange.

The ROESY cross-section in Figure 5-7b shows a number of positive peaks due to exchange. The large peak at ~6.4 ppm is attributed to the base amino resonances.

Exchangeable proton signals are also observed for the base imino protons with chemical shifts of between 11 and 14 ppm.

5.5. Residence Times of Water Environments Around Solute Protons

Cross sections of NOESY spectra at the water frequency show a wide range of differing sign and intensity of NOEs (Figure 5-7c). Within the major groove the signals C10H6 (peak g.) and A11H8 (peak a.) can be seen to have negative peaks, corresponding to positive σ^{NOE} values and indicative of a short lived hydration environment around these protons. Typically the water molecules in close proximity to these base protons would be expected to have a residency time of less than 500ps. In comparison positive peaks are observed for the resonances G3H8 (peak d), G9H8 (peak f), A5H2 (peak e), T8H6 (peak h), and weakly for A7H8 (peak c), These protons reside in the major groove of DNA and are in a long-lived hydration environment. Weak negative signals between H1' protons and water indicate the presence of some stable hydration in the minor groove. It is interesting to note that water NOEs to H1' signals are not usually detected in free DNA under these conditions suggesting little stabilising interactions in free DNA (Liepinsh et al. 1994; Conte et al. 1996). Further, a number of drug protons appear in close proximity to water. Both aromatic protons on the aglycone (H3, H11) and the anomeric proton on the N,N-dimethylaminoglucose are in an environment of high occupancy water.

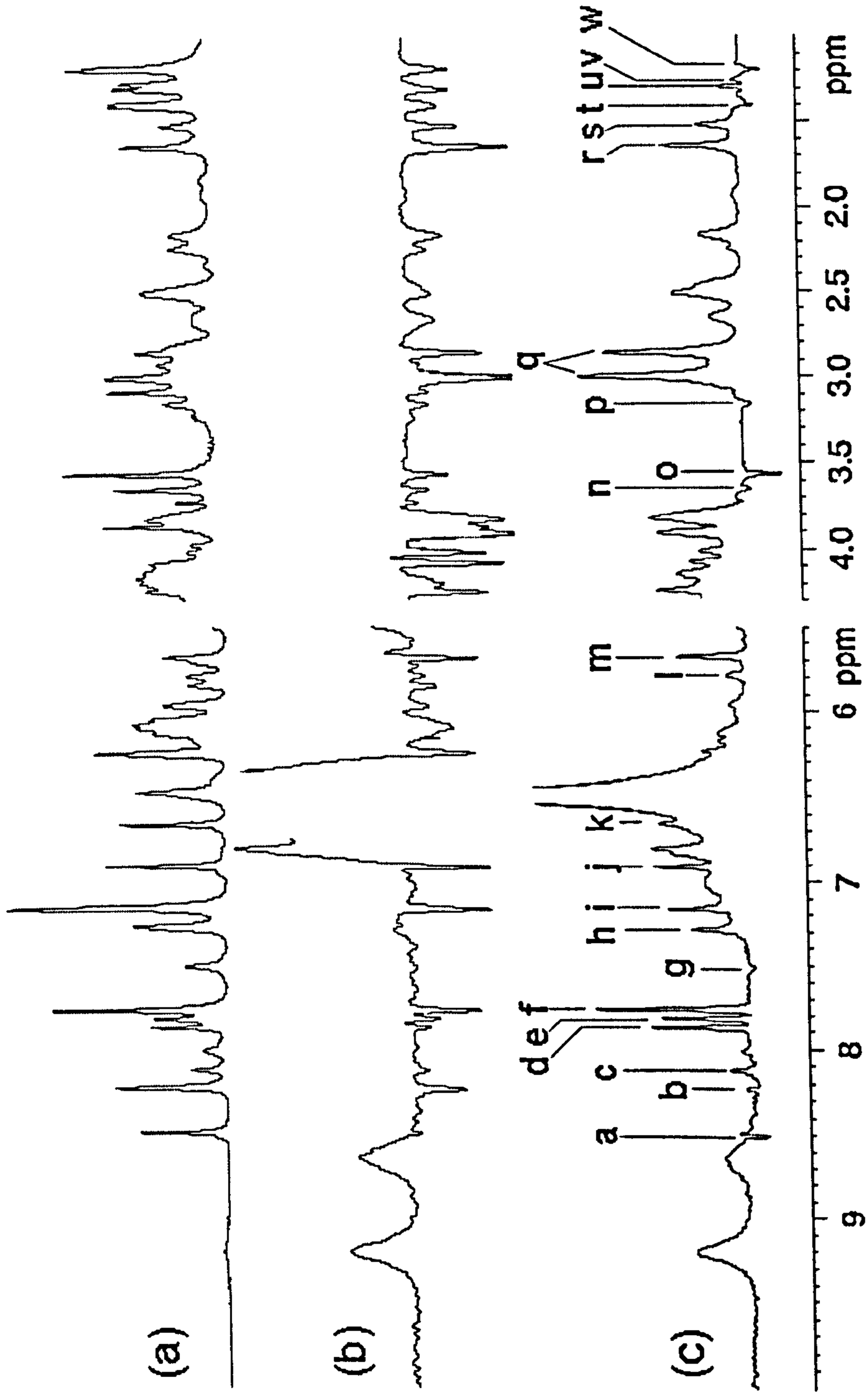


Figure 5-7: 1D 1H reference spectrum (a) and cross sections taken at the water frequency for ROESY (b) and NOESY (c). All spectra were recorded at 288K and calibrated to the internal trimethylsilylpropanoic acid (TSP) methyl protons. Peaks a-w were assigned as follows: a, A11 H8; b, A1/A5 H8; c, A7 H8; d, G3 H8; e, A5 H2; f, G9 H8; g, C10 H6; h, T8 H6; I, T2/C4/T6 H6; J. Nog. H3; k, Nog. H11; l, T8 H1'; m, Nog. H1"; n, 2'-OCH3; o. 4'-OCH3;p, Nog. H4'; q, Nog N(CH3)2; r, 5"-CH3; s, 9-CH3; t, 5'-CH3 / T6 CH3; u, T8 CH3; v, 3'-CH3;w, T2/T12 CH3.

Within the minor groove signals from the drug nogalose sugar 3'-CH_3 (Figure 5-7 peak v) 2'-OCH_3 (peak n), and 4'-OCH_3 (peak o) show clear positive σ^{NOE}. This suggests that the area surrounding these groups is experiencing a rapidly exchanging water environment. The hydrophobic sugar, although buried in the minor groove, shows some groups are accessible to solvent, in agreement with modelling and X-ray data that shows this same region has a high occupancy. Presumably, the high occupancy is derived from a combination of the pocket formed by the nogalose sugar and the floor of the minor groove, and weak hydrogen bonding from the methoxy groups on the nogalose sugar. However, the NMR information reveals that these protons are in an environment of low residence time water (<500 ps).

Significant long lived water interactions are only observed for two of the drug's methyl environments around the 5"-CH_3 on the α-D-3,6-dideoxy-3(N,N-dimethylamino)-glucose and 9-CH_3 on the A ring. The DNA thymine methyl group at T8 also shows a water-mediated interaction with a long residence time. Such an observation is unusual since the remainder of the thymine methyls are in regions of low occupancy water, and experiments with free DNA show only weakly bound solvent for all thymine methyls. This clearly suggests that the drug is able to modify and stabilise the solvent environment.

5.6. The role of exchangeable protons

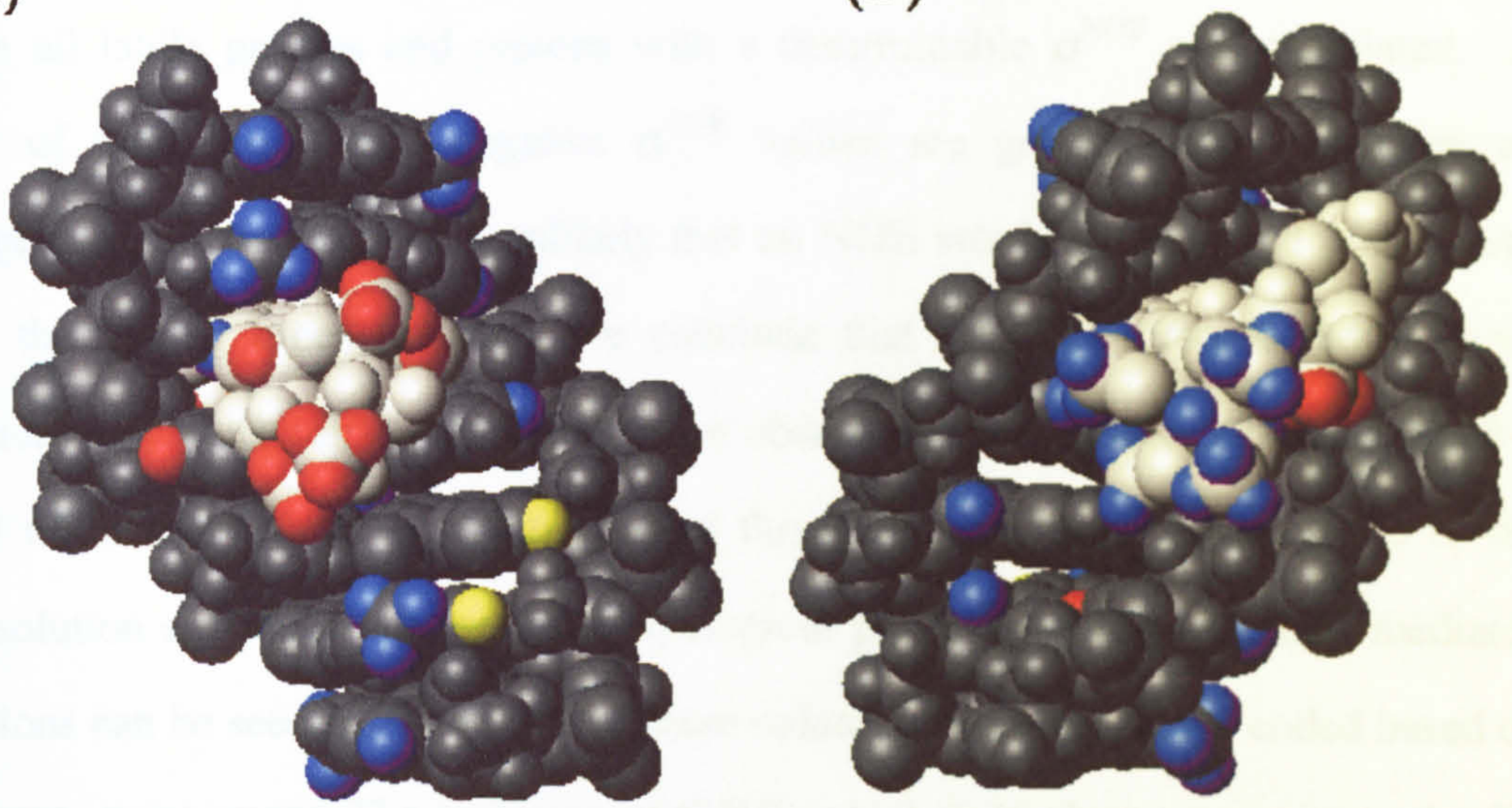


Figure 5-8: Graphical representation of NMR data from Table 5-1 and Table 5-2, with views into the major groove (a) and minor groove (b). DNA atoms are coloured black and drug atoms white. Protons which have been identified as close to long-lived hydration environments (>0.5ps) are coloured red, Protons which have been identified as close to short lived protons hydration have been coloured blue. Those protons that have been determined as being close to exchangeable groups based on the dynamics modelling are coloured yellow.

As discussed, the analysis presented above is complicated by the role of exchangeable protons. For signals that have been attributed to long lived water interactions of greater that 500ps, the effect could be mediated by a direct interaction with water or via an NOE to a proton that subsequently exchanges with water (Kubinec and Wemmer, 1992; Otting and Liepinsh, 1995). Although this effect is indistinguishable from the direct mechanism based on the NMR data alone, comparison with the high-resolution structure provides an insight into the proximity of protons to potentially labile OH and NH groups. This combination of techniques provides a powerful method for the

investigation of hydration structure by overcoming some of the limitations inherent with each individual technique. The average distance and one standard deviation between all labile protons and protons with a determinable σ^{NOE} was calculated. A number of interactions with negative σ^{NOE} values are greater than 5 Å from an exchangeable proton. Since it is unlikely that an NOE would be observed at distances ≥5Å at the mixing times involved, we conclude that for these protons that a direct interaction with water can be attributed to the observed NOE. Table 5-1 and Table 5-2 lists all water interactions and distinguishes those that are direct by reference to the high-resolution structure. The striking topological grouping of the hydration mediated interactions can be seen in Figure 5-8. Where solute protons are colour coded based on the solvent environment. The inclusion of G3H8 which is in close proximity to the 2"-OH (4.2 Å) is justified from dynamics data that suggests the G3N7 oscillates between a direct interaction and a solvent mediated one. This would imply that a strongly associated water molecule in part mediates the G3H8 interaction as discussed above.

Long residence time water observed from G9H8, T8H6 and T8CH_3 form a cluster in the major groove and water molecules in the pocket formed appear to mediate interactions between the groove and the charged N,N-dimethylammonium ion on the α-D-3,6-dideoxy-3(N,N-dimethylamino)-glucose group.

The solvent environment of the minor groove by contrast is dominated by only short residence times (Figure 5-8(b)). The only exception is the 9-CH_3, which is in close proximity to a hydroxyl group (9-OH) on the same ring carbon. Due to the complication of the adjacent labile proton it is not possible to determine the solvent environment around this group, however the presence of a hydrophilic group within a pocket would suggest that a long residence hydration would be feasible.

Table 5-1. NOEs between DNA and water protons in the d(ATGCAT)-Nogalamycin Complex. NOEs are assigned based on visual inspection as: s, strong; m, medium; w, weak; ~0, resolved signal but NOE absent. The nearest exchangeable proton [distance (+/- standard deviation)] that could give rise to a mediated exchange NOE with water is listed. Distances in excess of 4.5Å would not be expected to contribute to an exchange mediated effect and are not shown.

