NMR Studies of Nucleic Acids as Drug Targets

Cathal T. Gallagher

Thesis submitted to the University of Nottingham
for the degree of Doctor of Philosophy,
April 2004.

Contents

Abstract

Nogalamycin is a member of the anthracycline family of antitumour antibiotics. These are potent cytotoxic agents and are routinely used in cancer chemotherapy. Though nogalamycin is clinically insignificant, it does exhibit three distinct types of non-covalent binding to DNA. Since most other anthracyclines bind to DNA by only one or two of these mechanisms, nogalamycin is an excellent model with which to probe the interaction of this class of anti-tumour agents with DNA.

Here, we investigate the binding orientation and stoichiometry of nogalamycin in adjacent TpG(CpA) (and CpG(CpG)) intercalation sites using a combination of NMR techniques and NOE-restrained molecular dynamics simulations. These methods are also employed to investigate the structure of G*N*A hairpin loops, which are considered to have important biological functions, and assess how their structure and stability are influenced by the introduction of nogalamycin at an adjacent site. The effect of nogalamycin on extrahelical thymine bases incorporated onto either face of the intercalation sites is also investigated in this context.

Binding of quadruplex-specific antibodies to telomeric DNA in *Stylonychia lemnae* macronuclei has recently been detected using immunofluorescence, providing direct evidence for the formation of quadruplex DNA structures *in vivo*. Guanine-rich quadruplex structures have been extensively studied by NMR and x-ray crystallographic methods. Previous structural studies have failed to unambiguously resolve the conformation preferred by less-stable A-tetrads incorporated into DNA quadruplexes. Additionally, little effort has been made to address the exact number of ions bound to these adenine-containing structures. This forms the basis of our study into quadruplex DNA.

Finally, we endeavour to investigate the extent of hydration of both duplex and quadruplex structures using rMD methods, and to comapre hydration patterns in the liquid- and solid-state.

Acknowledgements

I would like to thank all those who proffered their time and support over the last three years.

Firstly, to the chief, Prof. Mark Searle, for the opportunity to gain a PhD, for his continual support and advice, and for making me feel left out while other students complained about their supervisors.

My thanks go out to all members of the Searle group, past and present, but especially to: Dr Huw Williams for allowing me to benefit from his extraordinary level of technical expertise, and for keeping the silicon-based group members in working order; Dr Michelle Colgrave and Dr Evripidis Gavathiotis, whose doctoral studies into DNA hairpins and quadruplex structures, respectively, formed the starting point for my own studies; and to my final year project students over the last three years: Stephen Simpson, Richard Grant and David Jarrom.

I am grateful to my parents, for their love and support, and for limiting their interest in my working life to regularly checking that I am happy.

Finally, I would like to acknowledge the EPSRC for funding my work, John Keyte for oligonucleotide synthesis, and the University of Nottingham Ju Jitsu club for giving me something to punch when things weren't going well.

Abbreviations and Symbols

1D	-	One-dimensional
2D	-	Two-dimensional
3D		Three-dimensional
A	-	Adenine
C	-	Cytosine
COSY	-	Correlated spectroscopy
CPG	-	Controlled pore glass bead
dA	-	Deoxyadenosine
dC	-	Deoxycytidine
dG	-	Deoxyguanidine
DNA	-	Deoxyribonucleic acid
ds	-	Double stranded
DQF	-	Double quantum filtered
dT	-	Deoxythymidine
EDTA	-	Ethylenediaminetetraethanoic acid
FID	-	Free induction decay
FT	-	Fourier transformation
G	-	Guanine
MD	-	Molecular dynamics
NMR	-	Nuclear magnetic resonance
NOE	-	Nuclear Overhauser effect
NOESY	-	Nuclear Overhauser effect spectroscopy
ppm	-	Parts per million
RF	-	Radio frequency
rMD	-	Restrained molecular dynamics
ss	-	Single stranded
T	-	Thymine
t1	-	First time domain
T1	-	Longitudinal relaxation
t2	-	Second time domain

T2	-	Transverse relaxation
TOCSY	-	Total correlation spectroscopy
UV	-	Ultraviolet

1. NMR Spectroscopy in Drug Discovery

1.1 The Global Pharmaceutical Industry

In 2003, the combined worldwide pharmaceutical market was US$491.8 billion (*1*). According to the PhRMA annual report 2000-2001 (*169*), pharmaceutical companies have almost doubled their Research & Development spending every five years since 1980. For every $5 of sales revenue, $1 is reinvested in R&D. Pharmaceutical companies need to ensure that there is a supply of new and better drugs in the pipeline to return on the substantial investments made in R&D. It has been estimated that, in order to maintain their positions within the marketplace, the top ten pharmaceutical companies need 4-5 new drugs approved each year (*1*). Most fall far short of this target, with only 1-2 new drugs receiving a marketing authorisation each year.

1.2 Approaches to Drug Discovery

The premise for the rational approach is that drug discovery based on knowledge of the three-dimensional structure of the chosen receptor molecule would reveal potential binding sites for drug molecules. Structural information and modelling data are used to design or refine a drug that will fit precisely within the binding site, like a key in a lock. Such information about the receptor structure can significantly improve the probability of obtaining a successful drug, while eliminating unlikely candidate structures at an early stage of the drug discovery process. In the 1980s, modern drug discovery was dominated by the idea of rational drug design (*228*).

With the emergence of combinatorial chemistry in the early 1990s came the promise of billions of molecules. This pledge swung the pendulum away from the notion of designing one perfect molecule towards one of making many, many molecules and screening them all to find a small number of viable leads. Initial combinatorial libraries consisted exclusively of peptides (*68;195*).

However, once a small number of organic libraries were prepared (*25;26;51;191*) this technology was enthusiastically embraced by the pharmaceutical industry at large.

By the end of the 1990s, the impracticality of synthesising and cataloguing such a vast number of molecules began to become very apparent, causing the pendulum to swing back in the direction of rational drug design. The idea of producing smaller subsets of directed combinatorial libraries was postulated in 1995 (*142*). The aim of this was to design smaller libraries, with a size range of 100-1000 molecules, which possess biological activities against specific targets. During this period, some groups began experimenting with novel approaches to discovering leads *in silico*, a process that is now known as 3D database searching. 3D database searching involves the use of a *virtual library* of computational descriptions of drug molecules that any given combinatorial reaction scheme can produce. The 3D structure of the drug-binding site can be used to select those molecules that best satisfy the restraints of the pharmacophore, and these molecules can be synthesised in vitro and submitted for biological assay. This is the currently a favoured method of drug design, which it is hoped will yield the greatest number of viable new drugs.

The standard techniques used for 3D structure determinations are x-ray crystallography and NMR spectroscopy. Modelling of potential drug receptors and drug-receptor complexes is studied using *in silico* molecular dynamics methods.

1.3 X-ray Crystallography

In order to determine the structure of drug compounds or drug-receptor complexes using x-ray crystallography, it is first necessary to have these molecules or composites available in crystal form. This can be a major drawback, as the process of obtaining crystals of DNA and proteins can be

difficult, involving techniques such as dialysis, vapour diffusion and interface diffusion (*55*). Techniques are being improved, however, and more than 13000 protein structures have been solved by crystallographic methods at the time of this writing. Once the macromolecule has been crystallised, and the crystal assessed for quality, a diffraction pattern is collected, from which a 3D electron density map can be calculated. This can be refined to give rise to a final structural model of the biopolymer being studied. The very nature of crystallisation, however, means that the observed molecules are frozen in space, like a camera snapshot, rather than in the natural liquid state as found in biological tissues.

When a ligand can be co-crystallised with its receptor molecule, the binding site can be easily determined within the complex. In instances where it is not possible to insert the ligand, the active site can often be inferred by comparison with other known structures. Once the structure of the active site is known, potential drug molecules can be generated using computation chemistry methods. This technique of inferring the location of the active site does not allow for conformational changes that may be induced in the target molecule by ligand binding. Examples of some clinically significant drugs discovered using x-ray crystallography include amprenavir (Agenerase®, GSK) and nelfinavir (Viracept®, Roche Products Ltd.), which were discovered by studying the interactions of potential drug compounds using crystal structure of both HIV protease (*236;245*) and lead compounds complexed with the protein (*130;208;209;222;259*). The anti-influenza drugs zanamavir (Relenza®, GSK) and oseltamavir (Tamiflu®, Roche Products Ltd.) were developed with extensive modelling of the crystal structure of influenza virus neuraminidase (*44;132;184;231*). Though surface neuraminidses are known to undergo mutations, which account for different strains of flu viruses, gene sequencing studies have revealed a conserved portion that does not mutate (*43;235*). X-ray crystallography revealed the structure of the conserved portion, and drug molecules were designed to fit into the cleft and prevent

neuraminidase form binding to sialic acid, which forms an important part of the life-cycle of the virus (*130*).

1.4 NMR Spectroscopy

Compared with x-ray crystallography, NMR has the advantage of being carried out in solution rather than requiring crystal samples. Since solution states are more representative of the native environment of DNA and protein drug targets, NMR is more amenable to the study of drug-receptor interactions. Macromoleclar structure determination using NMR spectroscopy typically involves four steps: Following initial recording of NMR spectra, correlations between atoms and resonance peaks are established by means of spectral assignment. The assigned spectra are then used to generate distance and angular restraints between atoms, which allow the three-dimensional structure to be calculated.

NMR is limited to molecules with molecular weights of <35 kDa due to the phenomenon of spectral overlap, which is effectively a function of spectral resolution. Since a Watson-Crick DNA base pair has approximately 24 protons, a DNA sequence containing 20 base pairs can result in a 1D proton NMR spectrum with more than 480 peaks. In such a case, assignment of the one-dimensional spectrum would prove impossible. Solutions to this problem include using higher magnetic field strengths, and spreading the spectrum into two or more dimensions. This size limitation has meant that NMR has been less useful than crystallography in rational drug design, as drugs have traditionally been targeted towards cellular proteins. However, with the emergence of gene therapy, more and more therapeutic agents act at the level of nucleic acids. The 3D structure of nucleic acids are primarily determined by the sequence of the bases on a local level, and so these biopolymers are more amenable to being broken down into smaller structural models than proteins, which often rely on long-range interactions between different domains to determine their tertiary structure. NMR methods have recently

been employed to describe the interaction and stabilisation of quadruplex DNA by a fluorinated polycyclic methylacridinium salt (*71;72*), which had previously been shown to inhibit the enzyme telomerase in human tumour cells through a mechanism thought to involve stabilization of four-stranded G-quadruplex structures formed by single-stranded telomeric DNA (*82;93*).

The techniques of x-ray crystallography and NMR spectroscopy, when combined, can provide invaluable information for drug designers. The precise binding topologies that can be generated using both techniques, together with dynamic properties derived from NMR, can be used to tailor-make drug molecules to fit is specific binding sites.

1.5 Scientific Collaboration

As the scope of what science as a whole can understand increases, so the extent of what individual scientists, limited by finite human memory and intelligence, can consider to be their field of expertise must decrease. A little more than a century ago, an individual worker could expect to have an in-depth knowledge of the extent of knowledge in chemistry, physics and biology. As the depth of current knowledge increases, so fields of expertise must become more specialised. What was once chemistry is now split into organic, inorganic, physical, biological and computational sub-categories, to name but a few. As we gain more and more knowledge about each of these facets of chemistry, so each subcategory can itself be divided further into new subject areas, such as computational surface chemistry. This increasing diversification and specialisation is not only evident in chemistry. Indeed, similar trends can be observed across all disciplines of science. In order that scientific endeavour does not become compartmentalised, it is essential that there is communication and collaboration between scientists working in each of these ever sub-dividing fields.

Within the field of cancer research, for example, an individual worker cannot be expected to possess the expertise to design, optimise and synthesise a potential chemotherapeutic drug, to conduct pharmacokinetic, pharmacodynamic and toxicological studies, and pre-clinical and clinical trials, to design a delivery system for the finished drug, to implement procedures, processes and equipment for large-scale commercial production, and to oversee the administration of the finished product. However, what is expected of each the hundreds of individual scientists involved in each step of this process is that they can communicate what they have learned and achieved to the other scientists involved in the project, each of whom may have very diverse scientific backgrounds. Most importantly, they must be able to effectively communicate their findings with those scientists involved in the steps directly before, and after, their own, as it is typically to these workers that their data may prove to be most useful. For those scientists, predominantly x-ray crystallographers and NMR spectroscopists, involved in structure determination of biological targets, it is vital to be able to relay structural data, in a usable form, to synthetic chemists and others who wish to rationally design, optimise and synthesise drugs specific to these targets. This is facilitated by the existence of a simple, standardised format, which can be used to describe all the relevant 3D topological data from macromolecular structure determinations.

1.6 The Protein Data Bank

The Protein Data Bank (PDB) is an archive of experimentally-determined three-dimensional structures of biological macromolecules, serving a global community of researchers, educators, and students. The PDB began during the late 1960s and early 1970s with community discussions about the need for such a resource (*17*). Protein crystallography was still in its infancy, but it was apparent to the producers of these structures as well as the potential users that every structure contained valuable information that needed to be archived and maintained. In October 1971, an article appeared in *Nature New Biology* (*5*)

announcing the formation of the PDB. The PDB file format was established early on (*18*) to contain the coordinates and related information, and still endures. The archives contain atomic coordinates, bibliographic citations, primary and secondary structure information, as well as crystallographic structure factors and NMR experimental data, all presented in the standardised PDB format, which is accessible to all members of the scientific community. The PDB is the single worldwide repository for the processing and distribution of 3D structure data of large molecules of proteins and nucleic acids and, as such, is in invaluable tool in rational drug design.

There are standardised procedures for depositing and exchanging three-dimensional structural data derived by NMR spectroscopy. Once deposited by a contributor, data are subjected to a series of computerised validation procedures: all the nomenclatures of the atoms and residues are standardised; the sequence records are checked for consistency and are checked against the sequence databases; the geometry of the macromolecule is checked against known standards for distances, angles, and chirality; stereochemical clashes and the chemistry of the small molecule ligands are checked; and the coordinate data are checked against the experimental data. It is the goal of every worker involved in macromolecular 3D structure determination to produce structures that meet the stringent criteria for inclusion in the archives of the PDB.

2. Binding Studies of Nogalamycin with DNA

2.1 Introduction

2.1.1 Anthracycline Antibiotics

In order for replication to occur, the nuclear DNA must first become dissociated from histones. After replication has occurred, it must then be returned to its tightly packed state. This constant winding and unwinding can lead to supercoiling; analogous to the flex on a cheap and often-used telephone. A family of enzymes called topoisomerases, of which there are two members, removes supercoiling. Type I topisomerases transiently cleave a single strand of DNA to allow passage of a second strand through the break, relieving supercoiling (*9*). Type II topisomerases perform a different reaction, cleaving and opening one DNA duplex, passing a second duplex through the opening, and resealing the break (*15;113*).

The mechanism by which anthracycline antibiotics kill the cancer cells is not fully understood, but the main mechanism is thought to be topoisomerase mediated strand breakage (*67;85*). The binding of the anthracycline antibiotics to the DNA is thought to stabilise the cleaved state of the DNA so that it is not resealed, resulting in a permanent double strand break, leading to cell death. Topoisomerases are most active during cell cycle periods of high DNA synthesis. Generally speaking, the topisomerase concentration of proliferating tumor cells is elevated substantially over that found in normal somatic cells. This makes topoisomerases a semi-selective target for cancer chemotherapy.

The presence of a glycoside ring at the C7 position of an anthracycline antibiotic, has been shown to play an important role in affecting topoisomerase specific poisoning (*202*). Those agents possessing a sugar ring, such as nogalamycin, daunomycin and doxorubicin generally poison topisomerase I, but not topisomerase II. In contrast, anthracyclines such as menogaril, a derivitave of nogalamycin without the nogalose sugar, are selective type I topisomerase poisons.

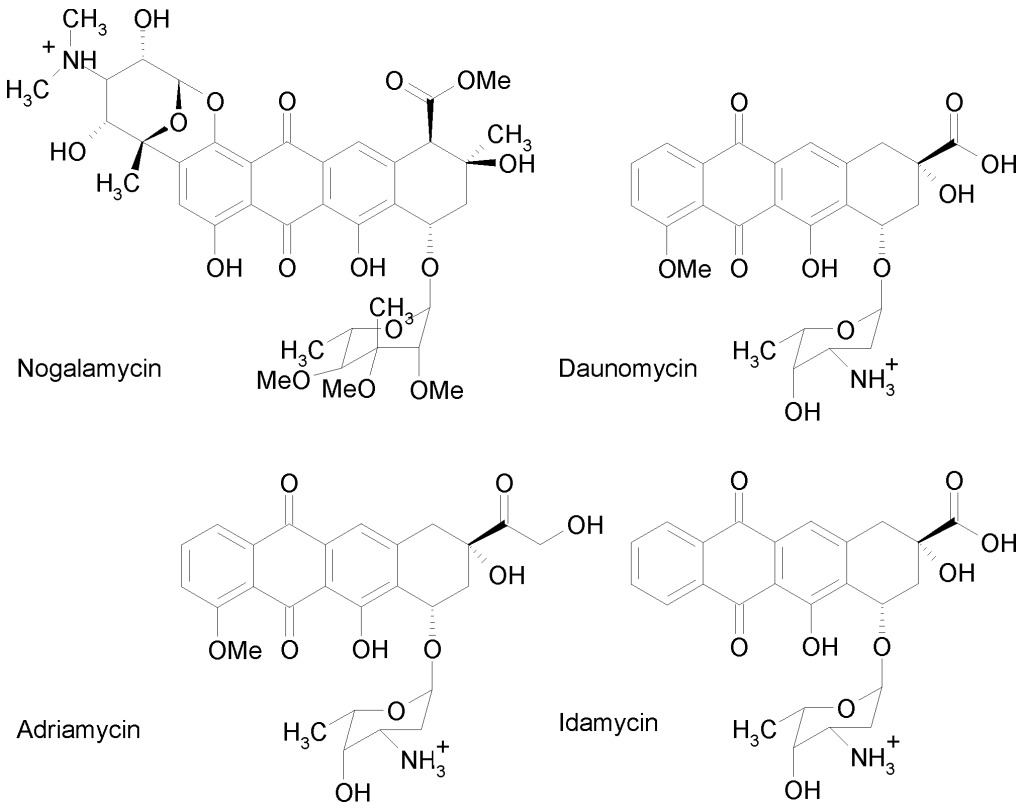


Fig.2.1: Chemical structures if the anthracycline anti-tumour antibiotics.

2.1.2 Nogalamycin

Nogalamycin (Fig.2.2) has been reported to show activity against experimental tumours *in vivo*, but is useless as a clinical chemotherapeutic agent due to the extent of the associated cardiotoxicity (*197*). Isolated from the species *Streoptomyces nogalator*, nogalamycin is unique in that it possesses bulky sugars (nogalose and aminoglucose) on rings A and D respectively of its tetracyclic chromophore (*7*), and yet still binds strongly to double helical DNA by intercalation. The nogalose sugar is attached to the C7 position of ring A, whilst the aminoglucose sugar is attached to the C1 and C2 positions of ring D of the aglycone chromophore. Nogalamycin forms a stable intercalation complex with slow association and dissociation rates (*64;65*).

Though the drug itself is clinically insignificant, the binding mechanism (*212;240;241*), stoichiometry and selectivity (*64;65*) of nogalamycin has been extensively studied.