DNA Protons	Sign and magnitude of σ^{NOE}	T_{eff} (ns)	Nearest exchangeable
A11 H8	+ (w)	<0.5	6-OH 4.79(0.12)[a]
A1/A5 H8[b]	- (w)	≥0.5	-
A7 H8	- (w)	≥0.5	-
G3 H8	- (m)	>0.5	2"-OH 4.21 (0.38)
A5 H2	- (m)	>0.5	T8 N3H 2.76
G9 H8	- (s)	»0.5	-
A7 H2	- (s)	»0.5	T6 N3H 2.76
C10 H6	+ (w)	<0.5	6-OH 3.91 (0.24)[a]
T8 H6	- (m)	>0.5	-
T2/T6/C4 H6[b]	- (m)	>0.5	-
T8 H1'	- (w)	≥0.5	-
G9 H1'	~0	≈0.5	-
C10 H1'	- (w)	≥0.5	6-OH 3.68 (0.21)[b]
			4-OH 4.77 (0.24)
A5/T12 H1'[a]	- (w)	≥0.5	-
C4 H1'	~0	≈0.5	-

T8 CH_3	- (w)	≥0.5	NM NH 5.18 (1.07)
T6 CH_3	+ (w)	<0.5	-
T2 CH_3	+ (w)	<0.5	2"-OH 5.6 (0.79)
T12 CH_3	+ (w)	<0.5	-

[a] Resonance overlap prevents unambiguous assignment
[b] The 6'-OH gives very weak exchange cross-peaks suggesting it is protected from solvent exchange. It is not considered likely that it plays a role in solvent exchange

Table 5-2. NOEs between nogalamycin protons and water protons in the d(ATGCAT)-Nogalamycin Complex. NOEs are assigned based on visual inspection as: s, strong; m, medium; w, weak. The nearest exchangeable proton [distance (+/- standard deviation)] that could give rise to a mediated exchange NOE with water is listed. Distances in excess of 4.5Å would not be expected to contribute to an exchange mediated effect and are not shown.

Drug Protons	Sign and magnitude of σ^{NOE}	T_{eff} (ns)	Nearest Exchangeable proton
H3	- (m)	>0.5	4"-OH 3.11 (0.37)
			4-OH 3.49 (0.08)
			NM NH 5.28 (0.42)
H11	- (w)	≥0.5	-
H1"	- (m)	>0.5	2"-OH 2.77 (0.30)
			NM NH 5.24 (0.38)
2'-OCH_3	+ (w)	<0.5	6-OH 3.04 (0.32)[a]
3'-OCH_3	+ (w)	<0.5	6-OH 4.30 (0.52)[a]
			9-OH 4.48 (0.26)

4'-OCH_3	+ (m)	«0.5	-
H4"	+ (W)	<0.5	-
5"-CH_3	- (m)	>0.5	4"-OH 3.16 (0.39)
			NM NH 5.28 (0.42)
NM-CH_3	- (s)	»0.5	NM NH 2.36 (0.20)
			2"-OH 3.22 (0.23)
			4"-OH 3.60 (0.69)
9-CH_3	- (m)	>0.5	6-OH 3.97 (0.17)[a]
			9-OH 2.34 (0.21)
5'-CH_3	+ (w)	<0.5	9-OH 4.53 (0.50)
3'-CH_3	+ (w)	<0.5	6-OH 4.30 (0.52)[a]

[a] The 6'-OH gives very weak exchange cross-peaks suggesting it is protected from solvent exchange. It is not considered likely that it plays a role in solvent exchange

5.7. Conclusions

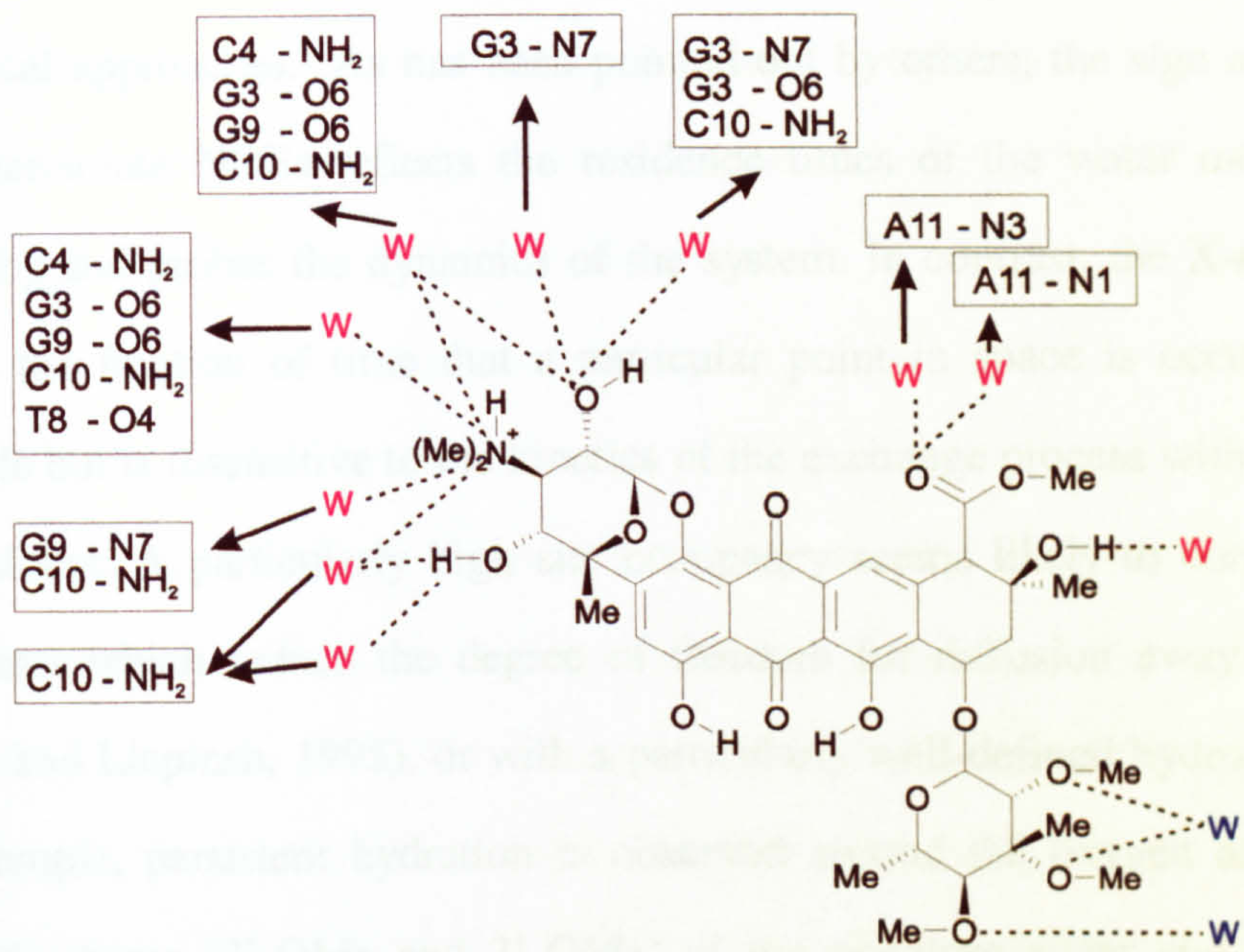


Figure 5-9: Possible interactions that can be mediated by water molecules at computed high-density sites. Hydrogen bonding calculated on an optimal 2.8Å distance. Water molecules that correlate with long residence time areas are coloured in red, water molecules that would occur in short residence time regions are coloured blue

While there is some correlation regarding the position of long-residence time water identified by NOEs and high occupancy sites identified in the X- ray structures and the rMD simulations, many high occupancy sites, particularly in the minor groove, are suggested by the NOE data to correspond to only short-lived interactions with water. In NMR studies of protein hydration nearly all NOEs from water with negative σ^{NOE} values have been shown to be mediated through the hydroxyl protons of Ser, Thr and

Tyr (Otting et al. 1991), and that only a few of the water molecules detected by X-ray analysis are also characterised by sizeable negative σ^{NOEs}, i.e., long residence times. It is important to emphasise the subtle differences between the different experimental and theoretical approaches. As has been pointed out by others, the sign and magnitude of the water-solute NOEs reflects the residence times of the water molecules in close proximity and probes the dynamics of the system. In contrast, the X-ray water density reflects the fraction of time that a particular point in space is occupied by a water molecule but is insensitive to the kinetics of the exchange process with molecules in the bulk solvent. A particularly high site occupancy seems likely to correlate with steric constraints which reduce the degree of freedom for diffusion away from these sites (Otting and Liepinsh, 1995), or with a particularly well-defined hydrogen-bonding site. For example, persistent hydration is observed around the oxygen atoms of the drug methoxy groups (2'-OMe and 3'-OMe) of the nogalose sugar and the 9-OH of the aglycone A-ring in the dynamics simulations and the X-ray structures (Smith et al. 1996), however, the NOE data show that these inter- actions are associated with short residence times. In proteins, sizeable σ^{NOE} around surface accessible peaks were observed around hydroxyl groups and an extensive stabilising solvent network is evident in the major groove in both the simulations and X-ray data. There is also see negative σ^{NOEs} that suggest that some of these interactions are also kinetically stabilised. Thus, it is the fraction of time that a particular point in space is occupied by water molecules, not the kinetics, that the MD simulations and X-ray data identify. Our MD results suggest good agreement between these two approaches.

In contrast to the conclusions from studies of protein surface hydration (Otting et al. 1991), both NMR and X-ray crystallographic analysis of the structural role of water in DNA A-tracts have identified water molecules to be of structural significance in the

form of a spine of hydration in the minor groove (Drew and Dickerson, 1981; Kubinec and Wemmer, 1992; Berman, 1994; Liepinsh et al. 1994; Eisenstein and Shakked, 1995; Chandrasekaran et al. 1995; Kochoyan and Leroy, 1995; Kochoyan and Leroy, 1995; Egli et al. 1996; Johannesson and Halle, 1998; Shui et al. 1998). Similar conclusions have been drawn regarding hydration of the 2'-OH of RNA (Conte et al. 1996) suggesting this to be an important factor in stabilising the A-form RNA helical structure. More recently, the role of solvent in protein-DNA interactions has been investigated by NMR and reveals that water molecules identified in polar cavities mediating protein-DNA recognition in X-ray structures can also be identified in solution structures, indicating a diverse use for water molecules in recognition processes (Billeter et al. 1996; Schwabe, 1997). In a number of cases a fluctuating network of hydrogen bonds involving a handful of water molecules appears to play an important role in determining specificity, or mediating peptide-protein recognition (Smith et al. 1996). We have seen from the dynamics simulation described in this work that a key interaction between nogalamycin and DNA (2"-OH to G3 N7), previously identified as an important component of binding specificity, may undergo dynamic fluctuations in which solvent mediates the interaction for at least part of the time.

5.8. Experimental Methods

5.8.1 NMR Sample Preparation

NMR studies to determine the residence time of water molecules within the Nogalamycin-d(ATGCAT)$_2$ complex were carried out using a Bruker Avance 500MHz spectrometer. All samples were prepared with 100mM NaCl and 10mM NaH_2PO_4 salt

solution and NaN_3 (6 mM) as an anti-bacterial agent and the tetrasodium salt of EDTA (0.6 mM) as a metal chelating agent. Samples were dissolved in H_2O doped with D_2O (10% by volume) to allow deuterium signal locking.

5.8.2 NMR Data Collection

All experiments were carried out at 288K. This avoided coincidental overlap of the water resonance with other non-exchangeable protons such as the deoxyribose protons at the 3' position. Both NOESY and ROESY experiments were acquired using the WATERGATE solvent suppression sequence. All experiments were acquired with between 80 and 128 scans and 4096 complex data points in the t2 dimension and 512 points in the indirect t1 dimension. Spectra were calibrated to 0 ppm by reference to the internal standard trimethylsilylpropanoic acid (TSP). After phasing and application of baseline correction spectra were strip transformed around the residual water signal in F1. The matrix obtained was zero-filled to 16384 by 2048 data points. Projection in F2 at the determined frequency of the H_2O resonance were obtained and baseline corrected to remove distortions around 4.8 ppm.

Analysis of the NOESY and TOCSY (recorded in D_2O) shows the location of all the 3' protons to be clear of the water signal at 4.8ppm. The use of mixing times of 100ms for ROESY and 75ms NOESY minimises effects of spin-diffusion that could occur during prolonged mixing times.

5.9. Bibliography

Berman, H.M. (1994) Hydration of DNA: Take 2. *Cur. Op. Struct. Biol.* **4**, 345-350.

Billeter, M., Guntert, P., Luginbuhl, P. and Wüthrich, K. (1996) Hydration and DNA recognition by homeodomains. *Cell* **85**, 1057-1065.

Chandrasekaran, R., Radha, A. and Park, H.S. (1995) Sodium-ions and Water-molecules in the structure of poly(dA)poly(dT). *ACTA Crysta. D* **51**, 1025-1035.

Conte, M.R., Conn, G.L., Brown, T. and Lane, A.N. (1996) Hydration of the RNA duplex r(CGCAAAUUUGCG)$_2$ determined by NMR. *Nuc. Acids. Res.* **24**, 3693-3699.

Drew, H.R. and Dickerson, R.E. (1981) Structure of a B-DNA Dodecamer .3. Geometry of Hydration. *J. Mol. Biol.* **151**, 535-556.

Egli, M., Portman, S. and Usman, N. (1996) RNA hydration: a detailed look. *Biochem.* **35**, 8489-8494.

Eisenstein, M. and Shakked, Z. (1995) Hydration patterns and intermolecular interactions in A-DNA crystal structures. Implications for DNA recognition. *J. Mol. Biol.* **248**, 662-678.

Fujiwara, T. and Nagayama, K. (1985) The wobbling-in-a-cone analysis of internal motion in macromolecules. *J. Chem. Phys.* **83**, 3110-3117.

Johannesson, H. and Halle, B. (1998) Minor Groove hydration of DNA in solution: Dependence on base composition and sequence. *J. Am. Chem. Soc.* **120**, 6859-6870.

Kochoyan, M. and Leroy, J.L. (1995) Hydration and Solution Structure of Nucleic Acids. *Cur. Op. Struct. Biol.* **5**, 329-333.

Kubinec, M.G. and Wemmer, D.E. (1992) NMR Evidence for DNA Bound Water in Solution. *J. Am. Chem. Soc.* **114**, 8739-8740.

Liepinsh, E., Leupin, W. and Otting, G. (1994) Hydration of DNA in Aqueous-Solution - NMR Evidence for a Kinetic Destabilization of the Minor-Groove Hydration of d-$(TTAA)_2$ Versus d-$(AATT)_2$ Segments. *Nuc. Acids. Res.* **22**, 2249-2254.

Otting, G. and Liepinsh, E. (1995) Selective Excitation of Intense Solvent Signals in the Presence of Radiation Damping. *J. Bio. Mol. NMR* **5**, 420-426.

Otting, G., Liepinsh, E. and Wüthrich, K. (1991) Protein Hydration in Aqueous Solution. *Science* **254**, 974-980.

Schwabe, J.W.R. (1997) The role of water in protein-DNA interactions. *Cur. Op. Struct. Biol.* **7**, 126-134.

Shui, X.Q., McFailIsom, L., Hu, G.G. and Williams, L.D. (1998) The B-DNA dodecamer at high resolution reveals a spine of water on sodium. *Biochem.* **37**, 8341-8355.

Smith, C.K., Brannigan, J.A. and Moore, M.H. (1996) Factors affecting DNA sequence selectivity of nogalamycin intercalation: the crystal structure of d(TGTACA)$_2$-Nogalamycin$_2$. *J. Mol. Biol.* **263**, 237-258.

Smith, K.J., Reid, S.W., Harlos, K., McMichael, A.J., Stuart, D.I., Bell, J.I. and Jones, E.Y. (1996) Bound water structure and polymorphic amino acids act together to allow the binding of different peptides to MHC class I HLA-B53. *Immunity* **4**, 215-228.

Chapter 6. DNA Hairpins as Novel Nicked Sequence Analogues - Drug Binding at DNA Strand Breaks

6.1. Introduction

Hairpin structural motifs are a well-known occurrence in RNA chemistry but relatively few examples are known for DNA. The conventional view on hairpin stability is that it is governed by the base-pairing stem regions and the turn or loop region is disordered contributing little to the hairpin stability. DNA sequences such as d(GCGAAAGC) and d(GCGAAGC), however form stable hairpin structures with the loop region formed by the GAAA and GAA sequences with only two G-C basepairs in the stem region. The unusually high stability suggests that these purine-rich loop sequences play an active role in hairpin stabilisation (Varani, 1995). The nucleotide sequence d(GCGAAGC) has been shown to adopt a stable hairpin arrangement in solution with a melting temperature of 76 °C in 0.1 M NaCl. NMR solution studies show the sequence forms two stable GC base pairs in the stem. The loop is facilitated by the $G^3A^4A^5$ sequence, with the A4 stacking over the G3 and the A5, which are involved in anti-anti GA base-pairing (Figure 6-1) (Hirao et al. 1994).

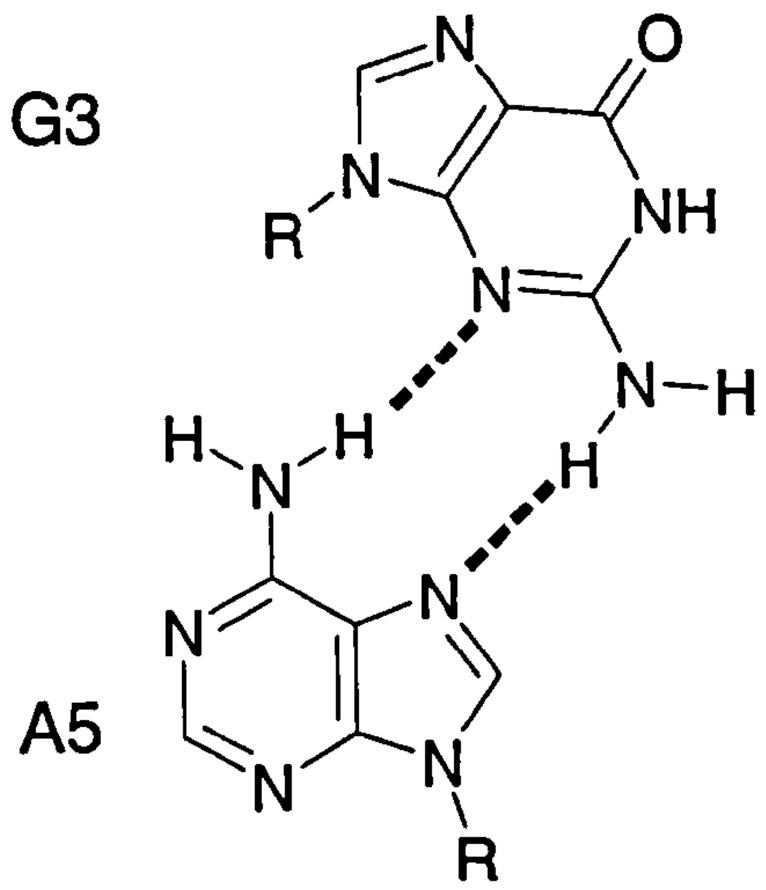


Figure 6-1: Anti-anti base pairing between G^3 and A^5 in the sequence d(GAA). A^4 stacks with the G^3 giving a characteristic chemical shift pattern.

Three basepairs are involved in the loop, however, the GA basepair appears critical to the stabilisation. The role of the middle adenine (G<u>A</u>A) is less critical and replacement with any of the other bases still confers a stable hairpin arrangement (Yoshizawa et al. 1997). Investigations using these mini-hairpin sequences to protect nucleotides from degradation from 3' nucleases digestion have already shown the hairpins form even when attached to longer DNA strands (Yoshizawa et al. 1994). The significantly increased resistance to nuclease degradation, has given rise to possible applications for protecting anti-gene sequences *in vivo*.

The stability of these mini-hairpins also gives rise to some interesting possibilities for engineering and studying nicked sequences of doubled stranded duplex DNA. The concatenation of two hairpin sequences within a single nucleotide strand provides a readily accessible model system for generating a nick site at a predefined position within the stem region (Figure 6-2). To this end, we have used the single stranded

oligonucleotide d(ACGAAGTGCGAAGC,) containing two d(GAA) loop sequences, this potentially generates the structure shown in Figure 6-2 with a TpG nogalamycin binding site at the centre of the sequence, and the CA step on the opposing strand carrying a strand break. In this chapter, we examine the stability of the structure shown schematically in Figure 6-2, and investigate the complex formed with nogalamycin bound at the TpG (CA) site. Our interest in pursuing this binding interaction is to shed some light on the mechanism by which nogalamycin is able to achieve a stable intercalation complex in regular double stranded DNA, analogous to the complex described in earlier chapters. It is well established that the binding kinetics for nogalamycin are very slow (Fox and Waring, 1984). Such a large activation barrier to binding has been attributed to disruption (local melting) of the duplex structure as the rate-limiting step. Thus, drug binding to a single stranded melted region of the sequence, followed by re-annealing of the duplex around the drug, appears to be a plausible hypothesis. We have investigated this possibility using our model system.

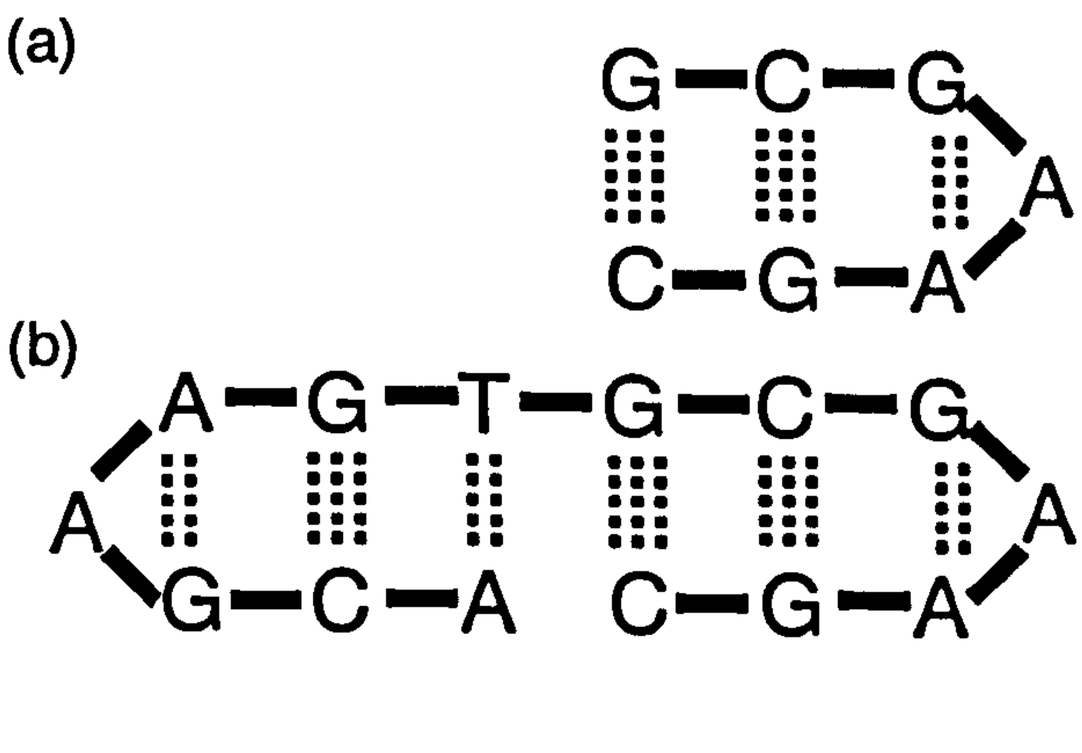


Figure 6-2: The structure d(GCGAAGC) (a) has been shown to form a stable mini-hairpin utilising the GAA turn sequence. The sequence d(ACGAAGTGCGAAGC) is postulated to also adopt a structure similar to that detailed above (b). This sequence has four Watson-Crick basepairs forming a stem region. The central TpG base step is a high affinity nogalamycin binding site.

6.2. Structure of a Novel Hairpin Loop

6.2.1 UV melting studies

The sequence d(ACGAAGTGCGAAGC) was synthesised and purified as previously described and the melting transition monitored by change in UV absorption. The annealed structure was heated from room temperature (20 °C) to 98 °C using a Peltier heating block and the UV absorbance at 260 nm monitored (Figure 6-3). The structure exhibited a broad but intense absorbance change with a melting point calculated from the gradient of the curve to be 69 °C, suggesting a temperature dependant transition from a duplex to single stranded form. This temperature is slightly below that of the isolated single hairpin loop described previously. However, the broadness of the transition observed here suggests that the two hairpins have closely similar, but not identical, melting transitions that can not be resolved under these conditions. The hairpin portion of the sequence carrying the T-A basepair would be expected to melt at a slightly lower temperature. The broadness of the transition monitored by UV suggests that a single highly co-operative melting of the whole structure is unlikely, and that the two hairpin portions behave somewhat independently, with a high degree of flexibility at the nick site. This later flexibility seems likely to account for the broad melting transition.

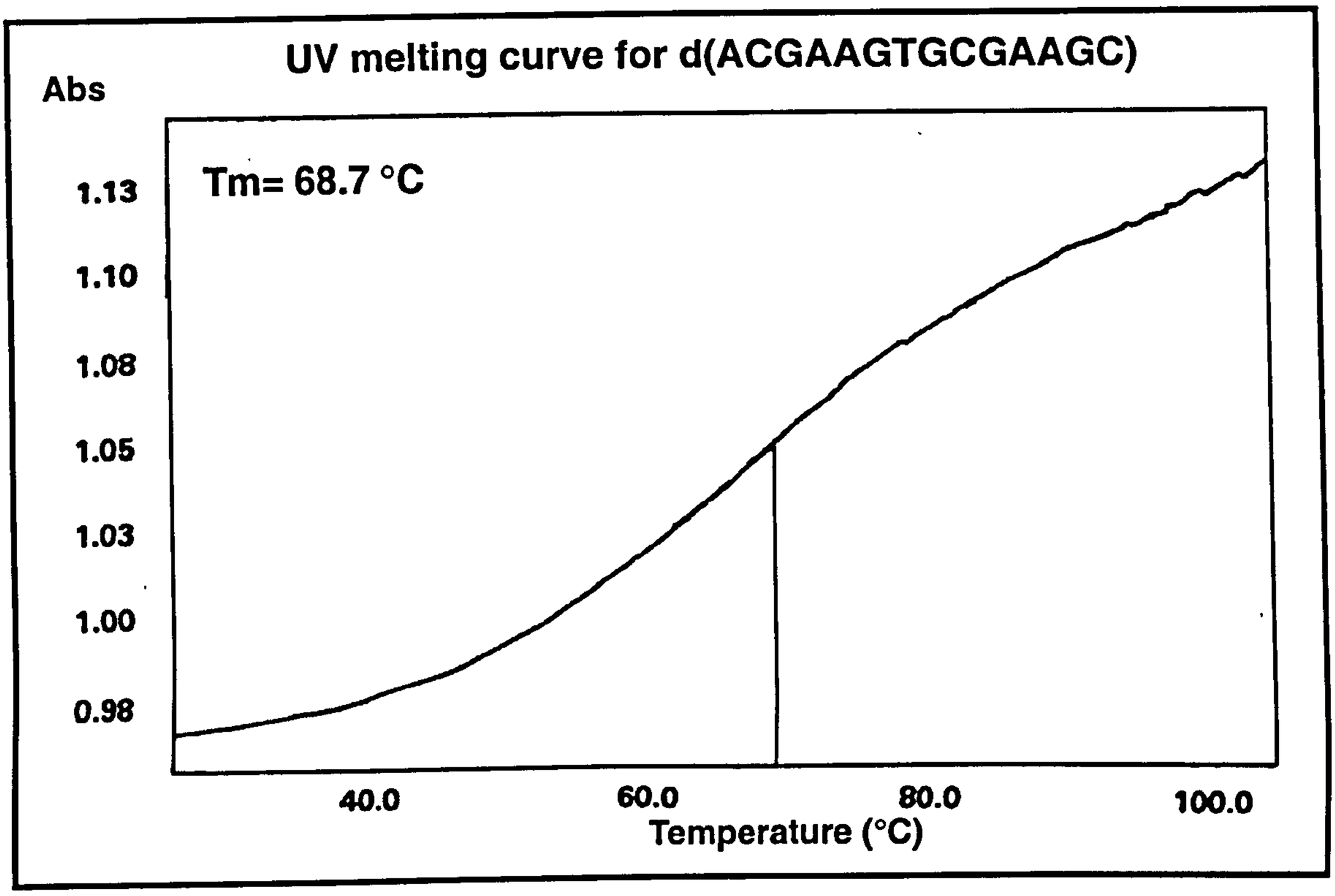


Figure 6-3: UV melting curve for d(ACGAAGTGCGAAGC). Absorbance is recorded at 260nm. An intense but broad transition is observed with the mid point of the curve corresponding to a temperature of 68.6 °C

6.2.2 NMR structural analysis

The NMR spectrum of the folded structure was assigned using the combination of NOESY, TOCSY and DQF-COSY experiments described in earlier chapters. The formation of a fully folded structure was immediately apparent from a number of unusual chemical shifts. Two of the three cytosine H2' resonances were up-field shifted to around 1.5 ppm, consistent with observations of the single hairpin (Hirao et al. 1994). The A4 H4' and A11 H4' were identified from TOCSY and NOESY peaks at a chemical shift of ≈ 2.1 ppm. This is an extreme up-field shift for these protons,

which normally occur at ≈ 4.5 ppm. A portion of the NOESY spectrum of the structure is shown in Figure 6-4 highlighting some of these assignments.