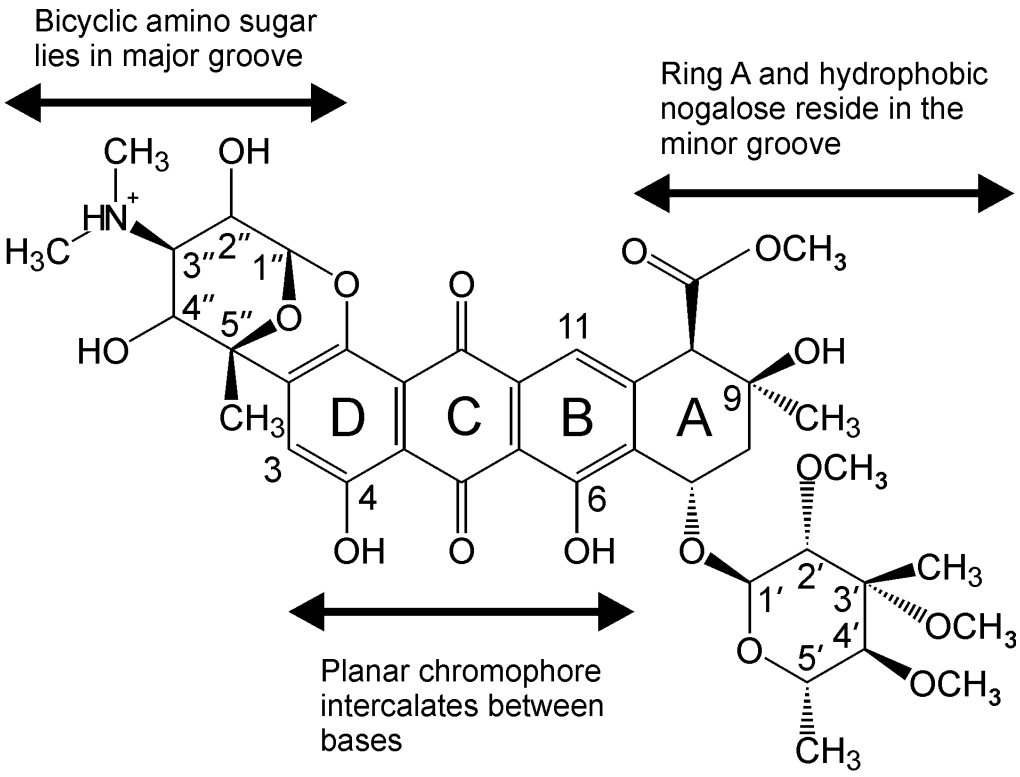


Fig.2.2: Chemical structure of nogalamycin, showing the numbering system employed on the planar chromophore, and the aminoglucose and nogalaose sugars.

Nogalamycin exhibits three distinct types of non-covalent binding to B-form duplex DNA. In common with all anthracyclines, the planar aglycone ring intercalates between stacked bases, causing localised unwinding of the helix and distortion of the bases (*243*). The nogalose sugar resides within the minor grove and the amino sugar in the major groove. Since most other anthracyclines bind to DNA by intercalation and either major or minor groove binding, nogalamycin is an excellent model with which to probe the interaction of this class of anti-tumour agents with duplex DNA.

A number of studies have been carried out to show the relationship between the structure of nogalamycin and its analogues and their biological activity (*70*). Nogalamycin is known to be a topoisomerase I poison (*202*). Its analogue, menogaril, which lacks the nogalose sugar, is a topoisomerase II poison. Both drugs exhibit antitumour activity. However, they have been shown to have different cellular activities and target different stages of the cell

cycle. It is known that nogalamycin can induce a 22° bend in DNA (*183*), and can trap topoisomerase I-cleavable complexes at specific sites on DNA (*202*). The stimulation of site-specific topoisomerase I-mediated DNA cleavage has been linked to the groove binding properties of anthracyclines and *bis*-benzimidazoles. It has been shown that these compounds preferentially bind and stabilise bent DNA domains suggesting a possible correlation between drug-induced bending and topoisomerase I poisoning (*172*). The replacement of the DNA binding sequence for these drugs, with a sequence that introduces a bend in the helix, was shown to result in the stimulation of topoisomerase I-mediated DNA cleavage under appropriate conditions (*248*). This led to the hypothesis that nogalamycin-induced topoisomerase I-mediated DNA cleavage is a direct result of DNA bending.

Other anthracyclines lack the bulky substituent on ring D that poses a mechanical problem for nogalamycin. This molecule, which is essentially the shape of a dumbbell, with a minimum bell-width of 5.5 Å, can intercalate between base pairs that are only 3.4 Å apart in *B*-DNA. Two theories that explain this phenomenon have been suggested: The first (*42*) postulates that transient melting of the DNA helix is required in order for the drug to insert between the bases. The helix then reforms around the drug. The second binding hypothesis (*243*) involves elongation of the DNA along its helical axis, with associated unstacking and buckling of the base pairs, but without any disruption of base pair hydrogen bonds. This requires the bulky sugar substituents on nogalamycin to flip from axial to equatorial positions to facilitate intercalation. Major conformational changes for both the drug and the DNA are required in the latter theory.

DNase I footprinting studies (*65*) have shown that the unusual structure of nogalamycin endows it with the ability to discriminate between different nucleotide sequences. The selectivity is determined primarily by the local dynamic properties of the helix. Nogalamycin has been shown to intercalate

only in the pyrimidine-purine steps, TpG (and symmetry related CpA) and CpG (CpG), and binds very poorly to purine-pyrimidine steps. ApT (ApT) and TpA (TpA) pairs are very stable and difficult to unstack, and are poor binding sites for the drug.

For nogalamycin, there is no steric preference for sugar conformation in either groove (*42*). However, in all reported structures of nogalamycin-DNA complexes, including duplexes and hairpin loops (Section **3.7.1**), the aminoglucose resides in the major groove and the nogalaose lies in the minor groove. A combination of nonbonded forces, including VDW interactions, and solvent mediated H-bonds has been shown to stabilise the drug-DNA complex (*212*).

2.1.3 Clinical Uses of Anthracyclines

Doxorubicin is one of the most successful and widely-used anti-tumour drugs, and it is used to treat the acute leukaemias, lymphomas, and a variety of solid tumours, such as adenocarcinoma and malignant mesothelioma (breast and lung cancers). It is given by fast running infusion, commonly at 21-day intervals. Pharmacokinetic studies, determined in patients with various types of tumors undergoing either single or multi-agent therapy have shown that doxorubicin follows a multiphasic disposition after intravenous injection. The initial distributive half-life of approximately 5 minutes suggests rapid tissue uptake of doxorubicin, while its slow elimination from tissues is reflected by a terminal half-life of 20 to 48 hours. Steady-state distribution volumes exceed 20 to 30 L/kg and are indicative of extensive drug uptake into tissues. Plasma clearance is in the range of 8 to 20 mL/min/kg and is predominately by metabolism and biliary excretion. Approximately 40% of the dose appears in the bile in 5 days, while only 5 to 12% of the drug and its metabolites appear in the urine during the same time period. Binding of doxorubicin and its major metabolite, doxorubicinol (Fig.2.3(c)) to plasma proteins is about 74 to 76% and is independent of plasma concentration of doxorubicin up to 2 mM.

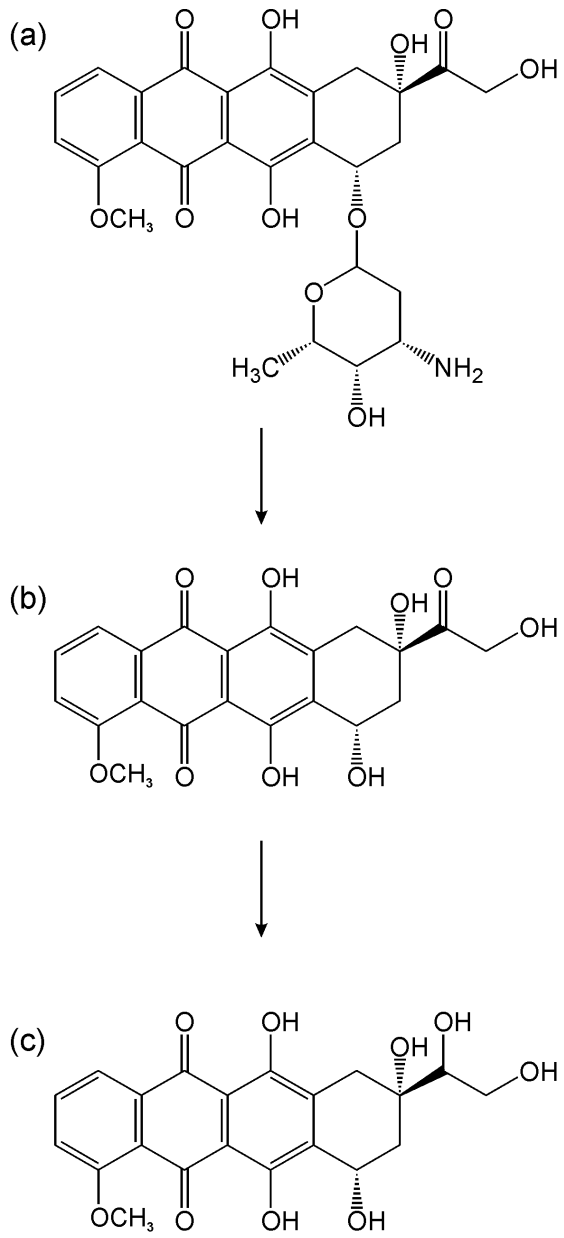


Fig.2.3: Pathway of doxycycline metabolism in human cardiac cytosol: Doxycycline (a) is converted to doxorubicinol (c) via the hydroxyaglycone (b).

In human cardiac cytosol, enzymatic reduction at the 7-position and cleavage of the daunosamine sugar yields aglycones (Fig.2.3(b);(c)) (*134*). High concentrations of cytosolic proteins cause doxorubicin hydrolysis as the first step and carbonyl reduction of the hydroxyaglycone as the second step. The formation of these metabolites is accompanied by free radical formation, the local production of which may additionally contribute to the cardiotoxic activity of doxorubicin.

Cardiotoxicity has been a major factor in limiting use of the anthracyclines, doxorubicin and daunorubicin. Toxicity is essentially of two kinds: acute, supraventricular tachycardia (*197*), which is a relatively uncommon complication, and a delayed, dose-related, cardiomyopathy (*53*), resulting in congestive heart failure. Higher cumulative doses are associated with cardiomyopathy. Special attention must be given to the cardiotoxicity induced by doxorubicin. Irreversible myocardial toxicity, manifested in its most severe form by life-threatening congestive heart failure, may occur either during therapy or months to years after termination of therapy. The probability of developing impaired myocardial function, based on a combined index of signs, symptoms and decline in left ventricular ejection fraction (LVEF) is estimated to be 1 to 2% at a total cumulative dose of 300 mg/m² of doxorubicin, 3 to 5% at a dose of 400 mg/m², 5 to 8% at a dose of 450 mg/m² and 6 to 20% at a dose of 500 mg/m² given in a schedule of a bolus injection once every 3 weeks. It is this specific cardiotoxicity that renders nogalamycin therapeutically useless. Because of the risk of symptomatic and potentially fatal heart failure, it is usual to limit total cumulative doses of doxorubicin to 450 mg/m² body-surface area.

Cardiomyopathy and/or congestive heart failure may be encountered several months or years after discontinuation of doxorubicin therapy (*80*). The risk of congestive heart failure and other acute manifestations of doxorubicin cardiotoxicity in children may be as much or lower than in adults. Children appear to be at particular risk for developing delayed cardiac toxicity in that doxorubicin-induced cardiomyopathy impairs myocardial growth as children mature, subsequently leading to possible development of congestive heart failure during early adulthood. Patients with pre-existing cardiac disease, the elderly, and those who have received myocardial irradiation should be treated cautiously. Cardiac monitoring, for example by sequential radionuclide ejection fraction measurement, may assist in safely limiting total dosage.

Treatment of doxorubicin-induced congestive heart failure includes the use of digitalis, diuretics, after load reducers such as angiotensin I converting enzyme (ACE) inhibitors, low salt diet, and bed rest.

Adriamycin®, (Doxorubicin Hydrochloride for Injection, BP) is available as a sterile red-orange lyophilised powder for intravenous use only, in 10, 20 and 50 mg single dose vials and a 150 mg multidose vial from Pfizer Ltd., (formerly Pharmacia Ltd.).

Caelyx®, the liposomal form of doxorubacin, is licensed by the MCA (*141*) for the treatment of Kaposi's sarcoma (KS), a cancer that develops in connective tissues such as cartilage, bone, fat, muscle, blood vessels or fibrous tissues, and also as a third-line treatment of metastatic ovarian cancer. For decades KS was considered a rare disease that mostly affected elderly men of Mediterranean or Jewish heritage, organ transplant patients, or young adult African men (*227*). In the last 20 years, however, the vast majority of KS cases have developed in association with human immunodeficiency virus (HIV) infection and the acquired immunodeficiency syndrome (AIDS), especially among homosexual men (*161*).The mechanism of action of Caelyx® is identical to that of doxorubicin, but the active drug is encapsulated in a lipid bilayer microsphere formulated with surface-bound methoxypolyethylene glycol (MPEG). The pegylated drug is protected from detection by the mononuclear phagocyte system (MPS) and, since it does not bind to plasma proteins and thus resides in the intravascular space, has a much smaller volume of distribution and much longer terminal half-life.

Pharmorubicin®, (Epirubicin Hydrochloride for Injection, BP), a sterile red-orange solution for intravenous use only, is available in 10 mg single dose vials from Pfizer Ltd., (formerly Pharmacia Ltd.). Epirubicin possesses essentially the same spectrum of activity as doxorubicin, as well as a quantitatively similar pattern of metabolism. This agent has lower cardiac

toxicity relative to doxorubicin (*197*). The cumulative maximum tolerable dose of epirubicin is 900 mg/m^2.

Daunorubicin is used in the treatment of acute lymphocytic and acute granulocytic leukemias. It is the drug of choice in the treatment of non-lymphocytic leukemia. Daunorubicin is administered I.V. with a biphasic elimination curve with a terminal half-life of 1.5-10 hours. The drug is primarily metabolised in the liver. Unlike doxorubicin, renal excretion of drug and its metabolites occurs within six hours and the urine may turn bright red in colour. The drug causes much the same type of cardiomyopathy as doxorubicin. The incidence has been reported to be 1.5% at a total dose of 600mg/m^2 and 12% at 1000 mg/m^2 (*197*).

Daunomycin is formulated as Cerubidin®, a sterile red lyophilised powder of Daunorubicin Hydrochloride BP for intravenous use only, in 20 mg single dose vials, by Beacon Pharmaceuticals Ltd. Daunorubicin is also available as a pegylated liposomal formulation for the treatment of KS (DaunoXome®, Gilead Sciences.).

2.1.4 Cooperativity in Drug-DNA Complexes

In terms of nucleic acids with multiple binding sites, cooperativity can be defined as a thermodynamic effect, which describes the effect that one bound ligand has on the binding affinity of a second ligand. Cooperativity arises when the changes in free energy of multiple species are mutually dependent. Cooperative interactions can be defined as either positive or negative, depending on whether the binding affinity of the second drug is enhanced or diminished by the presence of the first.

Negative cooperativity, or anticooperativity, is often related to steric effects, resulting from an overlap of the binding sites (*241;256*). Positive cooperativity is thought to be related to conformation changes in the DNA

(*92;230*). Binding of the first ligand may promote a structural change in the DNA, making the second site more favourable for the second ligand. This change may be permanent, as a result of drug binding, or transient, occurring only during part of the binding event.

Cooperative binding has been identified between drug molecules that bind to DNA with varying modes of interaction, including intercalation (*193*) and groove binding (*74;75*). The factors responsible for cooperativitity are presumably different, but ill-defined for each mode of binding. The intercalation event involves distortion of the DNA helix, during which the binding of the first molecule may cause a conformational change at the second site, increasing its affinity to bind a second ligand. Thus, the drug induced conformational change may be propagated from one site to the other.

2.2 Experimental Design

2.2.1 Project Overview

End effects encompass the range of unfavourable interactions, either with the surrounding solvent or within the crystal lattice, that occur when the bulky sugars that bookend the nogalamycin chromophore overhang the ends of the oligonucleotide in which it is intercalated. Here, we aim to assess the contribution of *end effects* on the orientation of nogalamycin bound to a terminally-located interacalation site by determining how far away the site must be from the terminus of an oligomer in order to negate their contribution. This will be assessed in terms of drug orientation within a symmetrical CpG intercalation site. To address the question of the orientational preference of nogalamycin, we attempt to assess whether nogalamycin, complexed in a 2:1 ratio with three oligonucleotides, each containing the core sequence $d(CGATCG)_2$, is oriented such that its sugars point towards the centre of each duplex or outwards towards its ends.

We are interested in investigating whether cooperativity exists in DNA sequences demonstrating the potential for communication between binding sites, but without the possibility of direct ligand-ligand interactions and, if so, in probing the cause of this phenomenon. We aim to distinguish whether nogalamycin binds to two equivalent intercalation sites, each separated from the other by an ApT spacer, in a cooperative or anticooperative manner, or whether each binding event is independent of the other. To achieve this, we have designed a simple qualitative model that can quickly assess whether there is communication between the sites or whether they are entirely independent of one another. If the sites are shown to interact, the complexity of the ^{1}H NMR spectra of the drug-DNA complex relative to the free DNA can be used to determine the nature of this interaction; either cooperative or anticooperative. It is important to state at this point that this is not a thermodynamic study and, as such, does not attempt to quantify the degree of cooperativity. Rather, we attempt here to qualitatively assess the relationship, *vis-à-vis* cooperativity, between the binding sites.

2.2.2 Sequence Design

Smith *et. al.* (*212*) have previously published the crystal structure of the 2:1 complex of nogalamycin with $d(TGATCA)_2$, in which two nogalamycin binding steps are separated by an AT spacer. In order to facilitate comparison between x-ray and NMR-refined structures (Section **5.4.2**), this sequence was chosen as a starting point for this study. To assess how the orientation of the drug is affected by end effects, the sequence $d(CGATCG)_2$, in which the TpG (CpA) intercalation steps are replaced with analogous CpG (CpG) steps, was also synthesised. Following the study of these two complexes, the sequences $d(TCGATCGA)_2$ and $d(CTCGATCGAG)_2$, in which each CpG site is flanked on its terminal face by one and two base pairs respectively, were synthesised to see what effect, if any, internalising the binding site would have on drug orientation. To act as a control, and to aid in the assigning the NOESY spectra of these sequences, the sequences $d(CTGATCAG)_2$ and $d(CCTGATCAGG)_2$

were also synthesised. For the TpG-containing oligomers, CG base pairs were used to extend the sequence lengths, as these offered the greatest stability. In the case of the CpG-containing sequences, one TA and one CG pair were used at each end. Although this would be potentially less stable than using two CG pairs as before, it should improve the chemical shift dispersion on key areas of the NMR spectra.

2.3 Materials & Methods

2.3.1 DNA Synthesis & Purification

All DNA sequences used were synthesized using an Applied Biosystems® 394™ phosphoramidite solid phase oligonucleotide synthesiser, by Mr John Keyte, School of Biomedical Sciences, University of Nottingham. In each case, the sample volume was approximately 1.5 ml of 10 μM ssDNA, equivalent to roughly 20 mg of duplex DNA in total. In preparation for HPLC separation, the DNA was centrifuged for 10 minutes at 13000 rpm. The supernatant was further cleaned by syringe filtration through a 450 nm aperture width filter, to remove all insoluble impurities from the sample.

The DNA was purified *trityl on* by reverse phase HPLC on a HyperSil® 10μm ODS, 10x250 mm packed C18 semi-preparative column, using an acetonitrile gradient (0-70 % over a 30 minute timescale) and a TEAA buffer (pH 7.0). Approximately 80 % of the total purified DNA was retrieved, by separation of six to eight 250 μl aliquots. All the recovered fractions were then combined and any remaining acetonitrile removed by rotary evaporation. The aqueous vestiges were freeze dried and reconstituted with ~10 ml distilled water, before being heated with an equal volume of 50 % aqueous acetic acid for 30 minutes, to remove the DMT protecting group. The acetic acid was sequentially extracted into three 200 ml portions of diethyl ether. The aqueous fraction was dialysed, first against distilled water (2 L), then against a similar volume of NaH_2PO_4 (100 mM), NaCl (1 M), to introduce sodium as the counter-ion. After further dialysis with H_2O, the sample was lyophilised and

redissolved in 600µl of H_2O. To this was added NaCl (1 M, 70 µl) and NaD_2PO_4 (100 mM, 70 µl), to aid duplex formation, 10µl of the metal chelator EDTA (6mM), NaN_3 (6 mM, 10µl), to protect against microbial contamination, and the internal standard TSP (1.5 µM, 5 µl). After adjusting to pD 7.4 with NaOD (100 µM) and DCl (100 µM), the sample was again lyophilised and, finally, made up to 700 µl with D_2O.

2.3.2 Nogalamycin Preparation

The nogalamycin was generously provided by Pfizer Inc. (formerly Pharmacia & Upjohn). 2.5 mg of solid nogalamycin was added to 300 µl of buffered (pD 7.2) D_2O solution (0.1 M NaCl, 10 mM NaD_2PO_4, 0.1 mM EDTA). The mixture was shaken is a cold room (4°C) for 24 hours. The pH was adjusted to 7.5 using 0.1 M solutions of DCl and NaOD. The sample was subsequently centrifuged to remove any undissolved solids.

2.3.3 NMR Analysis

NMR experiments were performed on either a Bruker Avance™400 or Bruker DRX™500 spectrometer. All one-dimensional NMR experiments were recorded over 32768 data points in the f2 dimension, with spectral width of 16.022 ppm and a relaxation delay time of 1.5 s. Phase-sensitive Double Quantum Filtered COrrelated SpectroscopY (DQF-COSY) (*171;177*), TOtal Correlated SpectroscopY (TOCSY) (*13*), and NOE SpectroscopY (NOESY) (*12*) experiments were performed, collecting either 1024 or 2048 points in f2, and between 400 and 512 points in f1. NOESY data at mixing times ranging from 75 ms to 300 ms were collected with a spectral width of 16.022 ppm in each dimension, with a delay time of 1.5 s. Quadrature detection in f1 was achieved using states time-proportional phase incrementation (states TPPI). All TOCSY experiments employed a spin-locking field of 5 kHz. All data were processed on an SGI® Indy R5000™ or O2™ workstation, using Bruker XWIN-NMR™ processing software. Two-dimensional data were zero-filled to 2048 x 1024 points prior to Fourier transformation, optimised with a shifted

sine squared function in both dimensions, and treated with automatic baseline correction.

2.3.4 Distance Restraints

The NOESY experiment identifies protons in close spatial proximity. For an isolated spin pair, the intensity of an observed NOE is inversely proportional to the sixth power of the distance between the protons causing it (*10*). In macromolecular systems, the intensity of the NOE can be altered by cross-relaxation to another nucleus during the mixing period (τ) of the experiment. This phenomenon is known as spin-diffusion. The NOE build-up curves for DNA systems such as these have been shown to have a strong linear characteristic for mixing times of less than 150 ms (*249*). Since the mixing time is under experimental control, the distances can be linearly extrapolated to zero mixing time by collecting three low mixing time (<120ms) NOESY spectra, thus alleviating the effects of spin diffusion.

In DNA, some fixed distances defined by the covalent structure of rigid areas in the macromolecule that correspond to well-resolved NOESY crosspeaks can be used to normailise other peak intensities using **Equation 2.1**. To increase the accuracy of this procedure, different internal distances (D_{ref}) are used to reference NOE intensities (I_{ij}) corresponding to proton interactions in different environments. NOEs involving sugar protons are referenced with deoxyribose H2'-H2" distances (I_{ref}), which are known to be separated by 1.85 Å. Similarly, base proton separations are referenced relative to the cytosine H5-H6 distance (2.45 Å), while the thymine CH_3-H6 reference distance (3.00 Å) is applied to NOEs involving methyl protons.

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

Equation 2.1: Determination of the distance (D_{ij}) between two protons i and j, with an NOE crosspeak intensity (I_{ij}), calibrated for a known interproton separation (D_{ref}) with a measured intensity (I_{ref}).

By combining this distance information with the fully assigned NOESY spectrum, a set of distance restraints for the complex can be generated. These restraints can subsequently be incorporated into the AMBER force field (Section **3.3.4**), restricting the number of conformations available to the structure.

Assignment of all NOESY spectra was carried out using ANSIG (*124*), a program for the assignment of protein and nucleic acid ^{1}H 2D NMR spectra using interactive graphics. The spectra were integrated using the Bruker XWIN-NMR™ software suite. The frequency-domain spectrum matrices, together with the sequence of the DNA complexes, were used as input data for ANSIG. Typically, low-resolution (1024 x 1024) contour maps of the NOESY (τ = 300 ms), TOCSY and DQF-COSY spectra of the sample were read into the program. The centre of each discrete crosspeak on the NOESY contour plot was then marked with a cross, which was labelled with one proton assignment for each of the two frequency domains, F_1 and F_2, using the procedure described in Section **2.4.2**. The crosspeaks were then connected in F_1 and F_2 until a complete spectral assignment was constructed. The resulting ANSIG output files consisted of a list of frequency domain coordinates, corresponding to the centre of each crosspeak, together with an assignment in F_1 and F_2 for each.

As for ANSIG, frequency-domain spectrum matrices were used as the input data for XWIN-NMR™. The integration regions around each crosspeak on the NOESY spectra were defined by enclosing them in boxes. Once all peaks were enclosed, the integrals were computed. For the crosspeak integration to work properly within XWIN-NMR™, it was vital that when two or more crosspeaks overlapped (Fig.2.4), they should be separated, as areas of overlap would cause discrete peak volumes to be added to the total volume multiple times (Fig.2.4(a)), and so affect the accuracy of the integration. In order to integrate overlapping peaks, a portion of each peak on the face opposite the

area of overlap was boxed (Fig.2.4(b)). Once the volume within this box had been integrated, it was multiplied by an appropriate factor to create an isolated peak with intensity approximately equal to that of the original peak, had it not been overlapped (Fig.2.4(c)). Obviously, this procedure gives rise to less accurate intensities than those obtained by direct measurement, and so the error bounds on the resulting distance restraints were increased accordingly.

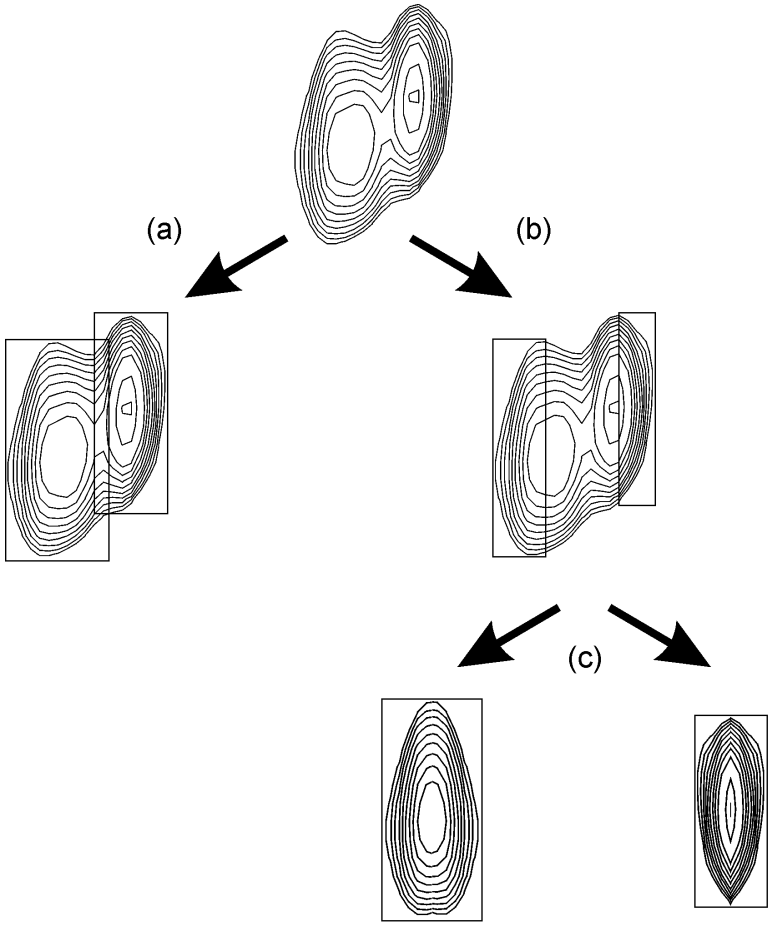


Fig.2.4: Procedure for resolving overlapped peaks in XWIN-NMR™.

The integration procedure was carried out for three NOESY spectra with mixing times of 120, 100 and 75 ms respectively. Since 2D NOESY contour maps possess 2-fold reflectional symmetry with a mirror line along their diagonals, the crosspeaks above, and below, the diagonal were treated as separate spectra, which can be mutually crosschecked. Both above- and below-diagonal integrals were then used to generate inter-proton distances at

each mixing time using **Equation 2.1**. There were not always three points on the NOE build up curve from which to extrapolate the zero-mixing time distance. Often, some weaker NOEs did not appear on the lowest mixing time spectra, as they were given insufficient time to develop. In such cases, there were only two points describing the build-up curve, and so its linearity could not be verified. Though there is no reason to suspect that these curves did not have a significant linear component, this did warrant increasing the error bounds on the resulting distance restraints. The XWIN-NMR™ output files comprised a list of two sets of coordinates, which correspond to the top-left and bottom-right corners of each integration region, together with the distance calculated from the integral within that region.

The output files from ANSIG and XWIN-NMR™ were subsequently combined to give rise to six files, corresponding to three low mixing time NOESY spectra for above, and three corresponding spectra for below, the diagonal. These files contained a list of labelled proton-proton distances. The protocol for combining the interproton distances, from XWIN-NMR™, with their assignments, from ANSIG, assumes that there will be only one assignment within each integration region, and that there is no overlap between adjacent regions. Multiple hits were therefore discarded, emphasising the importance of clean spectra with clearly defined crosspeaks when generating distance restraints for rMD simulation.

The labelled interproton distances from each of the 120, 100 and 75 ms spectra were extrapolated to zero mixing time. The resulting distances from the above- and below-diagonal spectra were parsed against each other. In instances where there was good correlation between two symmetry-translated crosspeaks, their associated distances were averaged. If a crosspeak was absent in either spectrum, only one value was used. Using this simple parsing method, erroneous and inaccurate peaks could be easily detected and removed.

For all of the rMD studies described in this thesis, the inherent error associated with distance calculations from the NOE data led to a 10% upper bound error being applied to all distances involving the non-exchangeable protons. All distances involving the exchangeable protons were given a 25% upper bound error due to the degree of exchange with the solvent. Distances obtained from indirectly integrated crosspeaks, and those extrapolated from incomplete build-up curves were also subject to a 20% error bound. Distance restraints were compared visually with the distances of a structure derived from unrestrained molecular dynamics simulation, by using MolMol (*123*) software for molecular graphics and display.

2.3.5 Drug Titration

Previous NMR studies of nogalamycin-DNA complexes (*181;256*) have been concerned primarily with the structural characterisation by 2D methods. In these studies, an excess (typically 3 equivalents) of solid nogalamycin was added to an aqueous solution of the duplex. The resulting solution was then lyophilised and redissolved before adjusting the pH to 7.0 to precipitate the nogalamycin, which was subsequently removed by centrifugation. In this study, the nogalamycin was titrated into the DNA solution in aliquots, and a 1D ^{1}H NMR spectrum of the forming complex was collected after each addition.

2.3.6 Starting Structures

A Cartesian coordinate model for the right handed Arnott B-form DNA structures, d(TGATCA)$_2$ and d(CCTGATCAGG)$_2$, were generated using the NUCGEN module of AMBER 6.0 (*29*). Using the LEaP module of the AMBER suite, a gap of approximately 10 Å was created at each intercalation step in order to accommodate the drug. A model of the drug molecule, which had been previously parameterised, was manually docked into each widened step. The orientation of the drug in each binding site was based on NOE data from the 2D spectra of the complexes (Section **2.4.3**), and by reference to

previously published structures (*210;212*). Sodium counter-ions were used to negate the negative charge on the DNA backbones, and the models were solvated to within a distance of 5 Å using boxes of 216 TIP3P H_2O molecules (*105*). The period box size for nogalamycin-d(TGATCA)$_2$ was approximately 37x40x43 Å, and 37x40x55 Å for nogalamycin-d(CCTGATCAGG)$_2$. In each case, the entire system was energy minimised to optimise bond lengths.

2.3.7 Structure Calculations

Minimisation was followed by ten rounds of unrestrained molecular dynamics simulation, using the protocol described in Section **4.3.5**.

2.3.8 Restrained Molecular Dynamics

Following minimisation and unrestrained MD simulations, NMR restraints were introduced to the equilibrated structure and the system allowed to undergo restrained MD to satisfy the NMR restraints and search adequate conformational space under the influence of water molecules and cations. A total of 248 NOE-derived distance restraints (124 per strand), each subject to a 10% inherent error, were applied to nogalamcyin-d(TGATCA)$_2$, while 338 restraints, 200 of which were subject to a 10% upper bound error and 138 with a 20% associated error bound, were applied to nogalamcyin-d(CCTGATCAGG)$_2$. Each system was subjected to rMD with all restraints active. The former system underwent 2 ns of rMD and the latter was subjected to 1 ns simulation. 162 DNA-DNA distance restraints, 50 drug-drug and 36 drug-DNA restraints were applied to the hexamer complex. The decamer was subject to 240 DNA-DNA distance restraints, 146 of which were subject to a 10% upper bound error. Of the 52 drug-drug and 46 drug-DNA restraints applied to the longer model, 10% upper error bounds were associated with 24 and 30 respectively. Autocorrelation plots, like that shown in Fig.2.5, showed that, in each case, independence from the starting structure was achieved after 350 ps. rMD was followed by energy minimisation. Snapshots were extracted throughout the rMD runs at picosecond intervals.

For the final averaged energy minimised structure of nogalamcyin-d(TGATCA)$_2$, none of the applied distance restraints were violated by more than 0.5 Å, while only three deviated by more than 0.3 Å. The mean pairwise RMSD, calculated for all heavy atoms, over the final 100 acquired snapshots was 0.60 (±0.1) Å. The NMR restraints for nogalamcyin-d(TGATCA)$_2$ were satisfied well within the two nanoseconds of simulation. Similarly, in the final structure of nogalamcyin-d(CCTGATCAGG)$_2$, only five of the applied high-quality (10% error) distance restraints were violated by more than 0.5 Å, while seven deviated by more than 0.3 Å. All but one of the deviated restraints were associated with the fraying end residues of the complex. The mean pairwise RMSD in this case, calculated over the final 100 acquired snapshots was 0.87 (±0.1) Å. This figure is reduced to 0.73 (±0.1) Å if only the core d(TGATCA) DNA atoms are considered.

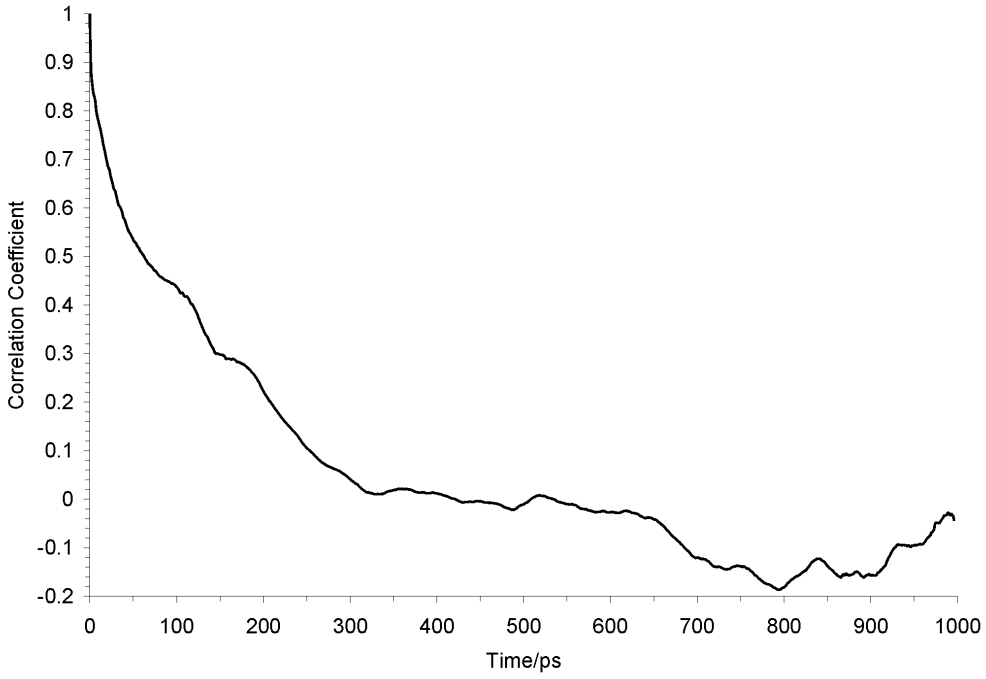


Fig.2.5: Autocorrelation plot for the first 1 ps of the 2 ns rMD simulation of nogalamycin-d(TGATCA)$_2$.

		nogalamycin-d(TGATCA)	nogalamycin-d(CCTGATCAGG)
NOE-derived restraints			
	drug-DNA	36	46
	DNA-DNA	162	240
	drug-drug	50	52
Restraint Violations			
	>0.3Å	3	7
	>0.5Å	0	5
RMSD			
	heavy atoms	0.60 (±0.1) Å	0.87 (±0.1) Å

Table 2.1: Summary of modelling statistics for the final 100 ps of rMD trajectories of nogalamycin-d(TGATCA) (2 ps) and nogalamycin-d(CCTGATCAGG) (1 ps).

2.4 Results & Discussion

2.4.1 Titrations

The 1D ^{1}H NMR spectra for the free uncomplexed DNA duplexes $d(TGATCA)_2$, $d(CTGATCAG)_2$, $d(CCTGATCAGG)_2$, $d(CGATCG)_2$, $d(TCGATCGA)_2$ and $d(CTCGATCGAG)_2$ are simplified, due to the self complementary nature of these sequences. In each case, upon titration with nogalamycin, a 1:1 complex began to form, giving rise to a more complex spectrum, since there were now many more species of protons in magnetically different environments (Fig.2.6;2.7). As the titration continued, the 2:1 species was formed, and each spectrum simplified as symmetry was returned. Thus the complexity of the NMR spectrum alone was used as an indicator of the relative amounts of each complex at a given step in the titration. The formation of a clean spectrum, with only one peak corresponding to each drug proton, indicates that a 2:1 complex, with dyad symmetry, was the final product. In all previous structural studies of nogalamycin complexed in a 2:1 ratio with symmetrical DNA oligomers (*133;181;183;210;212;256*), the intercalated nogalamycins have oriented themselves in opposite directions, so that the dyad symmetry of the free DNA is maintained in the complex. The preservation of this 2-fold symmetry shows that both drugs intercalate in a symmetrical manner, either both oriented towards the centre, or both towards the ends, of the duplex. In the case of the TpG-containing sequences, it can be

assumed that the drug is oriented towards the centre of the duplex. However, for CpG-containing sequences, no assumptions are made about the binding orientation, except that they are the same for both drug molecules.