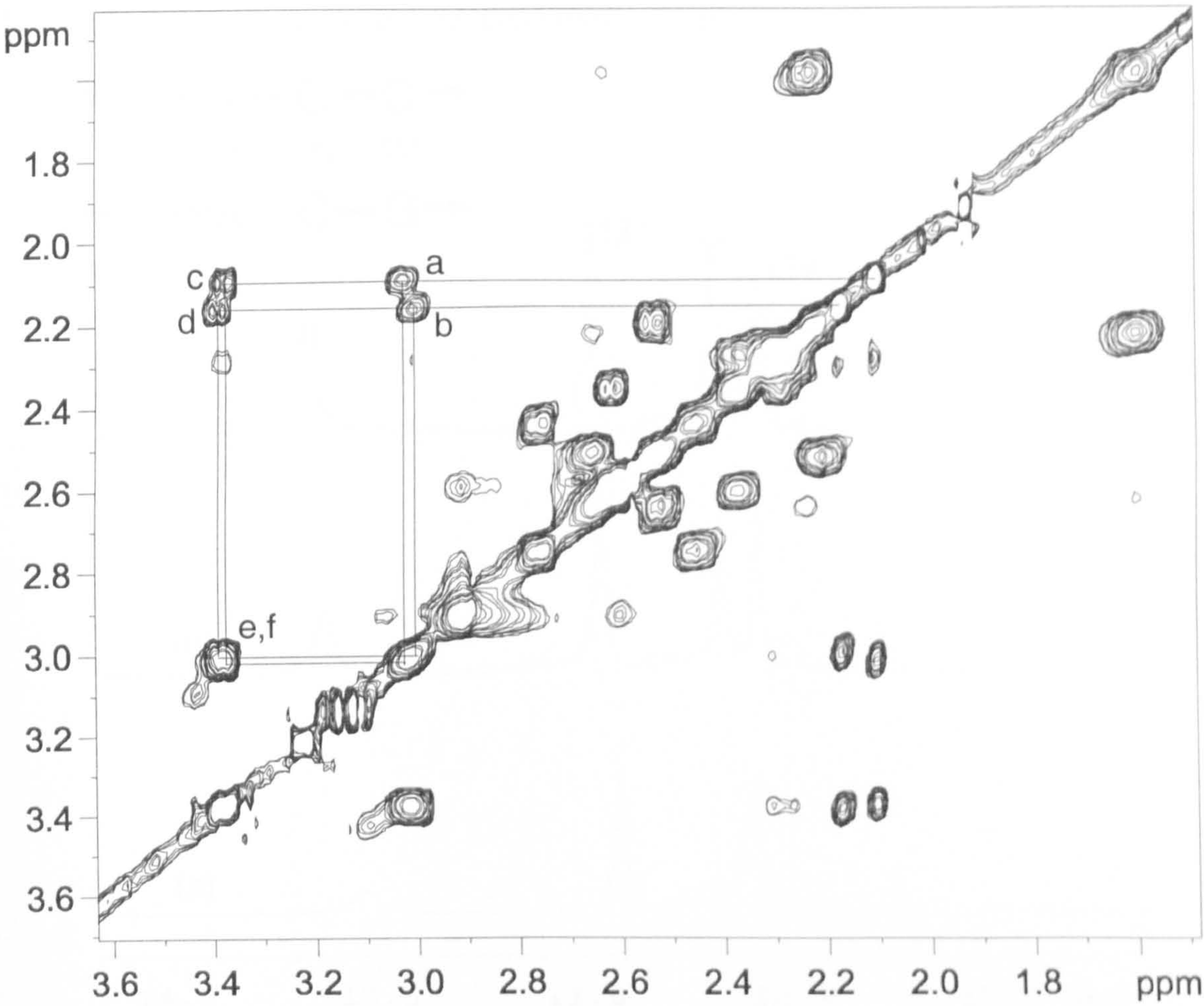


Figure 6-4: Portion of a NOESY spectrum, obtained at 298 K, showing the location of Adenine H4' and NOESY cross peaks to H2'. (a,c) A11 H4' → A11 H2'/H2''; (b,d) A4 H4' → A4 H2'/H2''; (e,f) A4/11 H2' → A4/11 H2''. These assignments are confirmed by TOCSY and DQF-COSY experiments.

These shifts are analogous to those identified in the isolated hairpins (Hirao et al. 1994; Yoshizawa et al. 1997). Such large perturbations are explained by the close proximity of the H4' of A4 and A11 to the non-basepaired bases of A5 and A12 respectively,

present in each of the loops. This unusual shift pattern is indicative of formation of a GAA turn. The similarities in the pattern of shifts (and NOEs) confirm that both hairpin loop sequences are formed under these conditions.

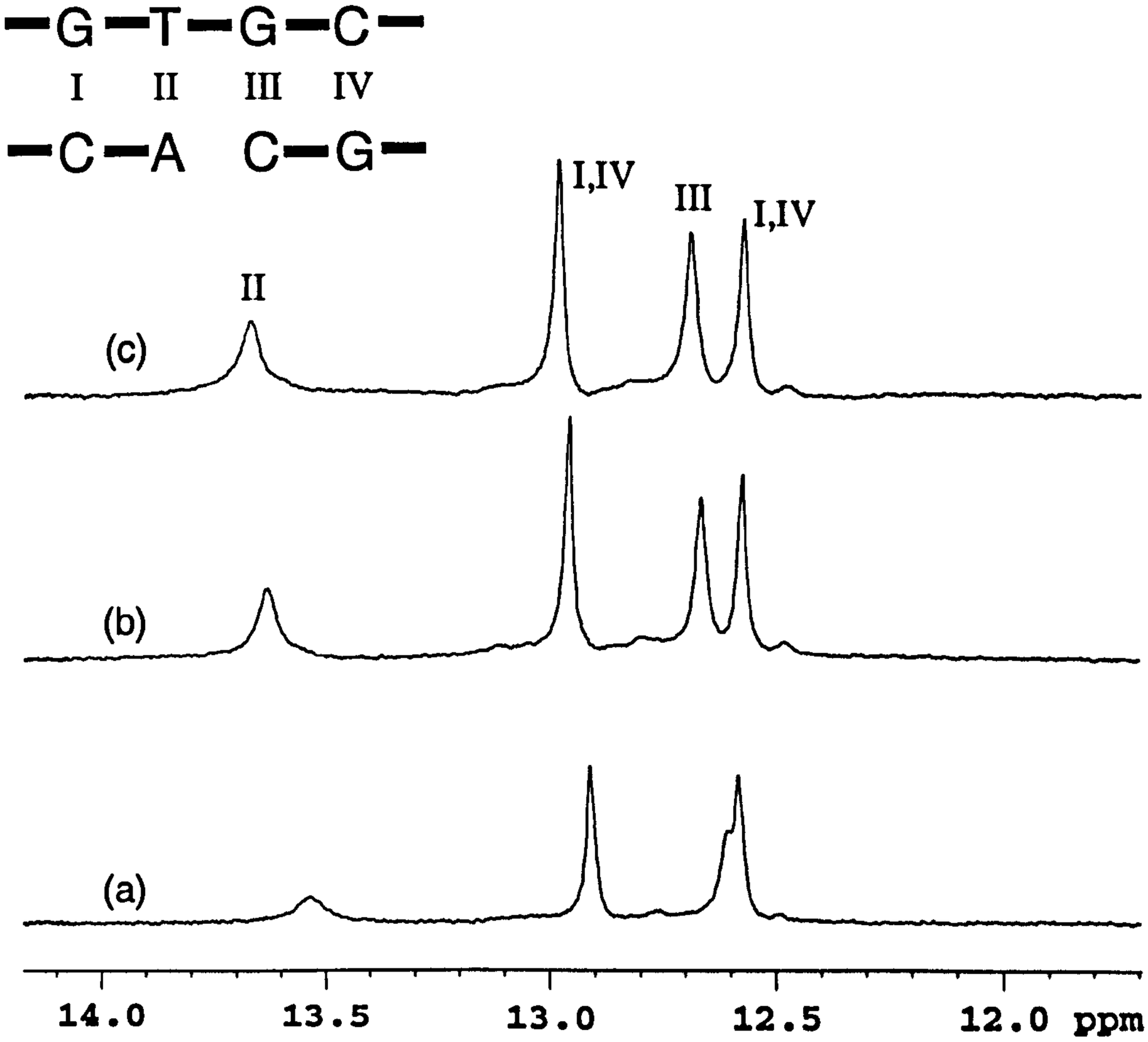


Figure 6-5: Imino resonances observed in 90% H_2O. Resonances formed are stabilised by base pairing. At 298K (a) four resonances are observed which appear to give shift changes consistent with an increase in stability at the lower temperatures 288K (b) and 278 K (c).

This conclusion is further substantiated by the NMR spectrum in H_2O solution, which enables the imino proton resonances to be identified for the Watson-Crick base pairs in the stem regions of the hairpins. The observation of broadened resonances for the base

pairs at the nick site indicates that these are fairly dynamic. These imino protons are not stabilised to the same extent as those at the centre of a normal DNA duplex, but are reminiscent of those at the ends of a duplex. Thus, the two hairpins appear to behave independently with relatively weak stacking interactions across the break site. The temperature dependence of the imino proton resonances shows them to sharpen considerably at low temperature (Figure 6-5).

The base pairing adopted around the GA turn does not protect the guanine imino protons from exchange and these are not apparent in the 12-15 ppm region of the spectrum.

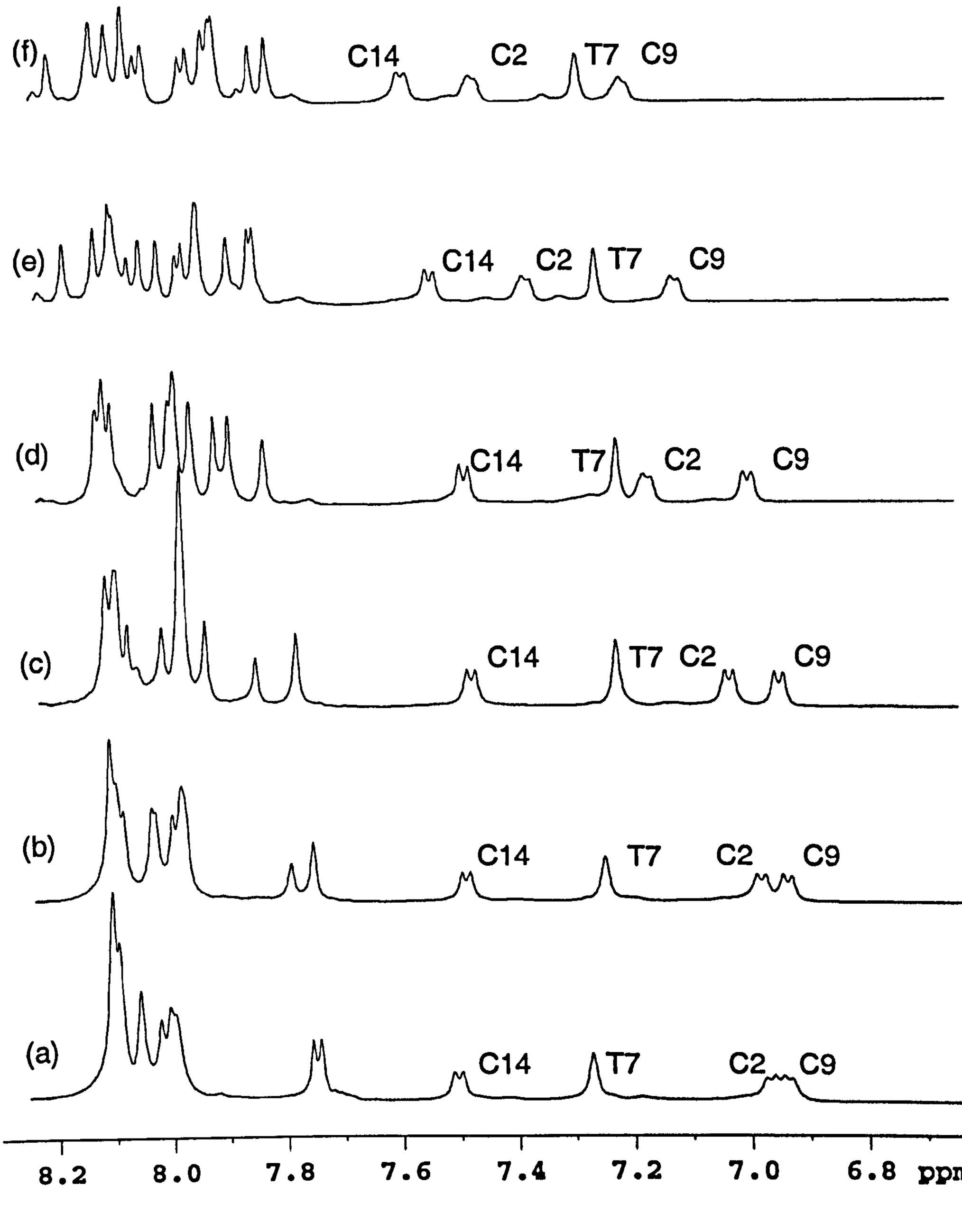


Figure 6-6: Aromatic region of the ^{1}H NMR spectrum of d(ACGAAGTGCGAAGC) showing the temperature sensitivity of chemical shift within the hairpin sequence. Spectra are recorded at (a)298 K ,(b) 308 K,(c)318 K,(d) 328 K,(e) 338 K,(f)343 K. The C2 H6 shows a marked dependence on temperature undergoing a shift change of 0.5 ppm