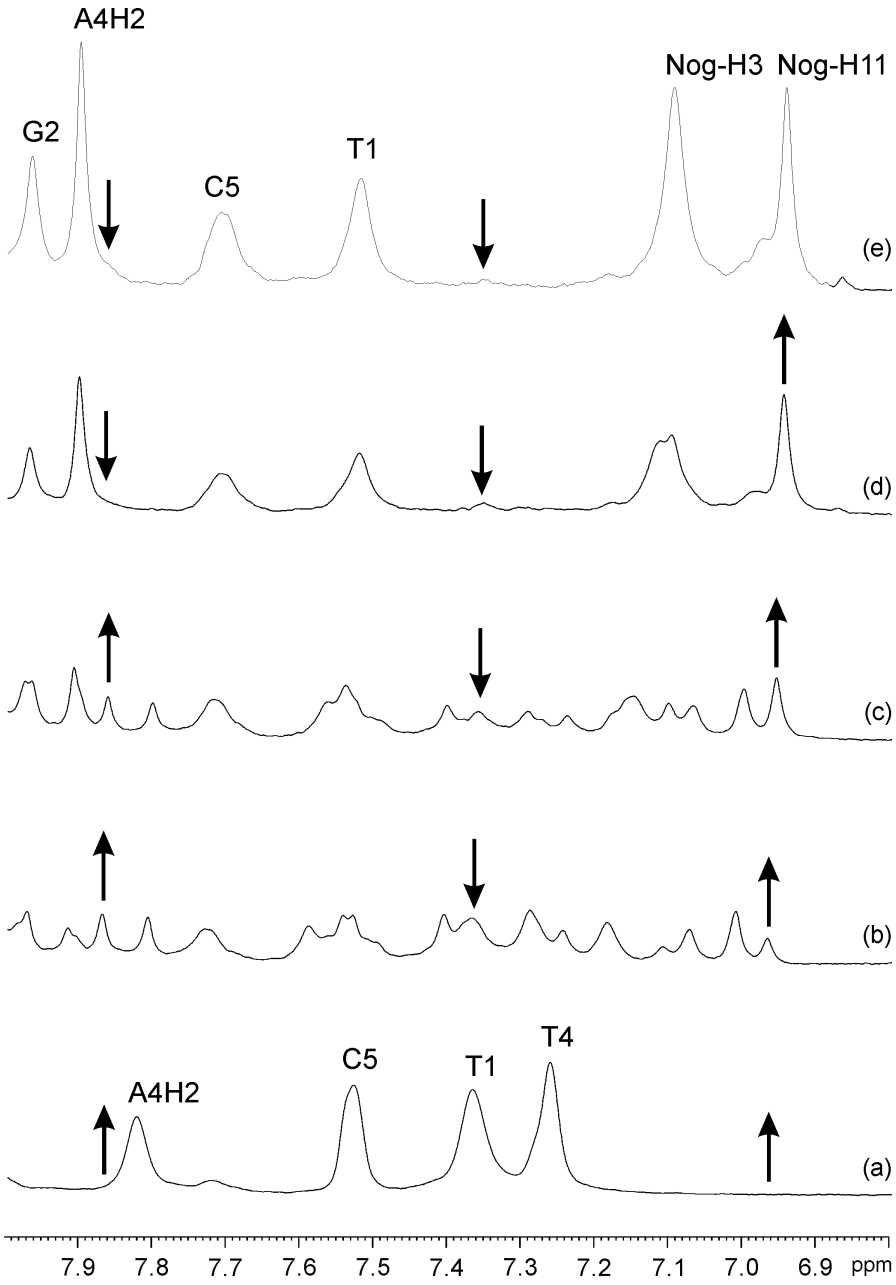


Fig.2.6: 1D ^{1}H NMR spectra of the titration of d(TGATCA)$_2$ with nogalamycin (298 K, D_2O), highlighting the aromatic region (6.8-8.0 ppm), at drug:DNA concentrations of (a) 0:1, (b) 1:2, (c) 1:1, (d) 4:3 and (e) 2:1. The downward facing arrows highlight clearly defined signals that decrease upon addition of the nogalamycin. Upward facing arrows indicate new signals that increase owing to complex formation.

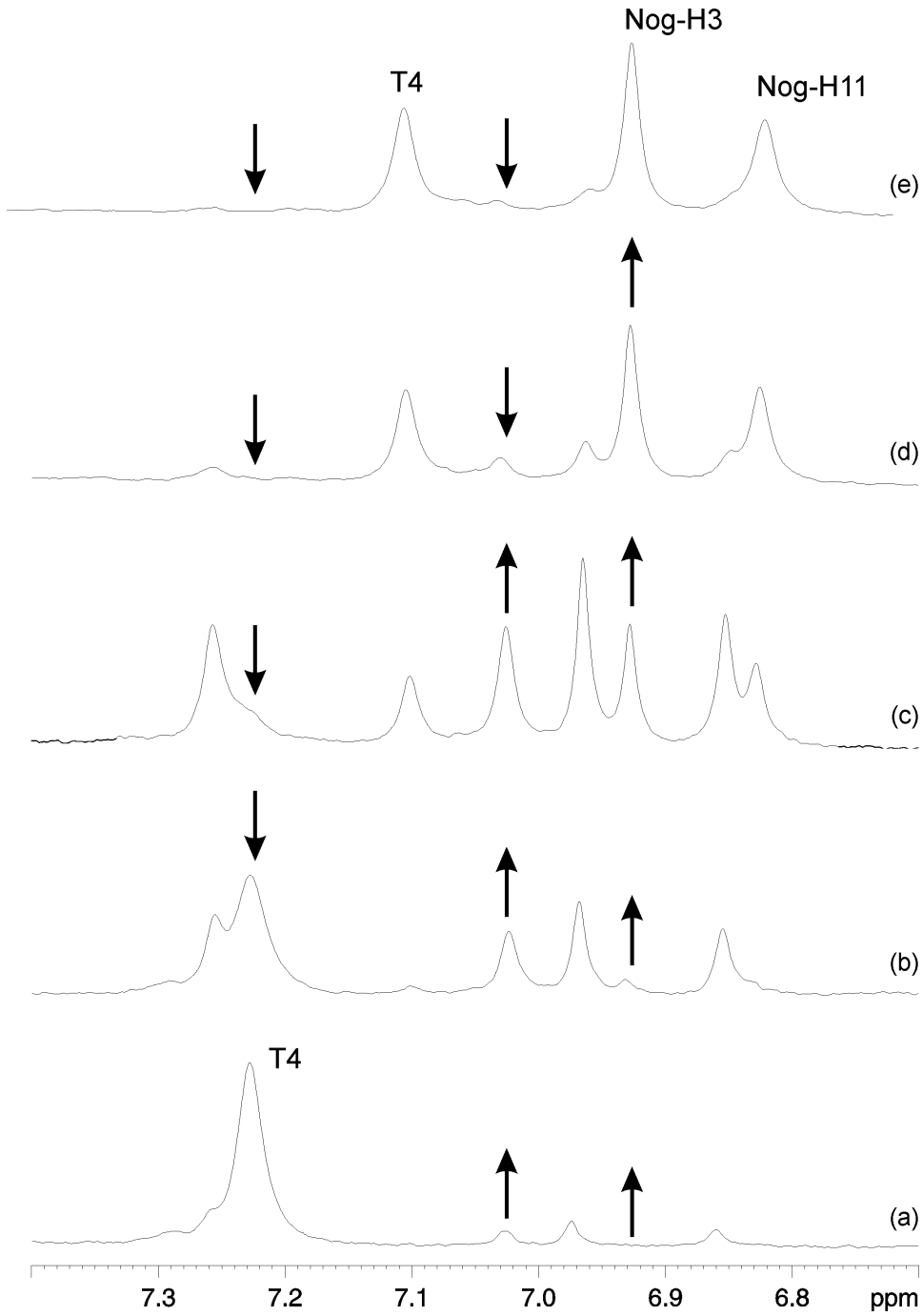


Fig.2.7: 1D ^{1}H NMR spectra of the titration of d(CGATCG)$_2$ with nogalamycin (298 K, D_2O), highlighting the aromatic region (6.7-7.9 ppm), at drug:DNA concentrations of (a) 0:1, (b) 1:2, (c) 1:1, (d) 4:3 and (e) 2:1. Arrows are used as in Fig.2.6.

2.4.2 Proton Assignments

The NMR investigations of these drug-DNA complexes were undertaken by first characterising the free oligonucleotides. This was followed by examination of the complexes formed upon reaction with nogalamycin. The

duplexes and drug-DNA complexes were fully assigned using a combination of *through space* and *through bond* interactions from NOESY (Fig.2.8), TOCSY and DQF-COSY spectra, using a method loosely based on that described by Chazin & coworkers (*34*).

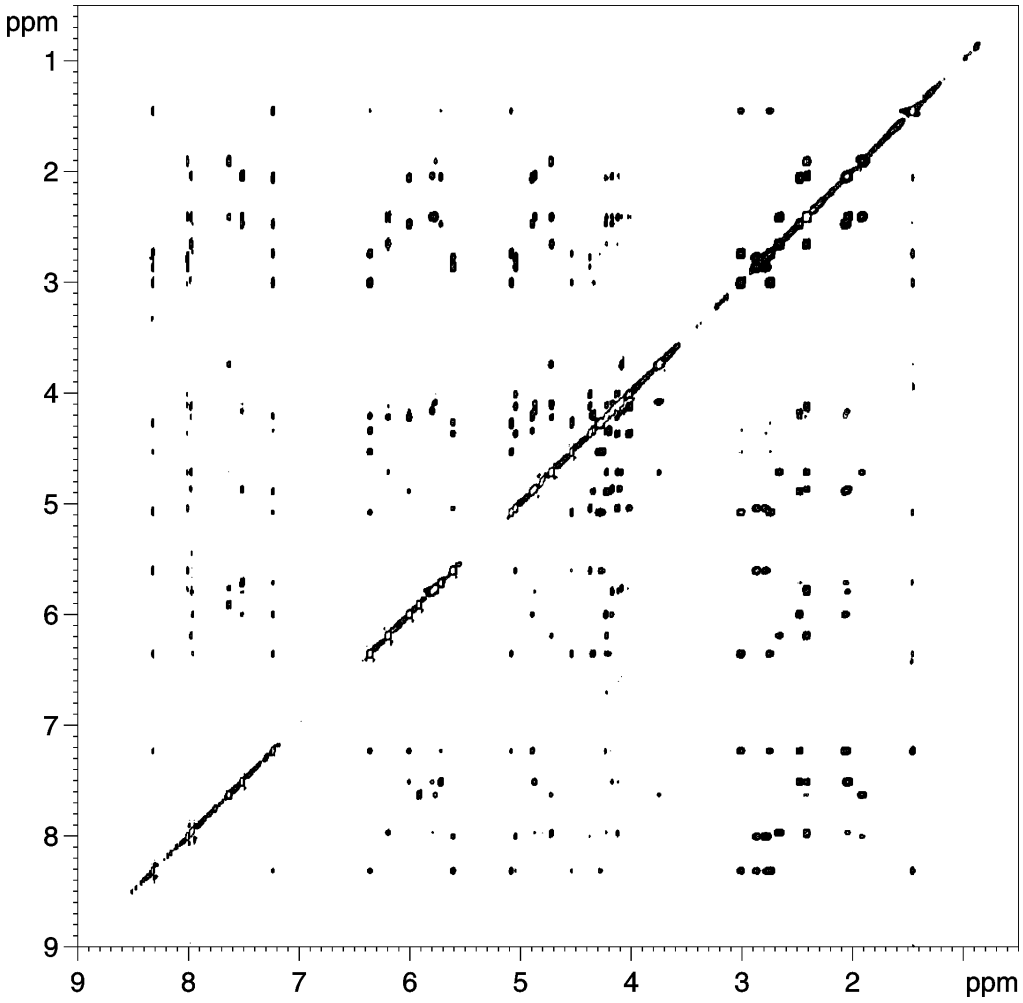


Fig.2.8: NOESY spectrum of d(CGATCG)$_2$ in D_2O solution, recorded at 298 K with a mixing time of 300 ms.

It has previously been shown that H1' protons of B-form DNA typically come into resonance between 6.5 and 5.0 ppm, while H6/H8 protons resonate in the range 8.5-7.0 ppm (*63*). Consequently, the starting point for the assignment of each of the uncomplexed DNA oligomers was the NOE interactions in this region of the NOESY spectrum i.e. the H6/H8-H1' assignment pathway (Fig.2.9). The H6/H8 of the 5'-terminal residue will interact through space with the H1' of its own deoxyribose only, whereas subsequent H6/H8 protons will interact with both their own H1' protons and with the sugar protons of the preceding residue.

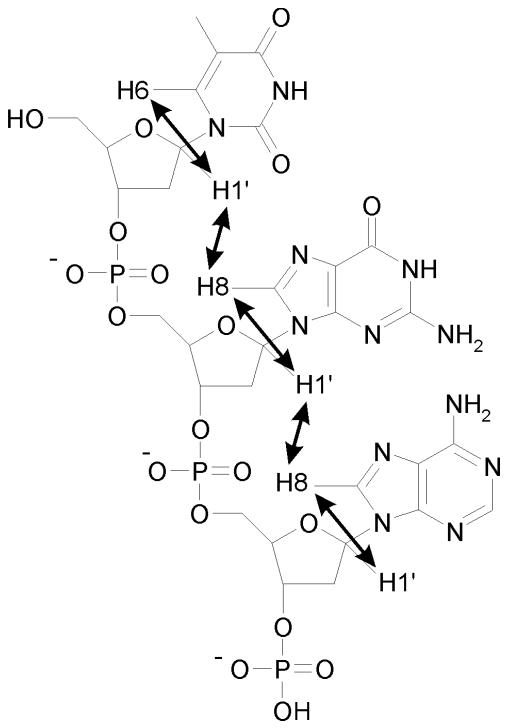


Fig.2.9: Schematic representation of the H6/H8-H1' assignment pathway starting at the 5'-terminal end of a DNA oligomer.

Assuming that all NOEs in the H6/H8-H1' region have been detected, a crosspeak without a partner in the F_1 dimension is likely to be the terminal intra-residue NOE. The H6/H8-H1' assignment pathway can then be traced out to its terminus at the 3'-terminal H6/H8-H1' crosspeak (Fig.2.10). Since their H5 protons resonate in the same frequency range as H1' protons, cytosine

residues will also show J-coupled peaks in this region of the DQF-COSY and TOCSY spectra. These are useful indicators that the initial assignment is correct, as are thymine Me-H1' dipolar connectivities on the NOESY spectra.

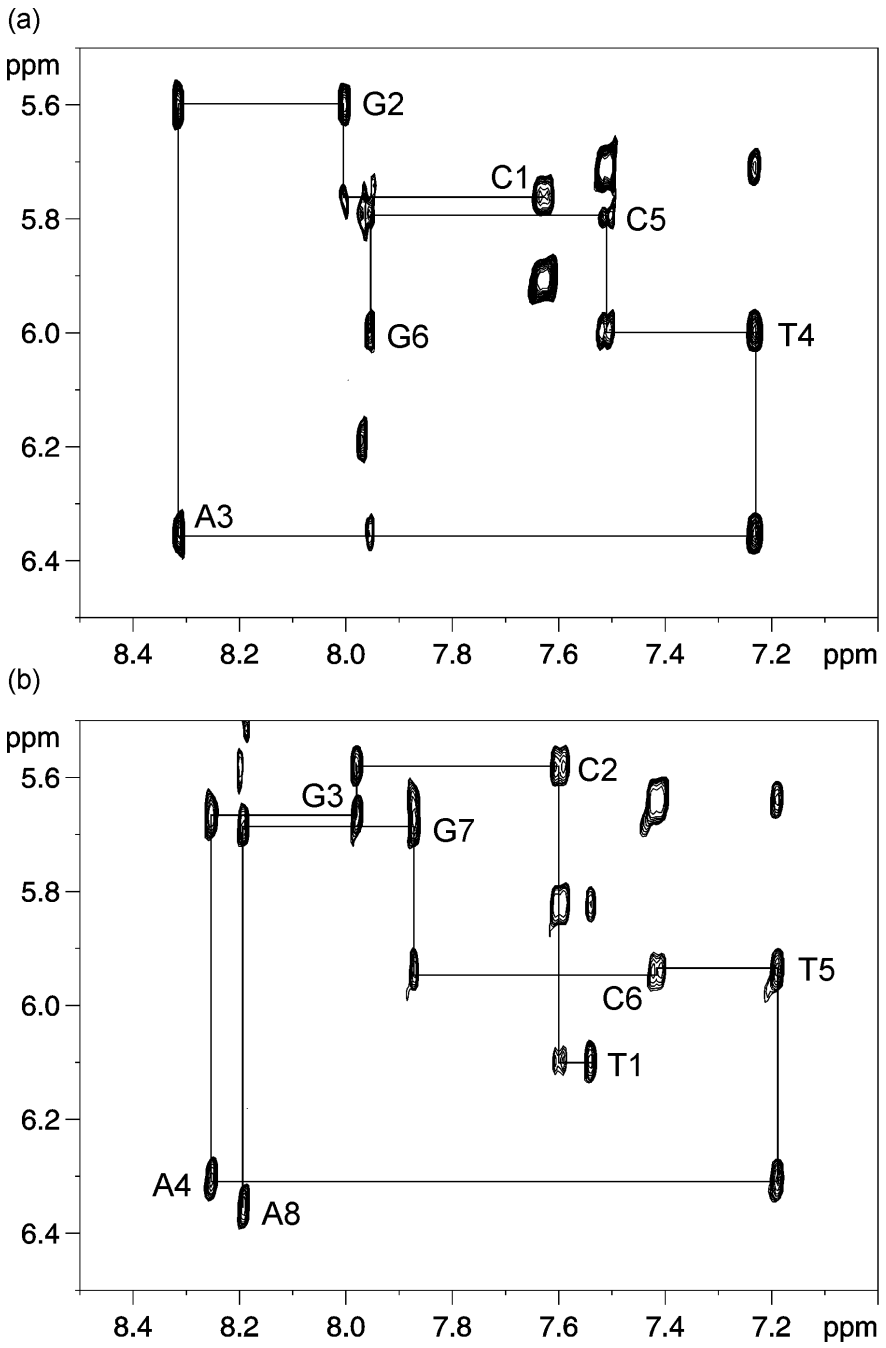


Fig.2.10: Portion of the NOESY spectra of (a) d(CGATCG)$_2$ and (b) d(TCGATCGA)$_2$, showing the H6/H8-H1' assignment pathway.

Once the initial connectivites have been verified, the ^{1}H spin systems of the individual nucleotides can be traced out (Fig.2.11). Often, all sugar protons can be identified on the DQF-COSY spectrum, but occasionally some H2" and H5" crosspeaks may be absent, dependant largely on the degree of sugar pucker (*34*). However, if one of the H2'/2" or H5'/5" protons can be identified, the other can be easily assigned, as both protons will appear as pairs in the F_2 dimension, separated by 0.0-0.9 ppm on the NOESY spectrum.

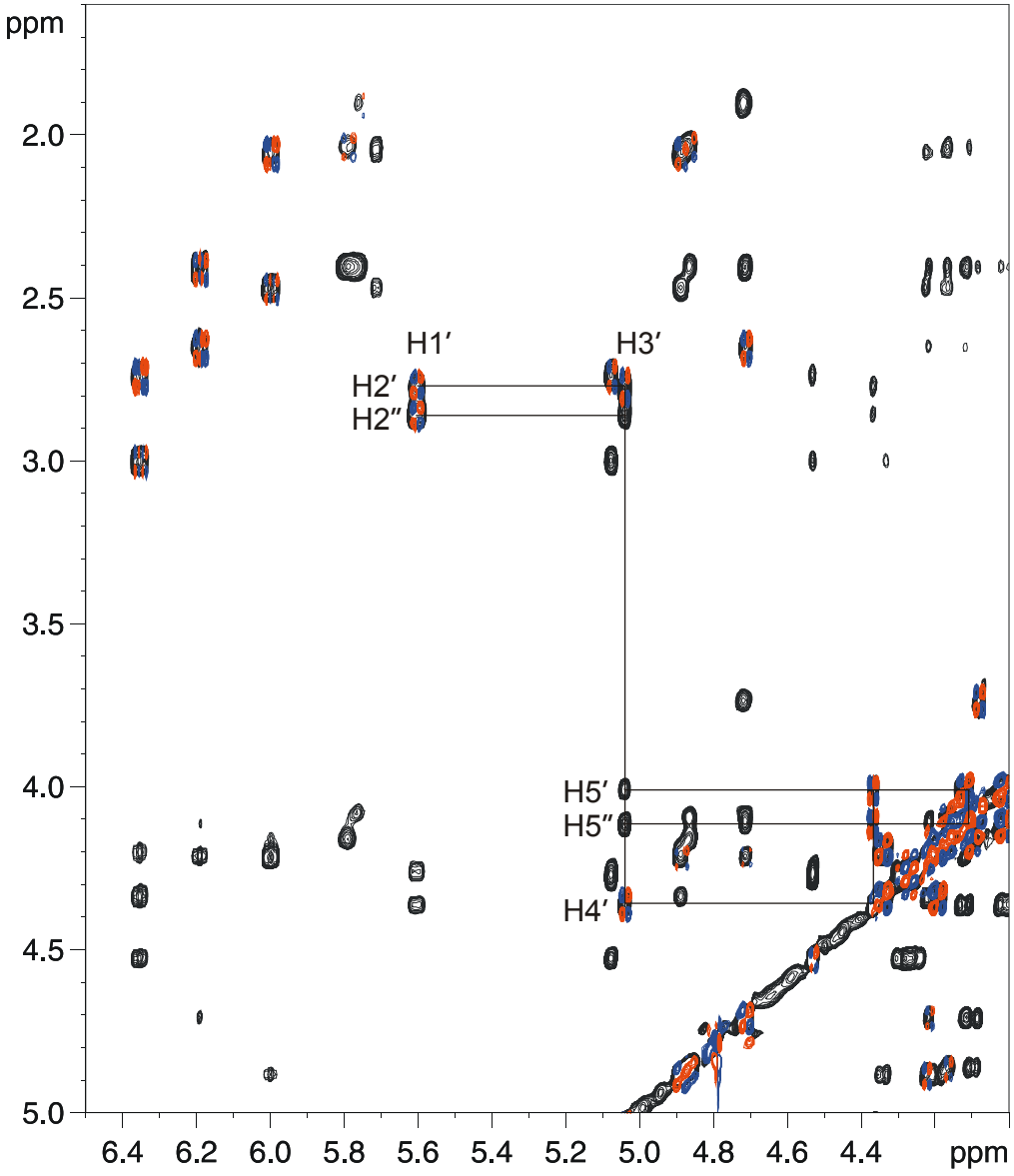


Fig.2.11: Superimposed NOESY (black) and DQF-COSY (red/blue) spectra of ^{1}H spin system of the G2 (guanosine) nucleotide of d(CGATCG)$_2$.

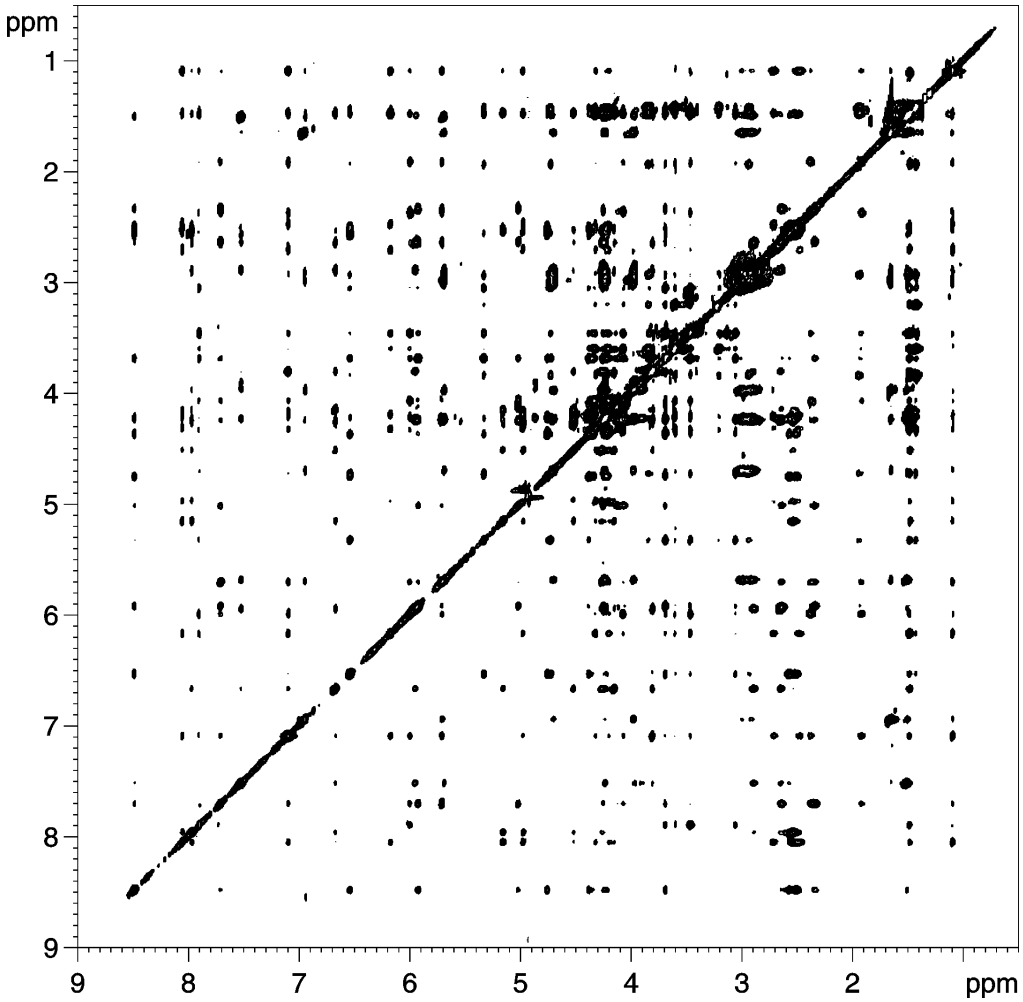


Fig.2.12: NOESY spectrum of nogalamycin-d(TGATCA)$_2$ in D_2O solution, recorded at 298 K with a mixing time of 300 ms.

In the drug-DNA complex, H6/H8-H1' assignment pathway will be broken in two places, as a result of the increased separation between the nucleotides at the binding site required to facilitate nogalamycin intercalation. This is caused by increased separation between the bases to facilitate the insertion of the drug. The intensity of NOESY crosspeaks has been shown to be inversely proportional to r^6 (*11*), and can typically only be observed for distances of 4.5 Å or less. However, the planar chromophore does have two non-exchangable protons, H3 and H11, which are detectable using the standard NOESY pulse sequence. The nogalamycin H11 proton bridges the gap in the assignment pathway created by the increased base separation (Fig.2.13): previous studies (*210;241*) have shown that the H11 proton is in close proximity to the H1' and

H6/H8 nuclei of the nucleotides at the intercalation site. The sugar H1' protons form strong inter-molecular NOEs with the drug proton, giving rise to a new, extended assignment pathway in the H6/H8-H1' region of the NOESY spectrum, which can be traced out and related to the other DNA protons as before (Fig.2.14). Additionally, it offers a good starting point for the assignment of nogalamycin's 42 other non-exchangable protons. Unlike $d(CGATCG)_2$ and $d(TCGATCGA)_2$, the titration of nogalamycin into $d(CTCGATCGAG)_2$ did not give rise to a clean ^1H spectrum indicative of a single species in solution. Rather the spectrum (data not shown) was complex and broad, indicating that several species were present. Given that the 1D ^1H spectra of the shorter complexes (Fig.2.15), and of uncomplexed $d(CTCGATCGAG)_2$ were so clean and sharp, this was considered to be as a result multiple orientations of the bound drug (Section **2.4.3**).

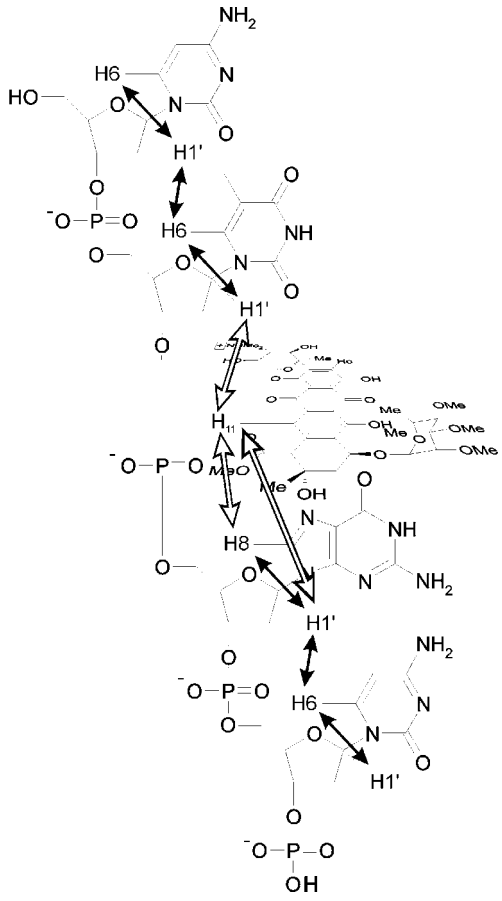


Fig.2.13: Schematic representation of the H6/H8-H1' assignment pathway of a DNA oligomer intercalated with nogalamycin.

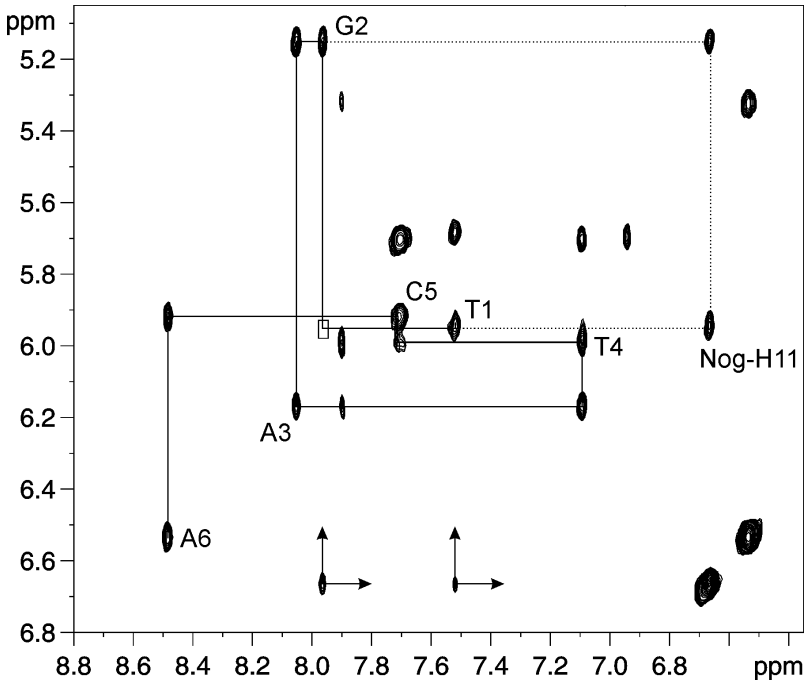


Fig.2.14: H6/H8-H1' region of the NOESY/DQF-COSY spectra of d(TGATCA)-nogalamycin (2:1) complex.

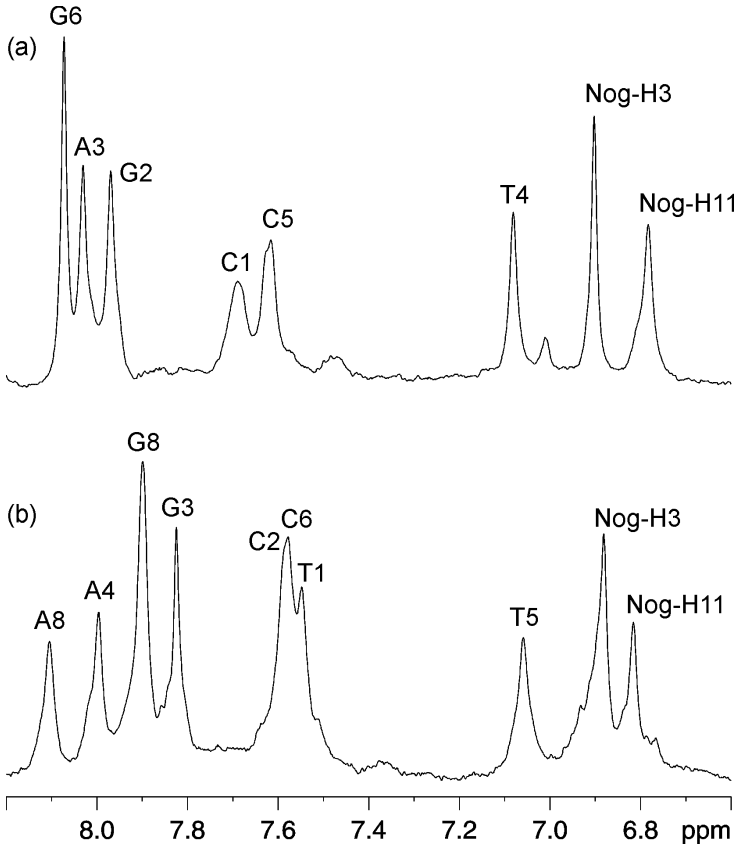


Fig.2.15: Aromatic region (ppm) of the 1D ^{1}H spectra of (a) nogalamycin-d(CGATCG)$_2$ and (b) nogalamycin-d(TCGATCGA)$_2$.

Nogalamycin always adopts the same orientation in a TpG intercalation site. The result of this is that the drug protons typically come into resonance in the same narrow band on the δ scale, their chemical shifts largely unaffected by the nature of the adjacent bases. Previous NMR studies of nogalamycin intercalated into DNA (*181;241*) prove extremely useful in assigning the complex studied here. Additionally, many nogalamycin nuclei show unusual properties that can be used to distinguish them from other protons with similar chemical shifts. For example, the tertiary amine of the aminoglucose sugar has a pK_a of approximately 9.0, and is consequently protonated at the neutral pH at which these experiments are carried out. At low temperatures, the *N*-methyl substituents of the drug intercalated in a TpG site show two sharp peaks at δ 2.8 and 3.0, because the rate of exchange of these two groups is slow. However, at room temperature, the rate constant of exchange increases and the peaks begin to broaden and coalesce. This phenomenon can be used to quickly identify these protons though they appear in a densely populated region of the spectrum.

2.4.3 Drug Orientation

Previous studies (*210;212*) have shown that when two nogalamycin molecules are intercalated into two symmetry-related TpG binding sites on a DNA hexamer, which are separated by a single ApT spacer, the drug molecules orient themselves such that their bulky sugars face towards the centre of the duplex. However, since nogalamycin always adopts the same orientation in a TpG intercalation site (Section **2.1.2**), there is no other possible orientation that the bound drugs can assume.

Since the CpG intercalation site has two-fold symmetry, nogalamycin can theoretically adopt two orientations within the site, both of which fulfil the necessary criterion of allowing each sugar to reside in the correct DNA groove. Though nogalamycin can theoretically bind in a terminally positioned CpG site such that its nogalose and aminoglucose sugars are oriented towards

the end of the oligomer, this is unlikely to happen due to so-called *end effects*. The NMR solution structures of the 2:1 complexes of nogalamycin-$d(CGTACG)_2$ have been determined by a quantitative treatment of two-dimensional NOE crosspeak intensities (*183*). The refined NMR structures retain major features of the crystal structure (*133*) in which the elongated aglycone chromophore is intercalated between the CpG steps with its nogalose and aminoglucose lying in the minor and major grooves, respectively, with each oriented towards the centre of the duplex. In this case, if the two sugars of each intercalated nogalamycins pointed away from each other towards the end of the complex, they would overhang the ends of the DNA. In the solid state, this would almost definitely affect crystal packing. In the solution state, the interaction between the aqueous solvent and the numerous hydrophobic groups on the drug sugars would be unfavourable and potentially preclude the formation of this conformation.

In the NMR-derived structures of the 2:1 complexes of nogalamycin with $d(GCATGC)_2$ (*181*) and $d(AGCATGCT)_2$ (*256*), the sugars of each intercalated drug point in opposite directions out towards the end of the DNA. Again, this is due to the unidirectional nature of the TpG intercalation sites, but it has been postulated (*212*) that even if these sites were replaced with their CpG cousins, to form $d(GCGCGC)_2$ and $d(AGCGCGCT)_2$, this observation would be the same due to steric effects that would arise because the intercalation sites are adjacent and not separated by an ApT spacer as before. This phenomenon has been demonstrated by Williams & Searle (*241*), who demonstrated that the hexamer $d(ATGCAT)_2$ is bound by nogalamycin in a 1:1 ratio despite the fact it possesses two adjacent intercalation sites. In this instance, steric effects caused by the insertion of the first drug molecule preclude binding of a second. Furthermore, the suggestion has been made that these CpG-containing complexes would be unaffected by end effects, as the intercalation sites are not at terminal locations.

When nogalamycin is bound at a TpG or CpG site, the methyl and methoxyl groups of the nogalose sugar have been shown to interact with the minor groove side of the DNA bases stretching up to three base-steps from the 3' face of the chromophore. This observation is used to determine the orientation of the drug within the intercalation site. As we have stated, when a CpG binding site is located at the terminal of a DNA oligomer, it has been shown that nogalamycin prefers to bind in an orientation with its sugars pointing towards the middle of the helix, because in the opposite orientation the sugars will protrude beyond the terminal CG base pair. However, in many of the hairpin structures studied recently (*41;69*), nogalamycin intercalates into a TpG site adjacent to the terminal base pair, with its sugars oriented towards the end of the oligomer. This raises the question of how far away a binding site must be from the terminus of an oligomer in order to negate the contribution of end effects to drug orientation, and whether there are other cooperative factors at work that influence the binding site orientation.

The orientation of the nogalamycin in the intercalation site can be verified by reference to drug-DNA NOEs in the NOESY spectra. An essentially complete assignment of drug and DNA resonances for all six of the complexes studied here provides the starting point for identifying intermolecular NOEs that define the position and orientation of the bound antibiotic.