The NMR temperature profile can be used to follow the melting of the two hairpins since a number of resonances have unusual chemical shifts in the folded state as a consequence of aromatic ring current effects. The C2 H6 in particular (6.98 ppm at 25 °C) shows a noticeable down-field shift as the temperature is increased, as is evident in the aromatic region of the spectrum shown in Figure 6-6. A plot of the chemical shift changes shows a melting profile for all shifts consistent with that observed from UV melting experiments. However, only the initial part of the melting curve is accessible in the temperature range possible in the NMR experiment. NMR melting curves (changes in shift versus temperature) are shown in Figure 6-7, with reference to shifts at 298K

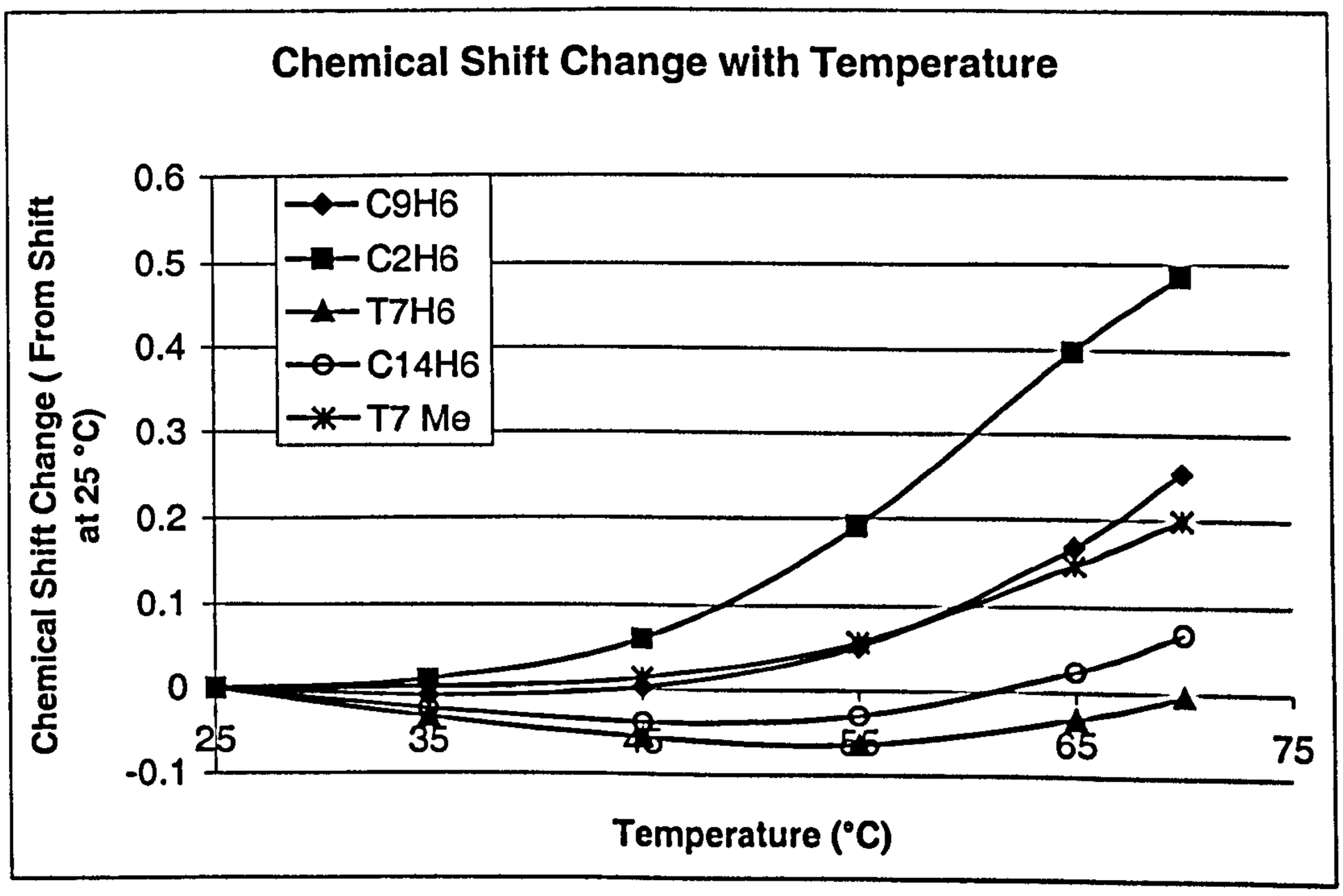


Figure 6-7: Chemical shift changes with temperature. All values are referenced to the chemical shifts obtained at 298K (25 °C).

Differences in the melting of the two hairpins are apparent by monitoring the chemical shifts of the H6 of C2 and C9 which reside in the two different hairpins. While the chemical shifts of these two resonances are similar at 298K, C2 is much more sensitive to temperature than C9, indicating a melting transition at a lower temperature for the hairpin carrying the A1-T7 base pair. Thus, the conclusions above from the UV melting studies that the broadness of the transition may reflect two independent melting processes with different T_m seems to be born out by monitoring the melting transition at the individual residue level.

6.3. Binding of Nogalamycin to the Nicked TpG Site.

The binding of nogalamycin to the double hairpin was monitored initially by NMR titration. A clear set of resonances for a drug-bound complex was evident at low drug:DNA ratios. The free DNA and complex appeared to be in slow exchange at 298K allowing the titration to the fully bound complex to be monitored readily by ^1H NMR titration experiments. The complexity of the spectrum of the 1:1 complex suggested that the drug was binding at a single site and that a single species was present in solution. Subsequently, a complete analysis of the complex was possible from 2D NMR analysis, using the methodology described in earlier chapters. The location of the drug in the complex was only apparent after detailed analysis and virtually complete resonance assignment.

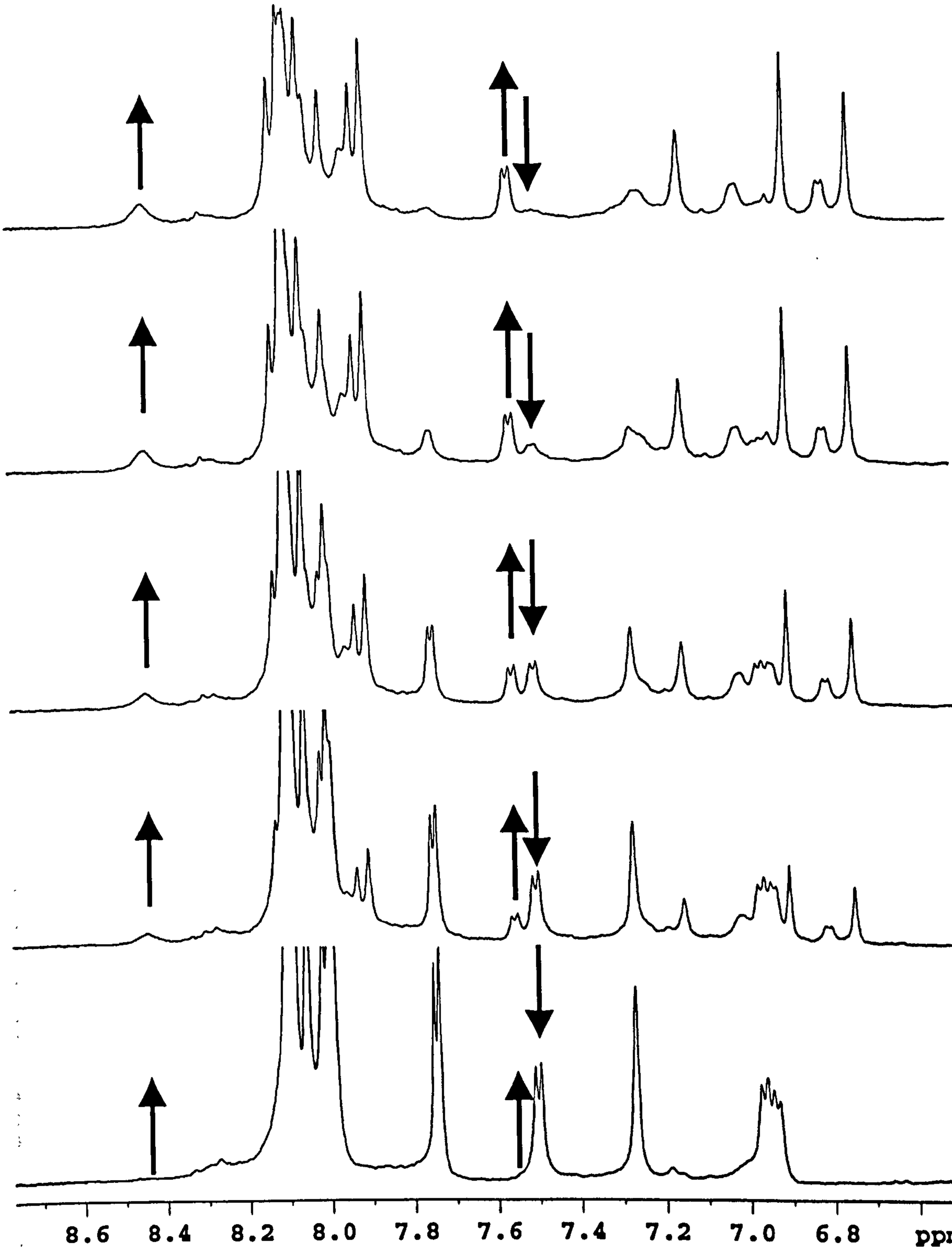


Figure 6-8: 1D NMR spectra of d(ACGAAGTGCGAAGC) with Increasing concentrations of nogalamycin. Signals present in the free DNA (bottom) are gradually replaced by those of complexed DNA (top). The end point corresponds to an approximate 1:1 complex of DNA to drug.

A number of significant chemical shift changes were initially evident. Both the A4 and A11 H4' resonances remain shifted up-field, indicating that the hairpin loops are still intact. The resonances of base pairs at the nick site undergo quite large changes consistent with intercalation at the TpG step. We see a significant stabilisation of the A1-T7 and G8-C14 base pairs as their imino protons appear to be exchanging significantly more slowly in the complex, suggesting stabilisation through stacking with the drug molecule.

UV melting studies carried out in a similar manner to those described for the free hairpin show an increase in melting point to 73 °C an elevation of 4 °C (Figure 6-9). The comparatively low melting point elevation due to nogalamycin binding is presumably due to the already high intrinsic stability of these sequences.

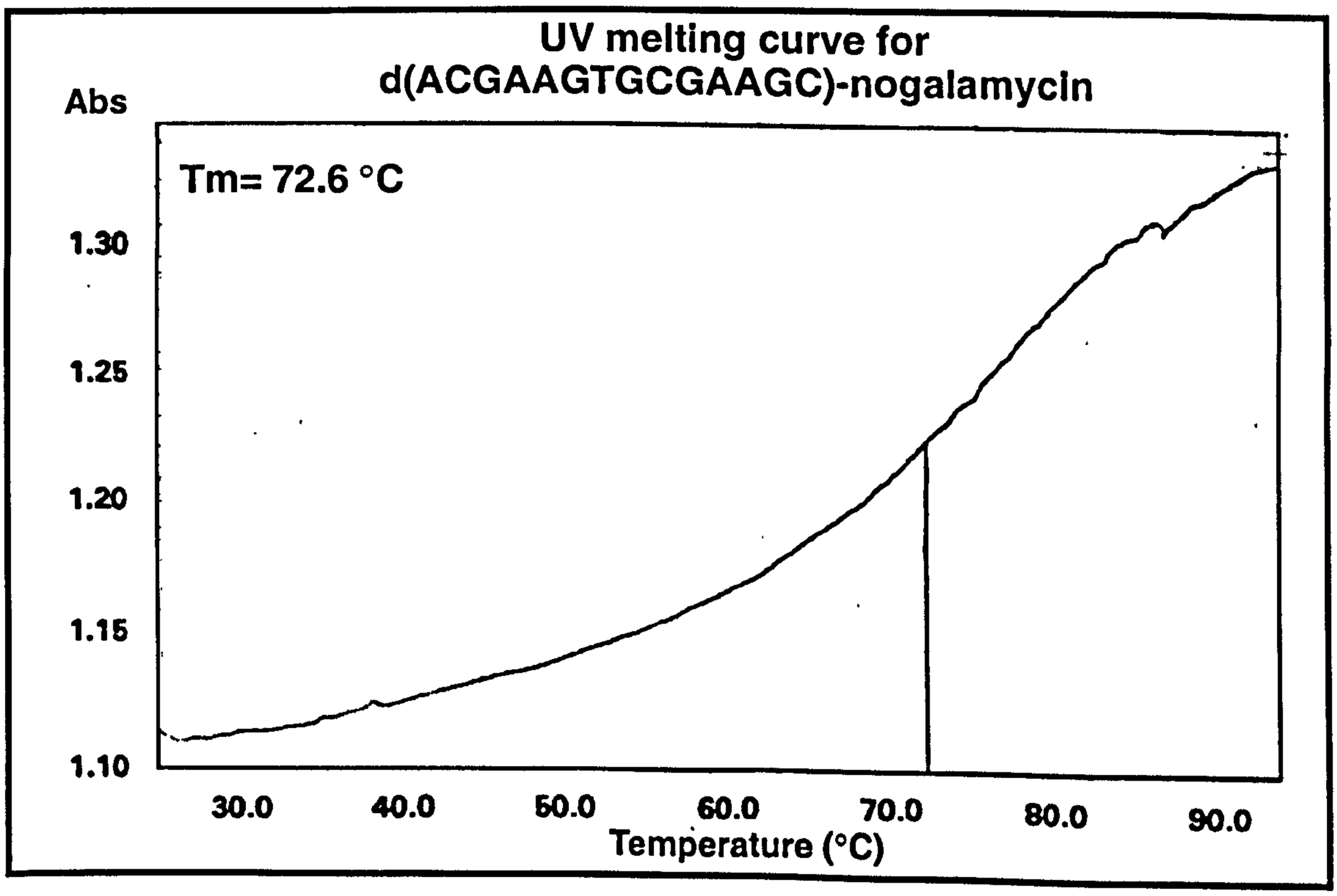


Figure 6-9: UV melting curve for d(ACGAAGTGCGAAGC)-nogalamycin. Showing an elevated melting temperature of 72.6 °C.