Since nogalamycin can adopt only one orientation within a TpG site, the pattern of NOEs on the fully assigned NOESY spectra of $d(TGATCA)_2$, $d(CTGATCAG)_2$ and $d(CCTGATCAGG)_2$ proved to be a useful tool for determining the orientation of the drug in each of the CpG-containing sequences. Similar patterns on the NOESY spectra of $d(CGATCG)_2$, $d(TCGATCGA)_2$ and $d(CTCGATCGAG)_2$ would indicate that the drug molecules were oriented towards the centre of these oligomers, while differing patterns would suggest a different orientation. For each pair of TpG- and CpG-containing oligonucleotides, $d(TGATCA)_2$ & $d(CGATCG)_2$,

$d(CTGATCAG)_2$ & $d(TCGATCGA)_2$, and $d(CCTGATCAGG)_2$ & $d(CTCGATCGAG)_2$, the pattern of NOEs in key regions of their NOESY spectra were broadly similar, indicating similar drug orientations in each. The pattern of these interactions will be described in the context of the CpG-containing sequences.

The H6/H8-H1' assignment pathway of the $d(CGATCG)_2$ complex (Fig.2.16(a)) provides a strong indication of the orientation of the drug. If the intermolecular H1'↔nogalamycin-H11 and nogalamycin-H11↔H6 NOEs occur, as they do, across the C1pG2 step of the intercalation site, then the nogalose must point towards the centre of the helix. If the nogalose and aminoglucose did overhang the ends of the DNA, these intermolecular interactions would be observed across the complementary C5pG6 step, and such is not the case. This is also the case for the $d(TCGATCGA)_2$ and $d(CTCGATCGAG)_2$ complexes (Fig.2.16(b)), giving us an early indication that the mode of complexation is the same in these longer sequences.

Since deoxyribose H1' and H3' protons form part of the floor of the minor groove in *B*-form DNA, the presence of NOE crosspeaks between these protons on specific residues, and substituents on the nogalose sugar can be used to verify that the nogalose resides in the minor groove and determine whether it is oriented towards the centre or towards the ends on the duplex. The pattern of intermolecular NOEs between the methoxy protons at the 2', 3' and 4' positions of nogalose, and the H1' and H3' protons on the central A3, T4 and G5 nucleotides of $d(CGATCG)_2$ (Fig.2.17(a)) can only be satisfied if the nogalose resides in the minor grove and points towards the centre of the duplex. This observation holds for the complex of nogalamycin with $d(TCGATCGA)_2$ (Fig.2.17(b)). Additionally, in the longer sequence, there are no crosspeaks evident between the drug methoxy protons and those in the minor groove at the termini of the oligonucleotides. In the NOESY spectrum (298 K, 300 ms) of $d(TCGATCGA)_2$, for example, there are no interactions

between the H1' protons on the terminal bases, T1 and A8, and the nogalose methoxy protons. This shows conclusively that, in this case, the nogalamycin is oriented towards the centre of the duplex.

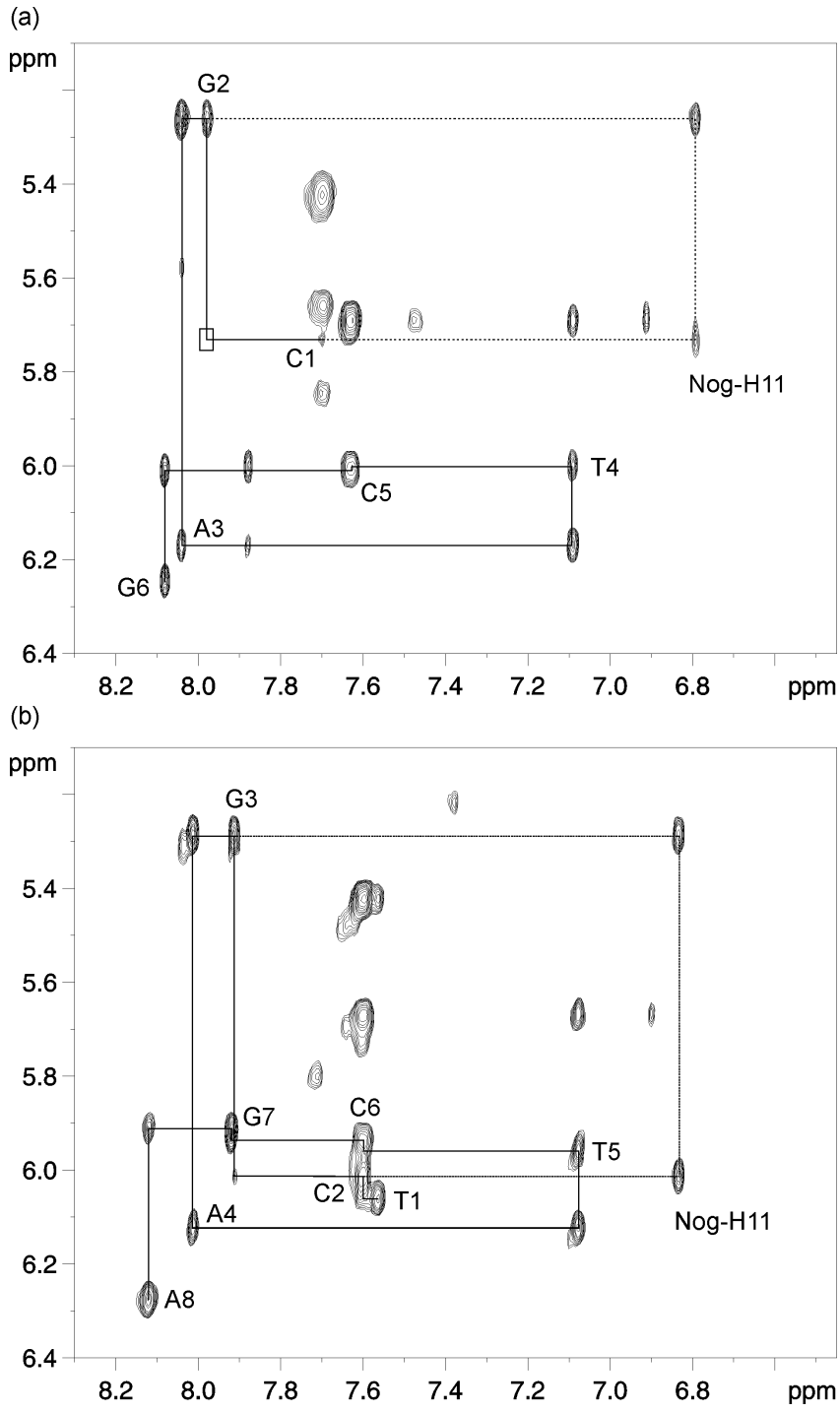


Fig.2.16: Portion of the NOESY spectrum (298 K, 300 ms) of (a) nogalamycin-d(CGATCG)$_2$ and (b) nogalamycin-d(TCGATCGA)$_2$, showing the H6/H8-H1' assignment pathway, and highlighting the drug H11 proton.

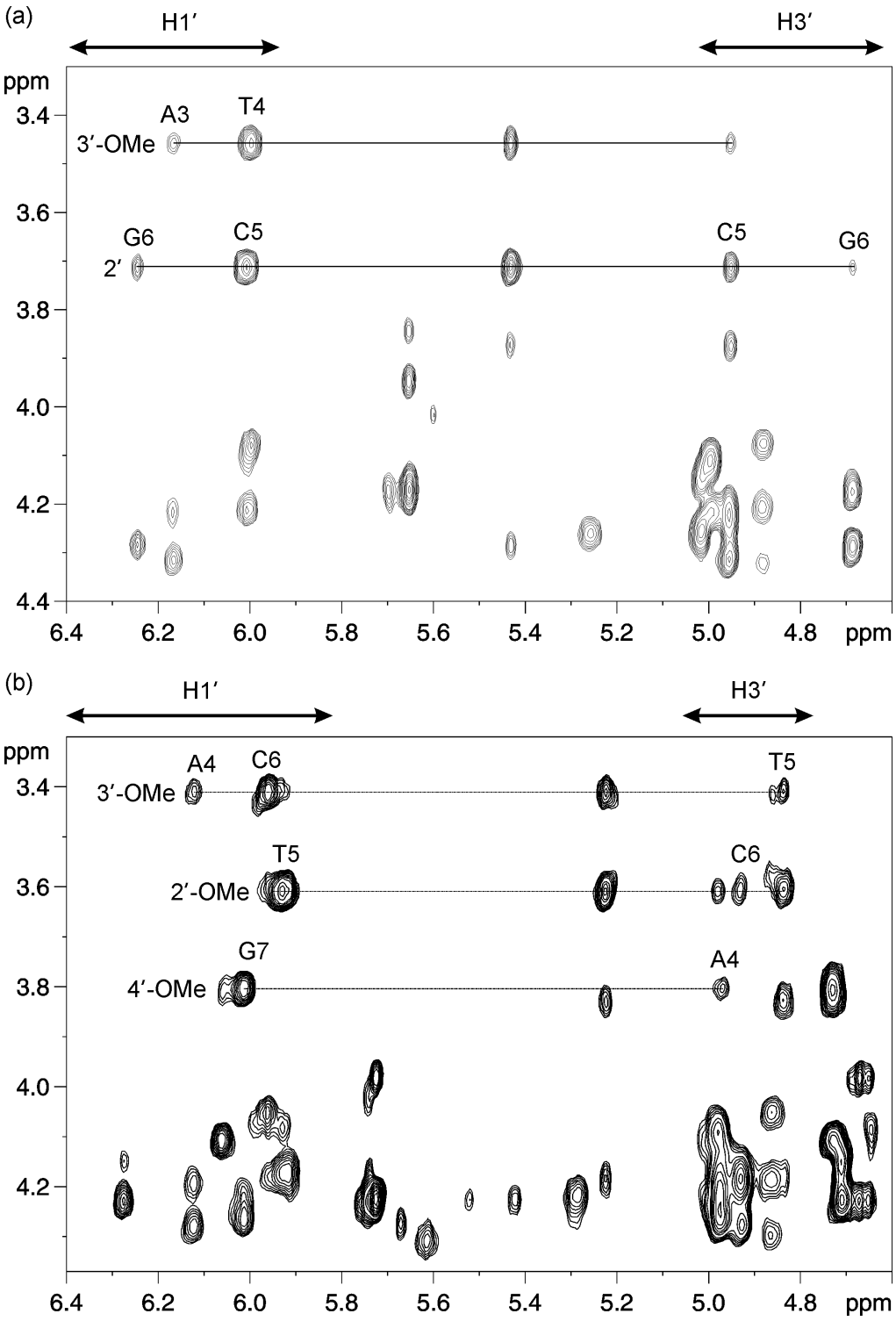


Fig.2.17: Portion of the NOESY spectra (298 K, 300 ms) of (a) $d(CGATCG)_2$ and (b) $d(TCGATCGA)_2$, highlighting interactions between nogalose methoxy protons and H1' protons on the floor of the minor groove in the conserved central portion of each oligomer.

The NOESY spectrum of nogalamycin-d(CTCGATCGAG)$_2$ proved impossible to fully assign, due to the presence of multiple species in solution. However, enough of the H1'-H6/H8 assignment pathway could be traced out to allow the interactions of nogalose with the floor of the minor groove to be examined. Although the NOESY spectrum (288 K, 300 ms, pH 6.8) is broad, tentative assignments can be made between the drug methoxy protons and DNA H1' protons, demonstrating that the drug intercalates in both orientations, apparently without discrimination (Fig.2.18). NOEs can be identified between the methoxy protons and both the central G4, A5 and T6 residues and the T2, C7 and A9 nucleotides at the ends of the oligomer, strongly indicating that at least two species, with nogalamycin oriented both towards the centre and towards the ends of the duplex are present.

In contrast to the extensive drug-DNA interactions in the minor groove, the number of contacts in the major groove is rather limited, though NOEs between the *N*-methyl protons and the central adenine and thymine H2" protons on the NOESY spectra of d(CGATCG)$_2$ and d(TCGATCGA)$_2$ (data not shown) are consistent with the orientation of the antibiotic suggested by the NOE patterns in the minor groove.

Since the orientation of nogalamycin remains unchanged by the addition of a single base pair at the ends of the d(CGATCG)$_2$ core, it would seem that this appendage is insufficient to encapsulate nogalamycin completely within the length of the oligomer when the drug is oriented towards its ends, and so counteract the end effects caused by the drug sugars overhanging the ends of the duplex and out into solution. However, when two base pairs are added at each end, the nogalamycin can be completely incorporated within the length of the oligonucleotide with no overhang, and so end effects no longer affect binding orientation. The presence of multiple species, with multiple binding orientations, indicates that each drug molecule is largely unaffected by the presence of the other in the adjacent site, and that end effects are the driving

force that determine the orientation of nogalamycin in terminally-located CpG intercalation sites. Since the sites are separated by two base pairs, it stands to reason that the drug molecules are far enough apart to prevent direct interaction between their sugars, as the same level of separation between drug and solvent is sufficient to overcome the end effects. The change to the topology of the DNA, which is necessary to facilitate intercalation, would appear to be localised, and not propagated to the adjacent site in a manner that affects binding orientation.

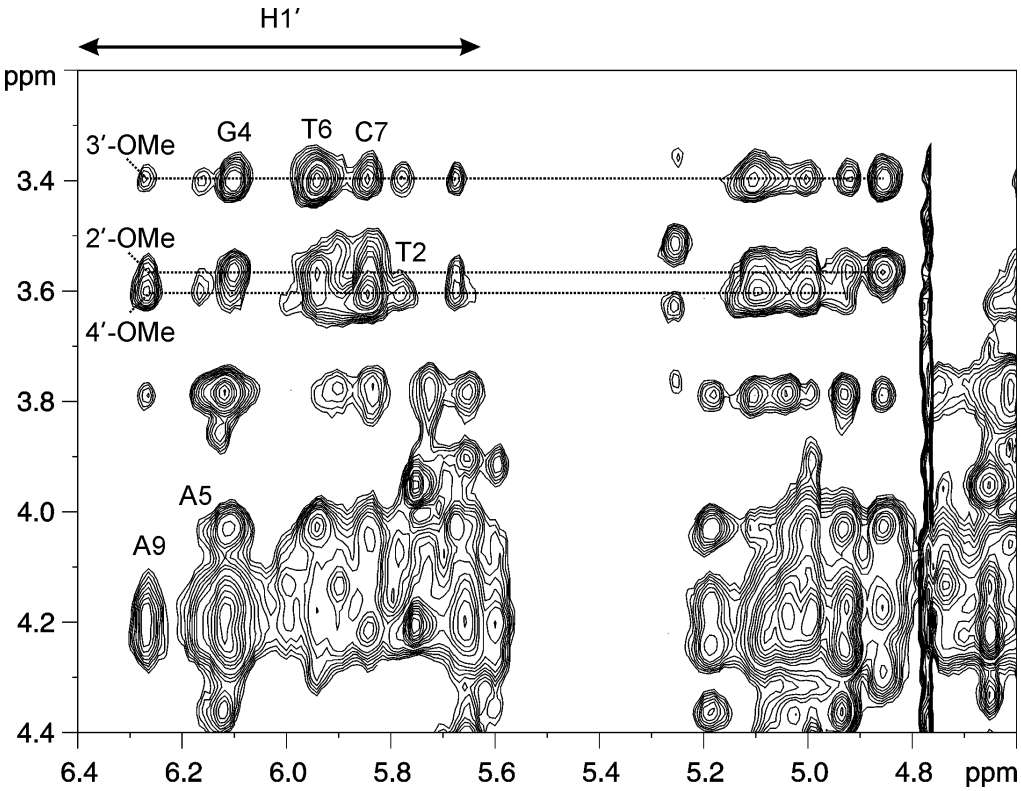
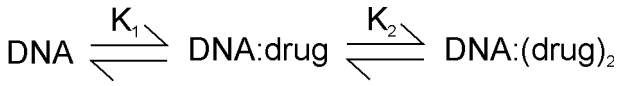


Fig.2.18: Portion of the NOESY spectrum (288 K, 300 ms) of d(CTCGATCGAG)$_2$, highlighting interactions between nogalose methoxy protons and H1' protons on the floor of the minor groove.

2.4.4 Evidence for Cooperative Interactions

If we consider a system in which two equivalents of a drug are titrated into a DNA oligonucleotide containing two binding sites specific to that drug, then the rates of formation of each species can be defined in terms of two equilibrium binding coefficients, K_1 and K_2, as described in Scheme **2.1**. By

defining a cooperativity coefficient, φ, equal to K_2/K_1, we can formulate a simple definition of what constitutes cooperativity: for independent sites $K_2=K_1$, and consequently φ=1. Similarly for cooperatively associated sites, $K_2>K_1$ and φ>1, while for anticooperative interactions φ<1.



Scheme 2.1: Binding of a drug to a DNA oligomer with two drug specific binding sites.

In an attempt to assess qualitatively the degree, if any, of cooperativity between two nogalamycin binding sites separated by an ApT spacer, we have constructed three binding models, which relate binding site occupancy at different drug:DNA ratios to cooperativity. In the first model the drug partitions between two equally probable binding sites (φ=1). In the second model, binding of the first drug increases the probability of a second binding event to unity (φ>>1), and in the third system, this probability is decreased to zero by the first binding event (φ<<1).

As described in Section **2.3.5**, 1D ^{1}H NMR spectra of the forming drug-DNA complexes were collected following the titration of aliquots of nogalamycin. Well-separated resonances of protons from each of the free DNA, 1:1 and 2:1 drug:DNA complexes were identified (Section **2.4.1**) and integrated for each step of the titration. To plot a graph of the binding profile of nogalamycin with $d(TGATCA)_2$, for example, three peaks, representing one proton from each of the free DNA, 1:1 complex and 2:1 complex, were chosen and monitored throughout the titration (Fig.2.6). The area under each peak was measured at each stage of the titration and used to quantify the percentage of each species present. This allowed each step of the titration to be treated as a separate system, independent of any other step, and thus unaffected by any possible errors in experimental procedure or data handling. The relative concentrations of nogalamycin and the oligonucleotides were plotted as a function of the

mole fraction of binding sites occupied. For comparison with the three binding models. If we consider a system with two equally probable binding sites that are not cooperatively linked (Fig.2.19(a)), and define *x* as the mole fraction of drug per binding site, we can derive binding curves that describe each.

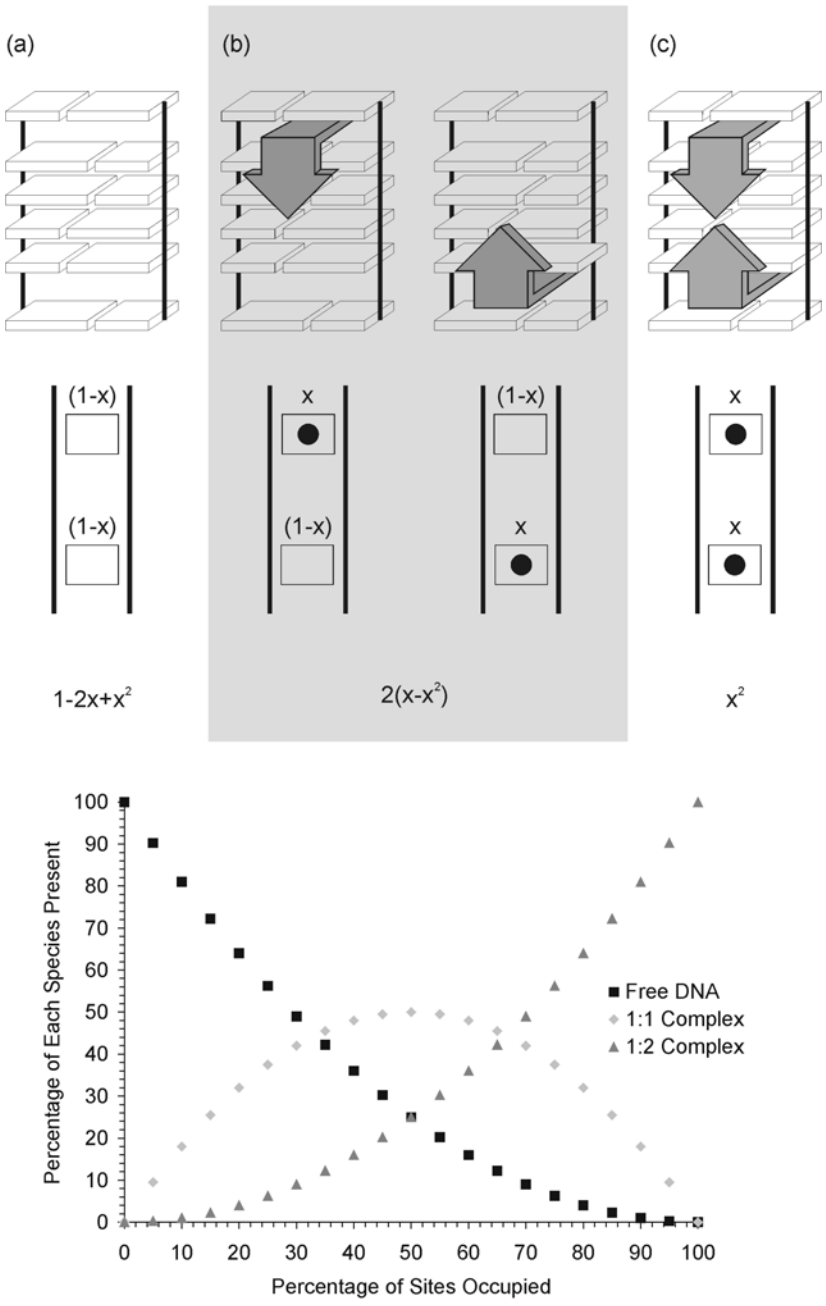


Fig.2.19: Theoretical partitioning of nogalamycin between two equally probable binding sites. The graph shows free duplex replaced by 1:1 and 2:1 complex.

Assuming that the two binding sites are entirely independent of each other, it would be expected that, as the titration continues, the amount of free DNA would fall continuously. Conversely, the amount of the 2:1 complex, would be expected to increase as the titration progresses. The 1:1 complex, being an intermediate species and the first step in the production of the final product, should initially rise, as there is an excess of the free DNA from which it is formed early in the reaction. However, as the titration continues, the free DNA becomes depleted and, consequently, the rate of production of the 1:1 complex drops. At the same time, following an initial lag-phase, the production of 2:1 complex will increase as the concentration of 1:1 complex decreases. The combination of these two factors will lead to a maximum concentration of the intermediate 1:1 species half-way through the reaction, followed by its inevitable depletion (Fig.2.19).

Cooperative drug binding in adjacent binding sites, as we have defined it for this qualitative study, can be identified by a reduction in intensity or absence of any recordable resonance of the 1:1 complex, and by the presence of the 2:1 complex from low drug:duplex ratios during the titration. Since neither site on an individual duplex can be occupied independently of the other, pairs of drugs and pairs of binding sites can be treated as single entities (Fig.2.20). Upon drug titration, resonances corresponding to the free DNA should decrease while new resonances corresponding to the 2:1 complex appear. At the point of one equivalent of drug per duplex, we should be able to detect resonances corresponding to both the free and bound DNA. Observing easily distinguished regions of the spectrum, such as the methyl region and the aromatic region, enables new resonances to be followed. Typically, resonances of the complex formed can also be observed in previously unoccupied spectral regions (*241*). Addition of further quantities of the drug would be expected to result in a decrease in the intensity of the signals of the free DNA until, at a ratio of two nogalamycin molecules per duplex, a single species corresponding to the 2:1 complex is retained (Fig.2.20).

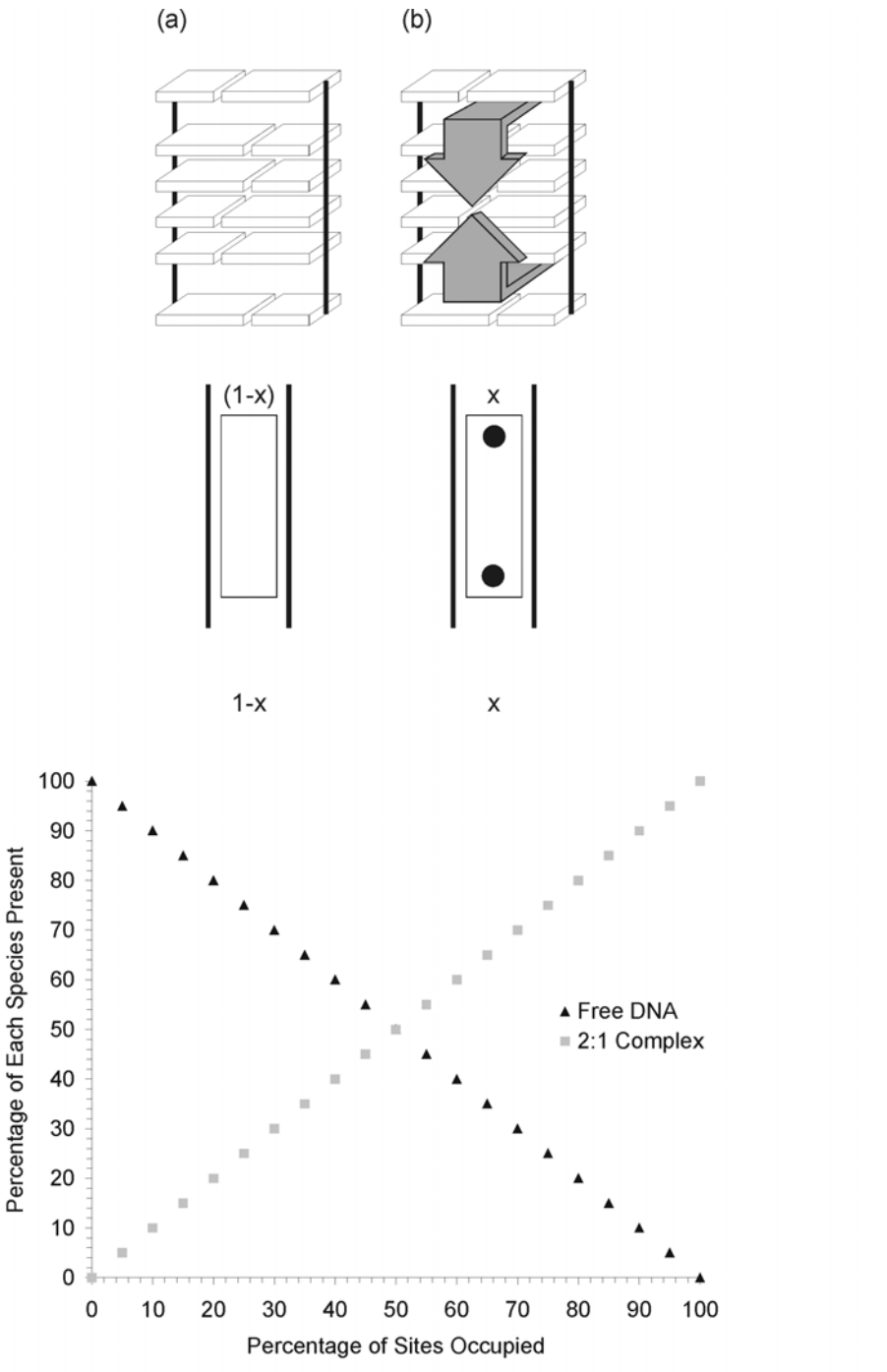


Fig.2.20: Cooperative binding of nogalamycin into two adjacent binding sites. The graph shows free duplex replaced by 2:1 complex.

The shape of the anticooperative and cooperative binding curves are very similar. However, given that two binding sites on the same duplex cannot be occupied simultaneously in the anticooperative model, only 50% of the total available sites will be occupied at the end on the titration (Fig.2.21). As such,

the slopes of the graphs describing anticooperative binding will have twice the magnitude of those in the cooperative model, in which all sites are occupied following titration with nogalamycin. This is based on *ad hoc* assumptions about the nature of the final products and, as such, cannot be used to distinguish between the two modes of binding. For the anticooperative model, the addition of 2 equivalents of nogalamycin to the duplex does result in loss of the 2-fold element of symmetry upon complex formation. Each signal on the 1D ^{1}H NMR spectrum should therefore be replaced by two corresponding resonances for the complex in an anticooperative system. This observation can be used to distinguish between the various modes of binding.

If no communication exists between the sites, then our first model (Fig.2.19) should hold. Since this model is empirically different to the other two, no additional means of identification is required. However, the curves representing cooperativity and anticooperativity are very similar and must be distinguished by a combination of graph-fitting and examination of the ^{1}H spectrum. If there is cooperativity between the sites, only two species will be detectable in the ^{1}H NMR, one corresponding to the free DNA and one to the emerging 2:1 complex, as a second drug introduced to the system after the first binding event will preferentially bind to the second site in the same DNA ligand, rather than bind to a different piece of free DNA. The ^{1}H spectrum of the complex should have one signal for each DNA proton, as the dyad symmetry is preserved. If there is an anticooperative relationship between the binding sites, again only two species will be detectable as the titration progresses. In this instance, formation of the 2:1 complex is precluded by the formation of the 1:1 complex. Therefore, each DNA proton signal in the uncomplexed species should have two corresponding resonances in the ^{1}H NMR of the complex, as the 2-fold symmetry do not exist in the 1:1 complex.

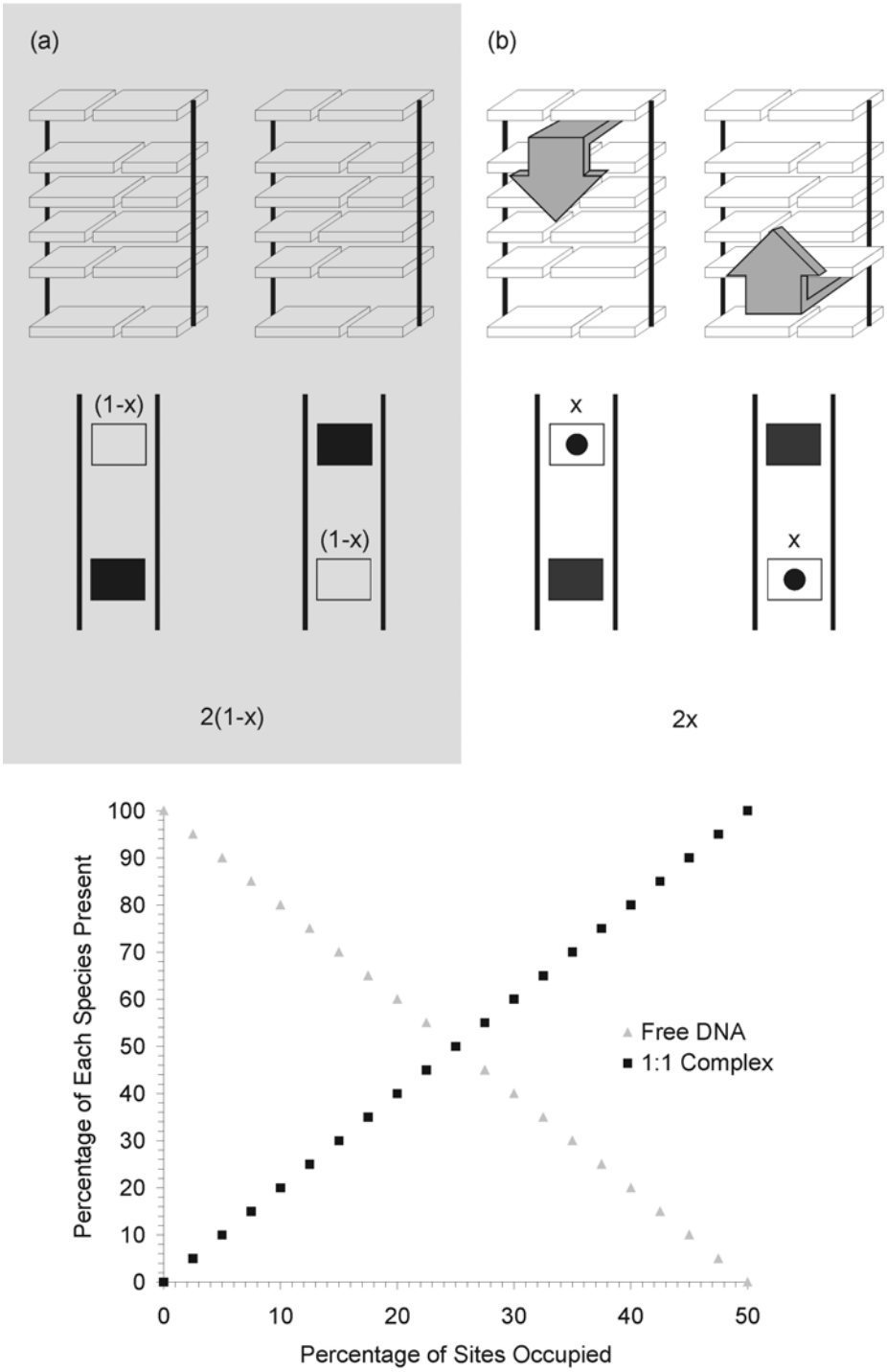


Fig.2.21: Anticooperative binding of nogalamycin into two adjacent binding sites. The graph shows free duplex replaced by 1:1 complex.

These models describe the extremes of cooperativity and anticooperativity ($\varphi >> 1$ or $\varphi << 1$). However, even these simple models can be applied to more complicated systems, where the degree of the cooperative interaction lies

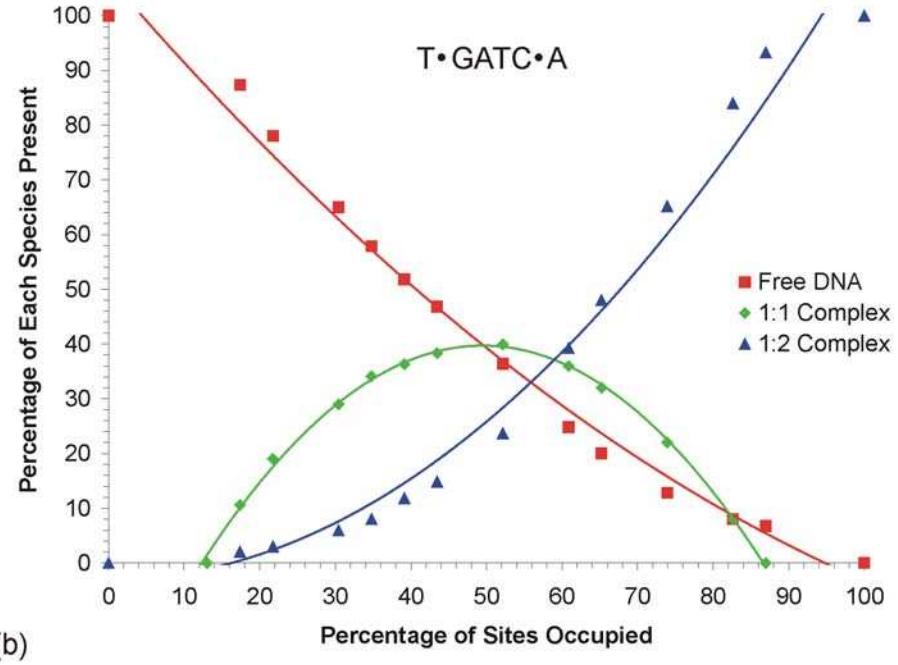
somewhere between the two extremes of total preclusion or necessary inclusion of the second binding event. As the degree of cooperativity increases, the apex of the curve representing the 1:1 complex (Fig.2.19) would lower, while the curves representing the free DNA and 2:1 complex would begin to lose their parabolic property, flattening out to become linear when the peak level of the 1:1 species is zero (Fig.2.20).

The oligonucleotides, $d(TGATCA)_2$, $d(CTGATCAG)_2$, $d(CCTGATCAGG)_2$, $d(CGATCG)_2$, $d(TCGATCGA)_2$ and $d(CTCGATCGAG)_2$, were each titrated with nogalamycin as described. In each case, as the titration continued, the amount of free DNA fell continuously, while the amount of 2:1 complex increased. The 1:1 complex was detectable throughout each titration. The 1:1 complex initially rose, and was then depleted, consistent with an intermediate species in the production of the 2:1 complex. The level of production of the 2:1 species was not constant throughout the duration of the complex, nor was the rate of depletion of free DNA. Rather, production of the 2:1 species increased markedly during the second part of the titration, when the 1:1 species began to deplete. Conversely, depletion of the uncomplexed species is greatest during the early part of the titration, when 1:1 complex production is greatest. The combination of these two factors will lead to a maximum concentration of the intermediate species half-way through the reaction. In each case, the peak concentration of 1:1 complex occurs at between 48% and 55% of total site occupancy, and is in the range 43-50%. The total concentration of the 1:1 species never exceeds 50%, which is in good agreement with our models for independent binding sites. This simple, empirical analysis of the titration curves clearly displays a system with two independent binding sites. The data were fitted by least squares regression analysis to a polynomial function (**Equation 2.2**). In each case the R^2 value was in excess of 0.98, indicating a good fit of the data to the curves. None of the curves described display any characteristics that would suggest any degree of cooperativity or anticooperativity.

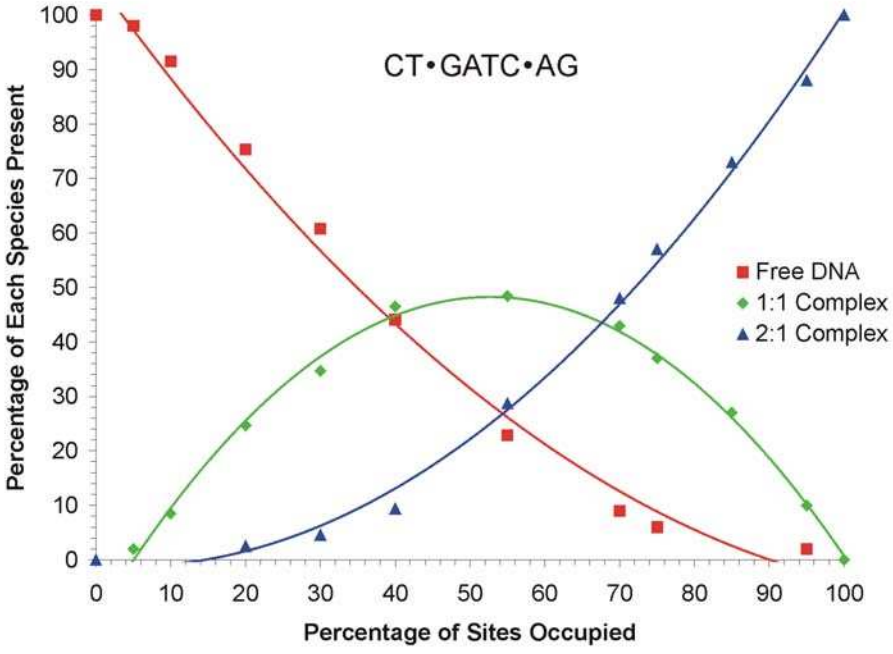
$$y = a_1 + a_2 x + a_3 x^2$$

Equation 2.2: Second order polynomial fitted to titration curves. a_1, a_2 and a_3 are integers.