A detailed analysis of NOEs identifies interactions across the A1 to C14 step, consistent with stabilisation of the nick site with the A1-T7 and G8-C14 base pairs clamping around the bound drug molecule (Figure 6-10, peak (f)). Many drug to DNA NOEs, have been observed, and these are analogous to those already described for binding at the TpG step. Portions of the NOESY spectrum illustrating resonance assignments and inter-molecular NOEs are shown in Figure 6-10. A full list of drug-DNA NOEs are provided in Table 6-1.

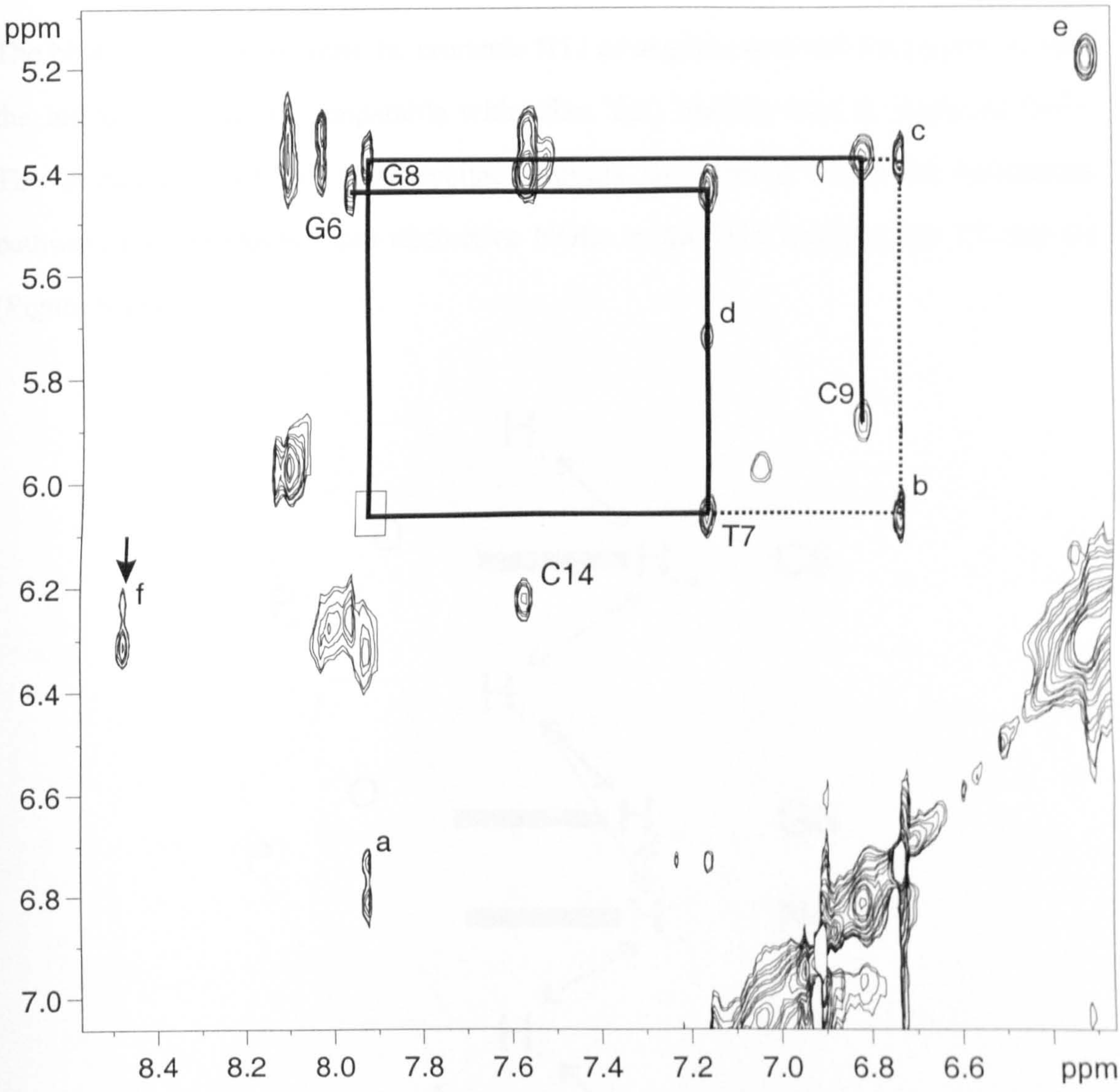


Figure 6-10: NOESY of the H6/8 – H1' region recorded at 288K. Only base intra-residue NOEs are labelled. The sequence of NOE assignments from G6 to C9 is interrupted between T7 and G8 (solid line) by the absence of a T7 H1' → G8 H8 peak (boxed region). An alternative pathway via the nogalamycin H11 proton is consistent with intercalation between the T7 and G8 (dashed line). A number of drug-DNA interactions are identified (a) – (e) these are assigned as follows: (a) H11 → G8 H8; (b) H11 → T7 H1'; (c) H11 → G8 H1'; (d) H1'' → T7 H6; (e) H1' → A1 H1'. The peak (f) is assigned to C14 H1' → A1 H8 indicating a tightening across the nick site.

The NOEs observed between the aromatic H11 of nogalamycin and the protons around the intercalation site is comparable with other TpG binding sites in duplexed DNA. The intercalation of the anthracycline disrupts the normal sequential assignment pathway, but introduces three distinctive NOEs to the H11 bridging the T7 and G8 (Figure 6-11).

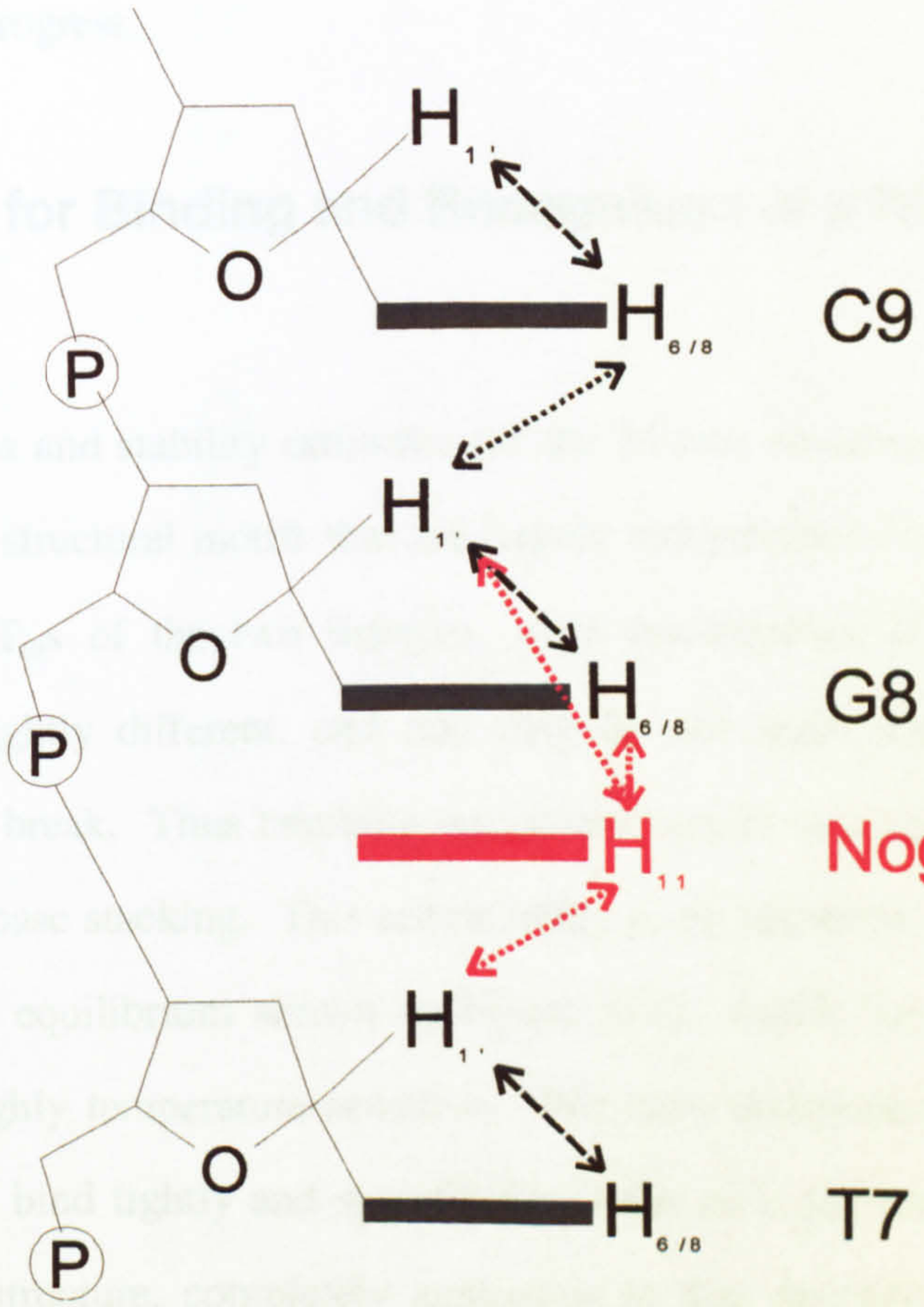


Figure 6-11: Schematic showing predicted NOEs due to intercalation between T7 and G8. Drug DNA NOEs, represented by red arrows, bridge the gap caused by intercalation. These NOEs are observed as peaks (a),(b) and (c) in Figure 6-10.

Several other inter-molecular NOEs are observed and these are summarised in Table 6-1. On the basis of the data, we are able to build a qualitative model of the complex

and can demonstrate that the drug molecule binds in a similar orientation in the TpG site to that described previously. The nogalose sugar is clearly located in the minor groove. The pattern of NOEs from the drug methyl groups confirms this. In the major groove the same pattern of NOEs from the bicyclicamino glucose sugar suggests that the nature of the electrostatic and hydrophilic contacts are also similar. A full structural refinement is in progress.

6.4. A Model for Binding and Recognition at a Nicked Site

Structural analysis and stability estimates for the 14-mer sequence suggest that it folds into two discrete structural motifs that are largely independent of each other. We have shown that the T_ms of the two hairpins, as a consequence of differences in stem sequence, are slightly different, and that they do not stack significantly end-to-end across the strand break. Thus breaking one strand results in a considerable increase in the dynamics of base stacking. This seems likely to be sequence-dependent. Thus, we can consider the equilibrium shown in Figure 6-12, which lies to the left, but also appears to be highly temperature-sensitive. We have demonstrated convincingly that nogalamycin can bind tightly and specifically to the nick site and stabilise a complex with a B-DNA structure, completely analogous to that described in earlier chapters. Thus, the integrity of the covalent connectivities along the DNA sugar-phosphate backbone is not essential for DNA recognition. Thus, as shown in Figure 6-12, nogalmycin is presumably able to bind initially to the fully open state allowing the complex to form around the drug molecule.

Table 6-1: Observed inter-molecular drug-DNA NOEs.

Residue 1	Atom 1	Residue 2	Atom 2	Residue 1	Atom 1	Residue 2	Atom 2
Nog	H6	C9	H6	Nog	H2'	C14	H1'
Nog	H2"	T7	H6	Nog	3'-OCH_3	C14	H1'
Nog	H2"	G6	H8	Nog	5'-CH_3	C14	H5'1
Nog	H1"	T7	H6	Nog	H3'	C14	H5'1
Nog	H2"	T7	H1'	Nog	H1'	C14	H1'
Nog	H6	C9	H1'	Nog	H3'	C9	H1'
Nog	11-OCH_3	T7	H1'	Nog	H1'	A1	H1'
Nog	9-CH_3	G8	H1'	Nog	2'-OCH_3	A1	H1'
Nog	H11	G8	H8	Nog	3'-OCH_3	C14	H1'
Nog	5"-CH_3	T7	CH3	Nog	H2'	C14	H1'
Nog	H6	C9	H6	Nog	H3'	C9	H2'2
Nog	2'-OCH_3	A1	H8	T7	H2"	Nog	H11
Nog	H1'	A1	H1'	T7	H2'	Nog	H11
Nog	2'-OCH_3	C14	H1'	T7	H1'	Nog	H11
Nog	2'-OCH_3	A1	H1'	T7	CH_3	Nog	H1"
Nog	H3'	C9	H1'	T7	CH_3	Nog	5"-CH3
Nog	H3'	G8	H1'	A1	H1'	Nog	H1'
C9	H2'	Nog	H6	C9	H2'	Nog	H6

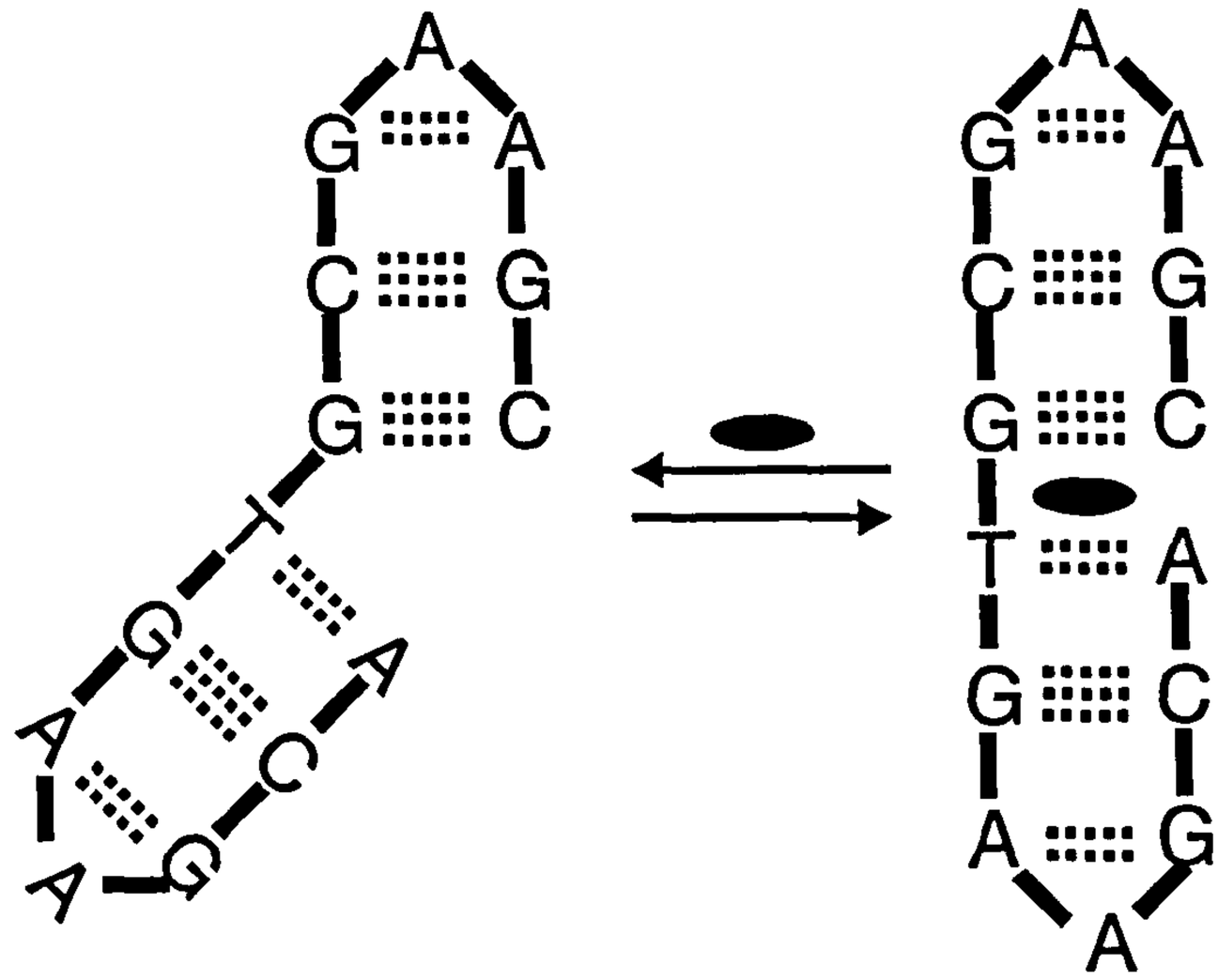


Figure 6-12: Proposed equilibrium with both hairpins formed. Binding of nogalamycin biases the equilibrium to the right.

The observation that the complex and the free DNA are in slow exchange indicates a large kinetic barrier to dissociation that seems likely to require destacking of the drug with adjacent base pairs to allow the drug to dissociate. Thus, the flexible open state of the free DNA provides easy access to the intercalation site. Although we do not have the data to support this, one speculation may be that the association rate constant may be significantly higher in this case because the structure already appears to be largely "premelted", allowing drug access to the thermodynamically preferred binding site.

Finally, in other studies underway in the group we have examined drug binding to the alternative sequence d(GCGAAGCAGAAGT) in which the strand break is now located between the T1 and G14 nucleotides. We find that the drug binds in a similar manner to this sequence, and that the position of the strand break (on the TG or CA strand) has little influence on the outcome of the complex formed.

6.5. Experimental

6.5.1 DNA Synthesis and NMR Sample Preparation

The DNA sequence d(ACGAAGTGCGAAGC) was synthesised and purified as detailed previously in Chapter 3.

The pH of the samples was adjusted to pH 6.8 at the end of the titration. Samples were lyophilised several times from D_2O. The sodium salt of trimethylsilylpropionate was used as an internal reference compound, sodium azide (0.6 mM) as an anti-bacterial agent and EDTA (0.6 mM) to complex any heavy metal ions.

6.5.2 NMR Analysis

NMR data were collected at 500MHz on a Bruker DRX500 spectrometer and processed on an R4600PC Silicon Graphics Indy workstation using XWINNMR software. Standard phase-sensitive 2D NMR pulse sequences were used throughout, including NOESY (States et al. 1982), ROESY (Bax and Davis, 1985b; Bothner-by et al. 1984), DQF-COSY (Piantini et al. 1982; Marion and Wüthrich, 1983; Rance et al. 1983), TOCSY (Braunschweiler and Ernst, 1983; Bax and Davis, 1985a) and WATERGATE NOESY and ROESY (Piotto et al. 1992) for solvent suppression in 90% H_2O solutions. NOESY spectra were acquired at mixing times between 250 and 300 ms for assignment and typically 1024 complex data points were collected for each of 400 t_1 increments with 64 transients for each.

6.6. Bibliography

Bax, A. and Davis, D.G. (1985a) MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy. *J. Mag. Res.* **65**, 355-360.

Bax, A. and Davis, D.G. (1985b) Practical aspects of two-dimensional transverse NOE spectroscopy. *J. Mag. Res.* **63**, 207-213.

Bothner-by, A.A., Stephens, R.L., Lee, J., Warren, C.D. and Jeanloz, R.W. (1984) Structure determination of a tetrasaccharide - transient nuclear overhauser effects in the rotating frame. *J. Am. Chem. Soc.* **106**, 811-812.

Braunschweiler, L. and Ernst, R.R. (1983) Coherence transfer by isotropic mixing: application to proton correlation spectroscopy. *J. Mag. Res.* **53**, 521-528.

Fox, K.R. and Waring, M.J. (1984) Evidence of different binding sites for nogalamycin in DNA revealed by association kinetics. *Biochim. Biophys. Acta* **802**, 162-168.

Hirao, I., Kawai, G., Yoshizawa, S., Nishimura, Y., Ishido, Y., Watanabe, K.A. and Miura, K.-I. (1994) Most Compact Hairpin-Turn Structure Exerted by a Short DNA Fragment, d(GCGAAGC) in Solution: an Extraordinarily Stable Structure resistant to Nucleases and Heat. *Nuc. Acids. Res.* **22**, 576-582.

Marion, D. and Wüthrich, K. (1983) Application of phase sensitive two-dimensional correlated spectroscopy (COSY) for measurements of H-1-H-1 spin-spin coupling- constants in proteins. *Biochem. Biophys. Res. Comm.* **113**, 967-974.

Piantini, U., Sorensen, O.W. and Ernst, R.R. (1982) Multiple quantum filters for elucidating NMR coupling networks. *J. Am. Chem. Soc.* **104**, 6800-6801.

Piotto, M., Saudek, V. and Sklenar, V. (1992) Gradient-tailored excitation for single quantum NMR spectroscopy of aqueous solutions. *J. Bio. Mol. NMR* **2**, 661-665.

Rance, M., Sorensen, O.W., Bodenhausen, G., Wangner, G., Ernst, R.R. and Wüthrich, K. (1983) Improved spectral resolution in COSY H-1-NMR spectra of proteins via double quantum filtering. *Biochem. Biophys. Res. Comm.* **117**, 479-485.

States, D.J., Haberkorn, R.A. and Ruben, D.J. (1982) A two-dimensional nuclear Overhauser experiment with pure absorption phase in 4 quadrants. *J. Mag. Res.* **48**, 286-292.

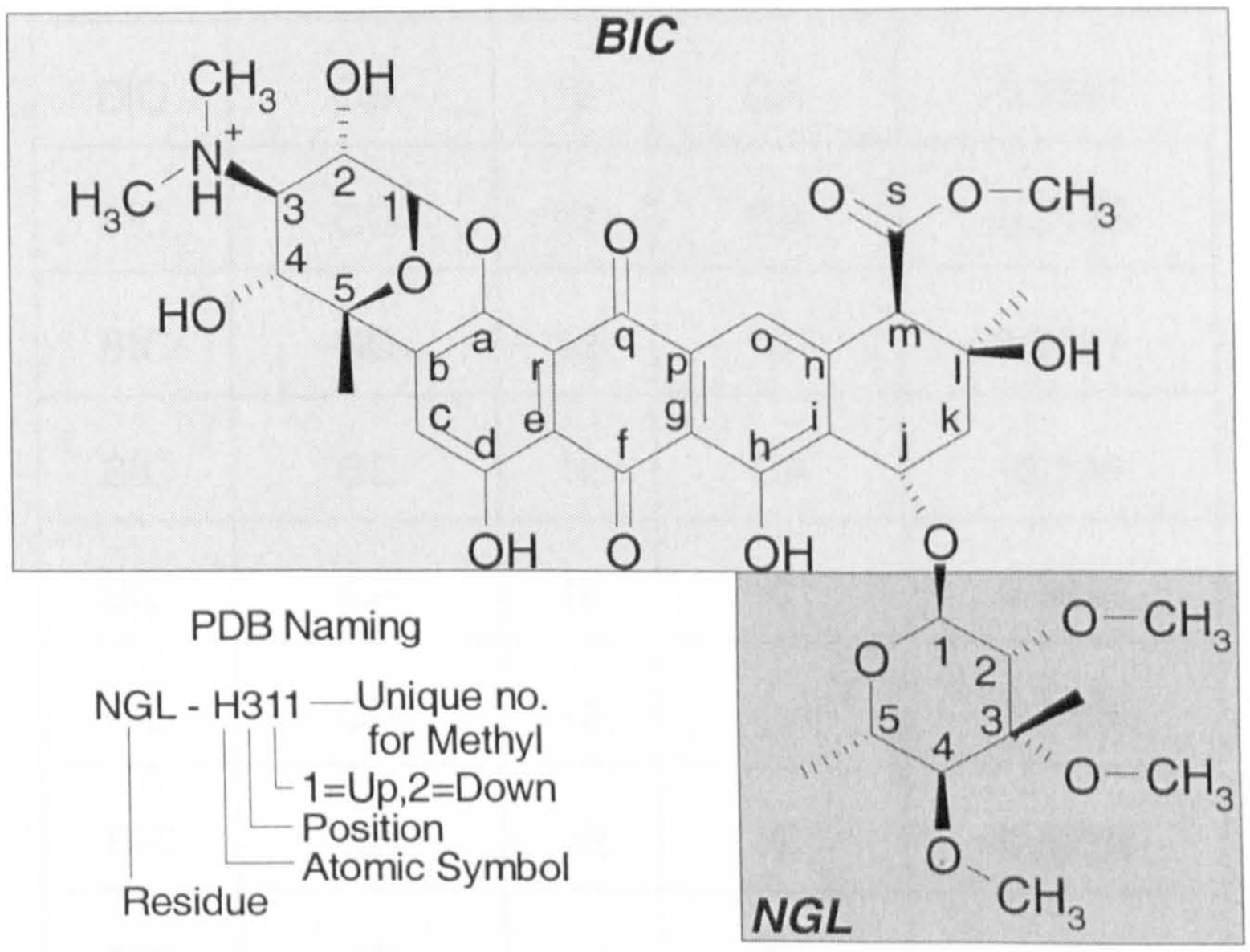
Varani, G. (1995) Exceptionally Stable Nucleic Acid Hairpins. *Annu. Rev. Biophys. Biomol. Struct.* **24**, 379-404.

Yoshizawa, S., Kawai, G., Watanabe, K.A., Miura, K.-I. and Hirao, I. (1997) GNA Trinucleotide Loop Sequence Producing Extraordinarily Stable DNA Minhairpins. *Biochem.* **36**, 4761-4767.

Yoshizawa, S., Ueda, T., Ishido, Y., Miura, K.-I., Watanabe, K.A. and Hirao, I. (1994) Nuclease resistance of an extraordinarily thermostable mini-hairpin DNA fragment, d(GCGAAGC) and its application to *in vitro* protein synthesis. *Nuc. Acids. Res.* **22**, 2217-2221.

Appendix 1: Nogalamycin Partial Charges and Atom Types

Notes on nomenclature: Due to restrictions imposed by the PDB format the following naming convention has been used throughout the appendices to remain consistent with the raw data files.



Residue	Atom Name	At No.	Atom Type	Partial Charge
BIC	C1	12	CT	0.4115
BIC	C2	12	CT	0.0055
BIC	C3	12	CT	-0.0919

BIC	C31	12	CT	-0.168
BIC	C32	12	CT	-0.168
BIC	C4	12	CT	0.1261
BIC	C5	12	CT	0.2618
BIC	C51	12	CT	-0.0831
BIC	CA	12	C	0.3917
BIC	CB	12	CA	-0.1391
BIC	CC	12	CA	-0.2196
BIC	CD	12	C	0.3677
BIC	CE	12	CA	-0.196
BIC	CF	12	C	0.5691
BIC	CG	12	CA	-0.3635
BIC	CH	12	C	0.4766
BIC	CI	12	CA	-0.253
BIC	CJ	12	CT	0.3444
BIC	CK	12	CT	-0.1696
BIC	CL	12	CT	0.307
BIC	CL2	12	CT	-0.1746
BIC	CM	12	CT	-0.5729

BIC	CN	12	CA	0.2359
BIC	CO	12	CA	-0.1295
BIC	CP	12	CA	-0.0752
BIC	CQ	12	C	0.5355
BIC	CR	12	CA	-0.2616
BIC	CS	12	C	0.8988
BIC	CS2	12	CT	0.1477
BIC	H+3	1	H	0.1699
BIC	H11	1	H2	0.1209
BIC	H21	1	H1	0.0689
BIC	H22	1	HO	0.3653
BIC	H311	1	HP	0.1175
BIC	H312	1	HP	0.1175
BIC	H313	1	HP	0.1175
BIC	H32	1	HP	0.1487
BIC	H321	1	HP	0.1175
BIC	H322	1	HP	0.1175
BIC	H323	1	HP	0.1175
BIC	H41	1	H1	0.0605

BIC	H42	1	HO	0.3537
BIC	H511	1	HC	0.058
BIC	H512	1	HC	0.058
BIC	H513	1	HC	0.058
BIC	HC	1	HA	0.1656
BIC	HD	1	HO	0.3574
BIC	HH	1	HO	0.3381
BIC	HJ1	1	H1	0.0644
BIC	HK1	1	HC	0.0766
BIC	HK2	1	HC	0.0897
BIC	HL1	1	HO	0.3545
BIC	HL21	1	HC	0.0439
BIC	HL22	1	HC	0.0439
BIC	HL23	1	HC	0.0439
BIC	HM2	1	HC	0.1976
BIC	HO	1	HA	0.1581
BIC	HS21	1	H1	0.0244
BIC	HS22	1	H1	0.0244
BIC	HS23	1	H1	0.0244

BIC	N3	14	N3	0.3114
BIC	O12	16	OS	-0.4723
BIC	O22	16	OH	-0.5225
BIC	O42	16	OH	-0.5475
BIC	OA	16	OS	-0.3529
BIC	OD	16	OH	-0.5059
BIC	OF	16	O	-0.4059
BIC	OH	16	OH	-0.4775
BIC	OJ2	16	OS	-0.4615
BIC	OL1	16	OH	-0.5674
BIC	OQ	16	O	-0.4804
BIC	OS1	16	O	-0.4841
BIC	OS2	16	OS	-0.5145
NGL	C1	12	CT	0.3161
NGL	C2	12	CT	0.0552
NGL	C22	12	CT	0.1167
NGL	C3	12	CT	0.2966
NGL	C31	12	CT	-0.2746
NGL	C32	12	CT	0.0907

NGL	C4	12	CT	0.1903
NGL	C41	12	CT	0.153
NGL	C5	12	CT	0.1792
NGL	C52	12	CT	-0.2055
NGL	H12	1	H2	0.1142
NGL	H21	1	H1	0.0886
NGL	H221	1	H1	0.0192
NGL	H222	1	H1	0.0192
NGL	H223	1	H1	0.0192
NGL	H311	1	HC	0.0909
NGL	H312	1	HC	0.0909
NGL	H313	1	HC	0.0909
NGL	H321	1	H1	0.0315
NGL	H322	1	H1	0.0315
NGL	H323	1	H1	0.0315
NGL	H411	1	H1	0.0462
NGL	H412	1	H1	0.0462
NGL	H413	1	H1	0.0462
NGL	H42	1	H1	0.0739

NGL	H51	1	H1	0.0628
NGL	H521	1	HC	0.0684
NGL	H522	1	HC	0.0684
NGL	H523	1	HC	0.0684
NGL	O1	16	OS	-0.4473
NGL	O22	16	OS	-0.4047
NGL	O32	16	OS	-0.4275
NGL	O41	16	OS	-0.4528

Appendix 2: Force Field Parameters

The following are additions made to the standard amber 94 atom types to allow parameterisation of nogalamycin. Atom type assignments are listed in Appendix 1

Table A1-1: Bonding

Bonds	Energy (kcal mol^{-1})	Equilibrium Distance (Å)
C –OS	450.0	1.376

Table A2-2: Additional Bond Angles

Angle	Energy (kcal mol^{-1})	Equilibrium Angle (°)
OS-CT-OS	60.0	109.50
O -C -OS	80.0	120.40
C -OS-CT	60.0	109.50
CA-CT-OS	50.0	109.50
CA-C -OS	70.0	120.40
C -CA-C	63.0	120.00
C -CA-CT	63.0	120.00

CA-C -O	80.0	120.00
C -CA-CT	63.0	120.00
C -CA-C	63.0	120.00
CA-CT-C	40.0	109.50
CT-C -OS	70.0	117.00
CA-CT-H1	50.0	109.5

Table A2-3: Torsion Angles.

Torsion Angle	Phase	Potential Energy (kcal mol^{-1})	Angle Offset (°)	Periodicity
O -C -OS-CT	1	2.50	180.0	2
CA-C -OS-CT	1	0.07	180.0	2
CT-C -OS-CT	1	0.07	180.0	2

Appendix 3: Restraints for d(ATGCAT)$_2$-Nogalamycin

Residue	Atom Name	Residue	Atom Name	Upper-Bound Restraint (Å)
NGL	H31*	NGL	H51	3.66
NGL	H52*	NGL	H51	3.68
BIC	HL2*	NGL	H51	3.93
NGL	H31*	BIC	HK2	4.01
NGL	H52*	NGL	H42	3.77
BIC	HL2*	BIC	HK2	4.78
BIC	HL2*	BIC	HK1	4.70
NGL	H52*	BIC	HK1	5.30
NGL	H42	NGL	H41*	3.78
NGL	H52*	NGL	H41*	3.77
BIC	H51*	BIC	HC	3.48
BIC	H51*	BIC	H11	4.83
BIC	HL2*	BIC	HO	4.98
NGL	H32*	NGL	H22*	4.15
BIC	HN2*	BIC	H41	4.28
BIC	HN1*	BIC	H41	3.58

BIC	*HN1**	BIC	H21	3.67
BIC	HN2*	BIC	H21	4.20
BIC	HN2*	BIC	H32	5.30
BIC	HN1*	BIC	H32	4.59
BIC	HN2*	BIC	H11	4.78
NGL	H31*	NGL	H42	4.73
NGL	H31*	NGL	H32*	3.81
BIC	HN1*	BIC	H11	4.78
NGL	H31*	NGL	H12	5.07
NGL	H32*	NGL	H12	4.78
NGL	H31*	NGL	H42	4.78
BIC	HL2*	BIC	HM2	3.62
NGL	H22*	BIC	HJ1	5.33
NGL	H31*	DC	H1'	3.86
BIC	HL2*	DC	H1'	4.83
BIC	HL2*	DG	H1'	3.90
NGL	H32*	DC	H1'	4.98
NGL	H32*	DA	H1'	4.41
NGL	H22*	DA	H1'	4.29

NGL	H22*	DC	H1'	3.73
BIC	HS2*	DT	H1'	4.53
BIC	HS2*	DT	H6	4.63
NGL	H31*	BIC	HH	4.06
BIC	H41	BIC	H21	3.28
BIC	H21	BIC	H32	2.76
BIC	H41	BIC	H32	3.11
BIC	HK1	BIC	HK2	2.21
BIC	HJ1	NGL	H12	2.44
BIC	H32	BIC	H11	3.70
BIC	H21	BIC	H11	2.38
BIC	H41	BIC	H11	4.33
BIC	H41	BIC	HC	4.24
NGL	H42	NGL	H51	3.05
BIC	HK2	NGL	H51	3.84
BIC	HM2	BIC	HO	2.65
NGL	H21	NGL	H12	2.71
BIC	HK2	BIC	HJ1	4.75
BIC	HK1	NGL	H12	3.64

BIC	HK2	NGL	H12	2.68
DC	H2'	DC	H6	3.00
DT	H2'	DT	H6	3.05
DC	H2"	DC	H6	4.16
DG	H2'	DC	H6	5.00
DA	H2'	DT	H6	5.00
DT	H2"	DT	H6	5.00
DT	H2'	DT	H6	5.00
DA	H2'	DT	H6	5.00
DA	H2"	DT	H6	3.32
DA	H2"	DT	H6	2.87
DT	H2'	DT	H6	3.19
DT	H2'	DT	H6	2.75
DA	H2'	DT	H6	5.00
DT	H2"	DT	H6	5.00
DA	H2'	DT	H6	3.64
DA	H2"	DT	H6	3.45
DA	H2"	DT	H6	3.03
DC	H2"	DC	H6	3.63

DC	H2'	DC	H6	5.00
DG	H2'	DC	H6	5.00
DG	H2"	DC	H6	2.92
DG	H1'	DC	H6	4.03
DC	H1'	DC	H6	4.64
DC	H5	DC	H6	2.94
DT	H1'	DT	H6	5.00
DT	H1'	DT	H6	4.13
DA	H1'	DT	H6	4.27
DA	H1'	DT	H6	4.09
DT	H1'	DT	H6	5.00
DA	H1'	DT	H6	5.00
DT	H1'	DT	H6	5.00
DC	H1'	DC	H6	5.00
DG	H1'	DC	H6	2.58
DC	H5	DC	H6	3.13
DC	H3'	DC	H6	4.55
DT	H3'	DT	H6	4.25
DT	H3'	DT	H6	3.46

DT	H2'	DG	H8	4.46
DT	H2"	DG	H8	5.00
DG	H2'	DG	H8	5.00
DG	H2"	DG	H8	3.69
DT	H2"	DG	H8	5.00
DG	H2'	DG	H8	5.00
DG	H2"	DG	H8	5.00
DT	H2'	DG	H8	5.00
DA	H2'	DA	H8	3.94
DA	H2"	DA	H8	5.00
DC	H2'	DA	H8	4.25
DC	H2"	DA	H8	3.69
DA	H2'	DA	H8	3.01
DA	H2"	DA	H8	3.67
DA	H2'	DA	H8	3.17
DA	H2"	DA	H8	4.28
DC	H2"	DA	H8	3.85
DC	H2'	DA	H8	5.00
DA	H2'	DA	H8	5.00

DA	H2"	DA	H8	4.81
DC	H1'	DA	H8	3.90
DA	H1'	DA	H8	4.51
DA	H1'	DA	H8	3.89
DA	H1'	DA	H8	4.06
DC	H1'	DA	H8	4.08
DA	H1'	DA	H8	3.54
DG	H1'	DG	H8	4.83
DC	H5	DG	H8	4.00
DT	H1'	DG	H8	3.79
DG	H1'	DG	H8	4.32
DA	H3'	DA	H8	4.06
DA	H3'	DA	H8	3.98
DA	H3'	DA	H8	4.37
DT	H7*	DA	H8	4.30
DT	H7*	DA	H8	3.92
DT	H7*	DA	H8	5.04
DT	H7*	DA	H8	4.16
DT	H7*	DT	H6	3.73

DT	H7*	DT	H6	3.63
DT	H7*	DT	H6	3.75
DT	H7*	DT	H6	3.90
DA	H2'	DA	H1'	2.70
DA	H2"	DA	H1'	2.12
DA	H2'	DA	H1'	2.79
DA	H2"	DA	H1'	2.28
DA	H2'	DA	H1'	2.90
DA	H2"	DA	H1'	2.28
DT	H2'	DT	H1'	2.01
DA	H2'	DA	H1'	2.84
DA	H2"	DA	H1'	2.12
DT	H2"	DT	H1'	2.10
DT	H2'	DT	H1'	2.51
DC	H2"	DC	H1'	2.17
DC	H2'	DC	H1'	2.60
DG	H2'	DG	H1'	3.01
DG	H2"	DG	H1'	2.29
DT	H2'	DT	H1'	3.17

DT	H2"	DT	H1'	2.32
DC	H2'	DC	H1'	3.27
DC	H2"	DC	H1'	2.19
DG	H2'	DG	H1'	4.20
DG	H2"	DG	H1'	4.20
DA	H4'	DA	H1'	3.28
DA	H5'	DA	H1'	4.20
DT	H4'	DT	H1'	2.60
DT	H4'	DT	H1'	2.64
DA	H4'	DA	H1'	3.17
DA	H4'	DA	H1'	4.20
DT	H4'	DT	H1'	5.00
DC	H4'	DC	H1'	2.66
DG	H4'	DG	H1'	2.89
DG	H4'	DG	H1'	3.32
DA	H4'	DA	H3'	2.97
DG	H4'	DG	H3'	5.00
DT	H4'	DT	H3'	5.00
DA	H4'	DA	H3'	2.07

DA	H4'	DA	H3'	2.58
DA	H5'	DA	H3'	5.00
DC	H4'	DC	H3'	5.00
DA	H5'	DA	H3'	5.00
DT	H4'	DT	H3'	2.65
DT	H4'	DT	H3'	2.66
DA	H5'	DA	H4'	5.00
DA	H5'	DA	H4'	2.41
DA	H5'	DA	H4'	2.87
DA	H2'	DA	H2"	2.21
DA	H2'	DA	H2"	2.43
DA	H2'	DA	H2"	1.99
DA	H2'	DA	H2"	2.06
DT	H2'	DT	H2"	2.84
DC	H2'	DC	H2"	2.09
DC	H2'	DC	H2"	1.92
DT	H7*	BIC	H11	2.74
DT	H2"	BIC	HO	3.46
DT	H2'	BIC	HO	3.11

BIC	HM2	DG	H1'	2.83
NGL	H12	DA	H1'	2.94
DG	H4'	BIC	HO	3.97
BIC	HJ1	DA	H1'	2.87
DG	H1'	BIC	HO	3.78
DT	H1'	BIC	HO	3.23
BIC	HO	DT	H6	4.37
BIC	HO	DG	H8	5.00
DC	H5	DC	H6	2.94
DC	H5	DC	H6	2.73
DC	H2"	BIC	HH	4.05
DA	H1'	BIC	HH	5.00
DA	H8	BIC	HH	4.81
BIC	HK2	BIC	HH	3.83
NGL	H21	BIC	HH	3.80
BIC	HC	BIC	HD	4.20
BIC	HH	BIC	HD	4.84
BIC	HJ1	BIC	HH	4.68

Appendix 4: Chemical Shift Assignments for $d(ATGCAT)_2$

Table A4-2: Assigned chemical shifts (ppm) for $d(ATGCAT)_2$[a]

	H6/H8	H2/H5/CH_3	H1'	H2'	H2"	H3'	H4'
A1	8.043	7.797	6.102	2.625	2.769	4.83	---
T2	7.29	1.295	5.714	2.107	2.412	4.824	---
G3	7.886	---	5.884	2.632	2.684	4.984	4.379
C4	7.365	5.404	5.695	2.002	2.413	4.826	4.403
A5	8.268	7.81	6.265	2.733	2.868	4.999	4.416
T6	7.219	1.431	6.087	2.214	---	4.544	4.026

a. *All data referenced to internal trimethylsilylpropionate at pH 7.0, 288K, 100 mM sodium chloride, 10 mM sodium phosphate buffer.*