(a)

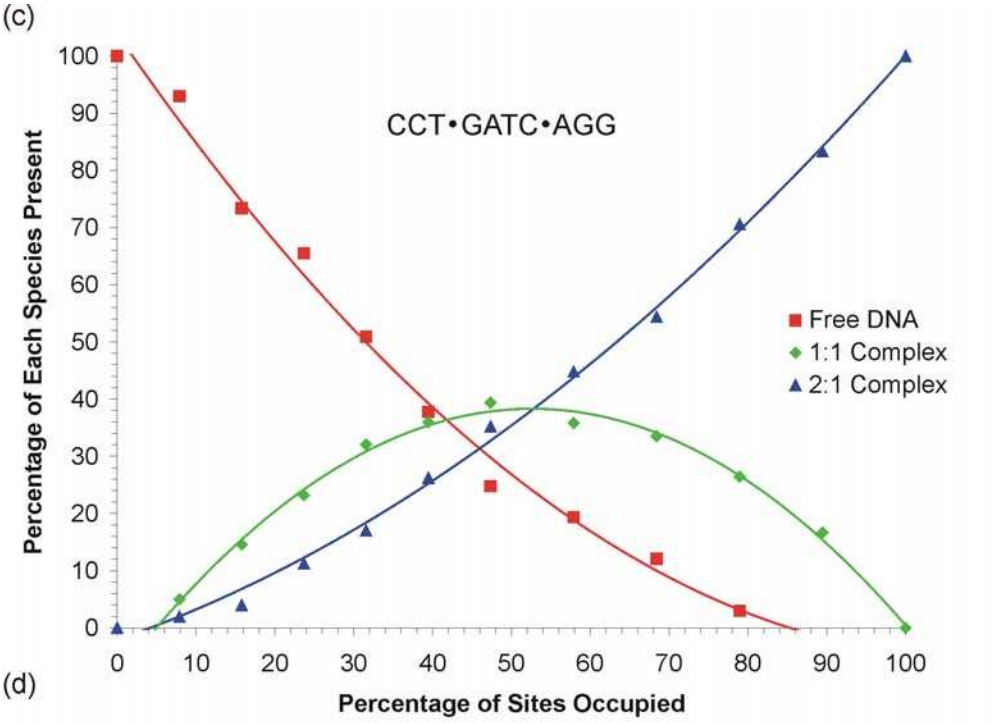


(b)

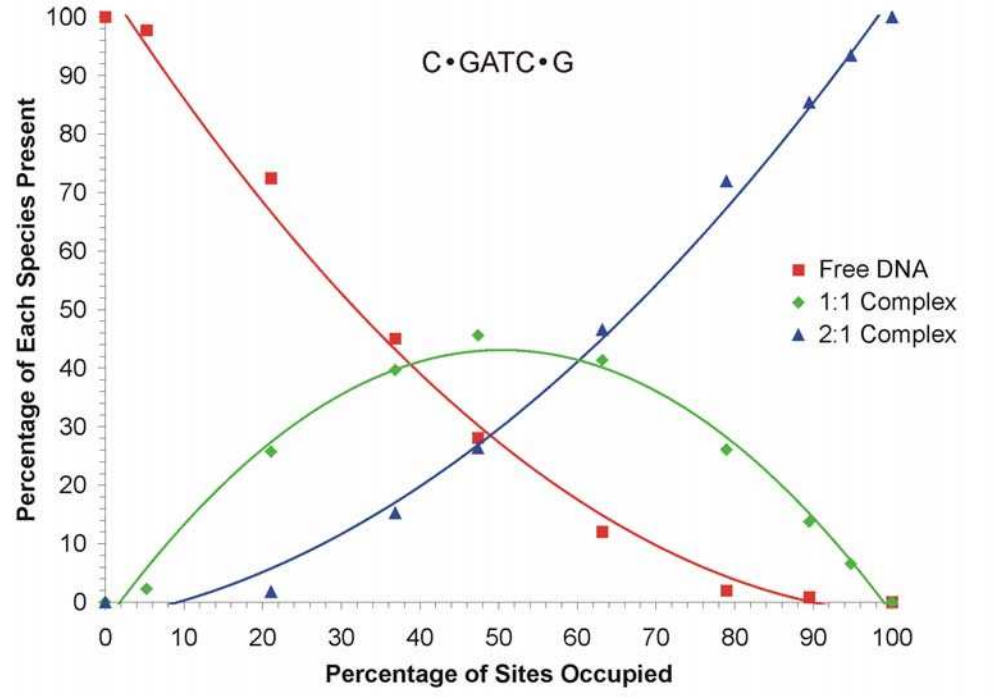


contd.

(c)


CCT•GATC•AGG
Free DNA
1:1 Complex
2:1 Complex
Percentage of Each Species Present
Percentage of Sites Occupied
0
10
20
30
40
50
60
70
80
90
100

(d)


C•GATC•G
Free DNA
1:1 Complex
2:1 Complex
Percentage of Each Species Present
Percentage of Sites Occupied
0
10
20
30
40
50
60
70
80
90
100

contd.

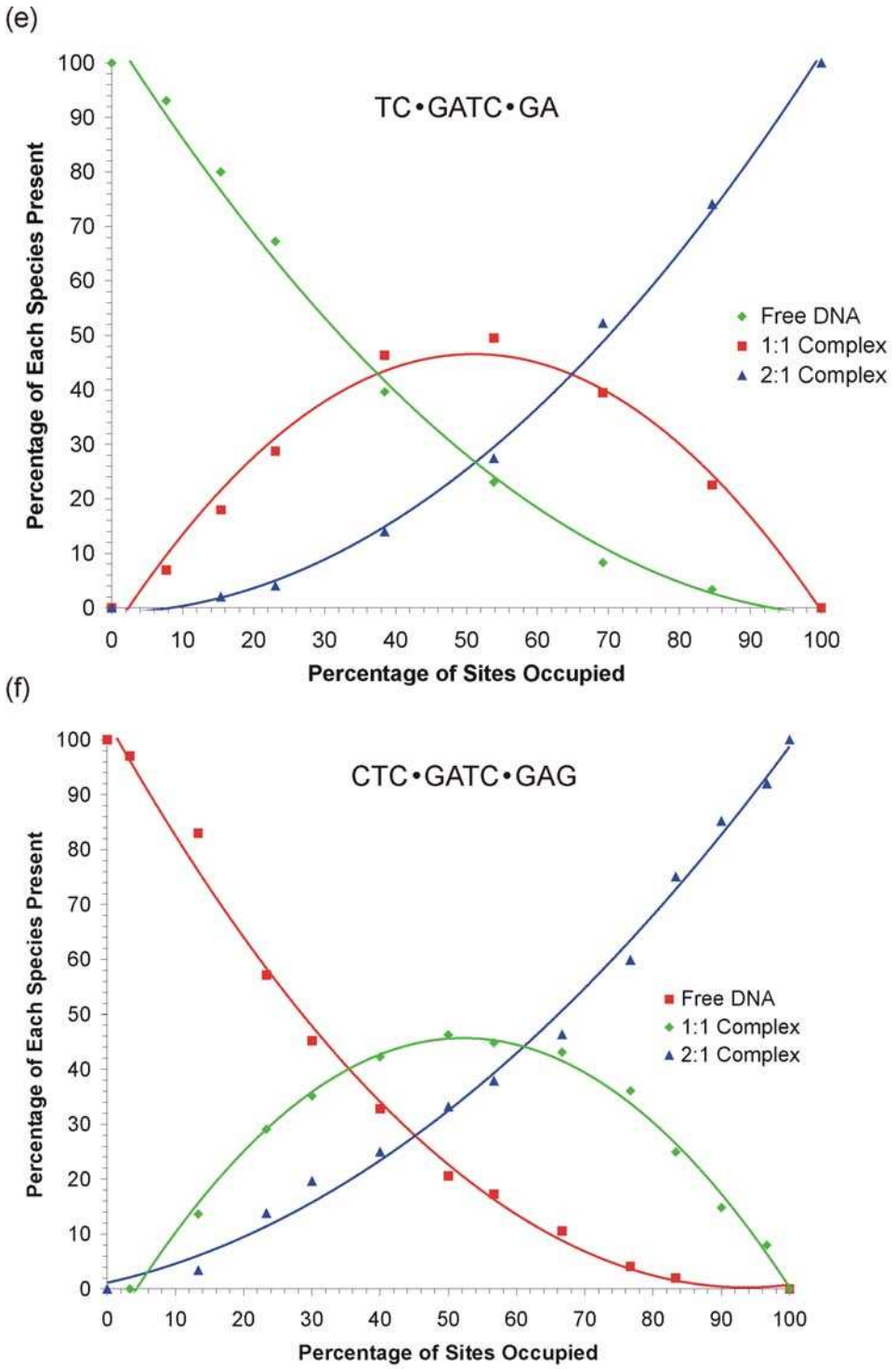


Fig.2.22: Actual partitioning of nogalamycin between two adjacent binding sites on (a) $d(TGATCA)_2$, (b) $d(CTGATCAG)_2$, (c) $d(CCTGATCAGG)_2$, (d) $d(CGATCG)_2$, (e) $d(TCGATCGA)_2$ and (f) $d(CTCGATCGAG)_2$.

2.5 Conclusions

We have demonstrated here that end effects are the primary driving force for determining the orientation of nogalamycin in neighbouring, terminally located, CpG intercalation sites. Two base pairs provide sufficient separation to completely disconnect nogalamycin from the solvent when it is oriented such that its sugars point toward the end of a DNA oligomer. This allows each drug to intercalate in two orientations in its symmetrical CpG(CpG) intercalation site, each of which satisfy the requirement for nogalose to reside in the minor groove. It is unclear whether a conformer exists where both drugs are oriented towards the same end of the oligomer. No previous studies have been reported in which this has been observed. We can, however, state that two base pairs provide adequate separation between binding sites such that the intercalation event at one site does not affect the other.

Non-adjacent intercalation sites bind nogalamycin in a non-cooperative manner. The inclusion of an ApT spacer between two binding sites, either TpG or CpG, effectively allows them to achieve independence in terms of drug binding, which cannot be achieved in sequences with adjacent sites (*256*). This property of site independence is unrelated to the orientation of the bound drugs, as demonstrated by the binding curves of d(CTCGATCGAG), in which the drug can adopt multiple orientations. A minimised structure of the 1:1 complex of nogalamycin with $d(TGATCA)_2$ shows that the unoccupied TpG site is qualitatively similar to those found in the uncomplexed *B*-like DNA (data not shown). The absence of any cooperativity between the sites would suggest that the second TpG site is largely unaffected by the intercalation event at the first, or at least not in any way that affects its binding affinity for the drug.

In this study we have determined a family of high-resolution structures of each of the complexes of nogalamycin with $d(TGATCA)_2$ and $d(CCTGATCAGG)_2$. The structures of nogalamycin-$d(TGATCA)_2$ and nogalamycin-

d(CCTGATCAGG)$_2$ are broadly similar, and the conserved central residues and intercalated drugs can be superimposed with an RMSD fit of 0.67 (±0.1) Å.

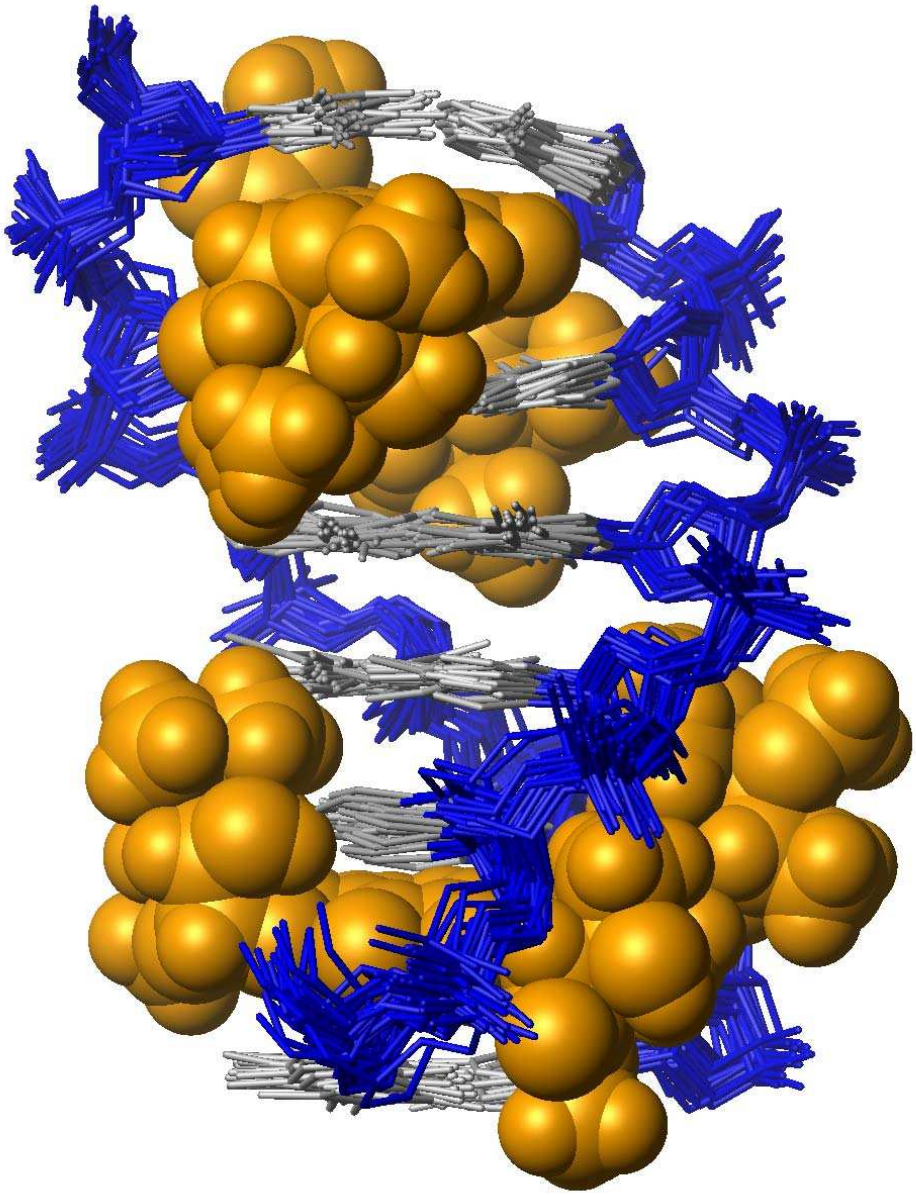


Fig.2.23: 20 randonly chosen DNA structures from the last 500 ps of a 2 ns rMD simulation of nogalamycin-d(TGATCA)$_2$ superimposed on the averaged structure of nogalamcyn from that portion of the rMD simulation.

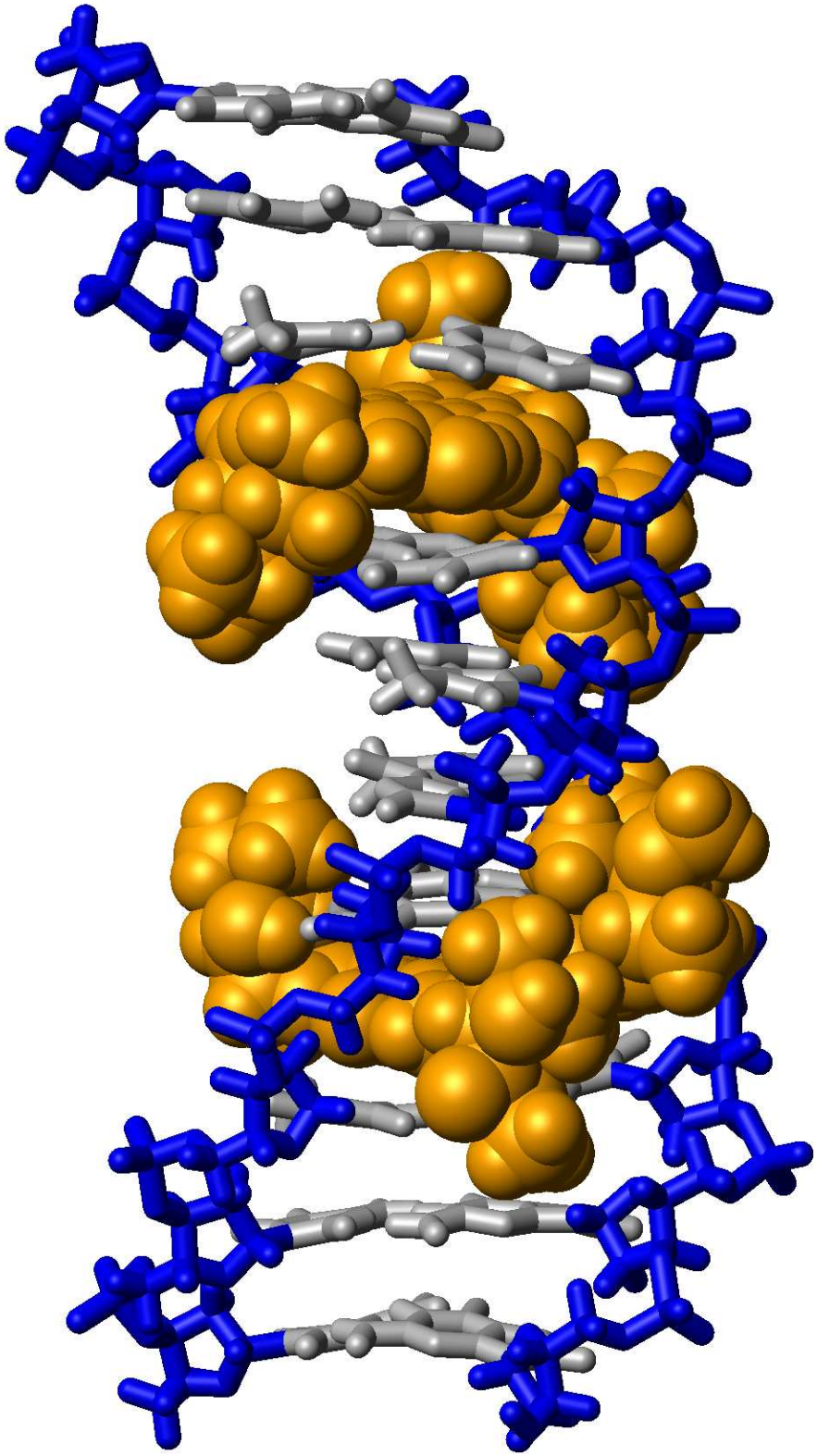


Fig.2.24: The average minimised structure from the last 100ps of a 1ns rMD simulation of nogalamycin-d(CCTGATCAGG)$_2$.

3. DNA Hairpin Loops and Bulges; Implications for Frameshift Mutagenesis

3.1 Introduction

3.1.1 Hairpin Loops in Nucleic Acids

Hairpin loops are a basic unit of nucleic acid structure and have been shown to play important roles in a great number of important biological functions (*251*). Hairpins comprise a single-stranded loop region closed by a base-paired stem (Fig.3.1). The stability of hairpin structures have been shown to increase with the number of base pairs in the stem region (*27*). Thus, long complementary regions flanking either side of the loop region are good predictors of loop formation. However, recent studies have shown that certain DNA and RNA hairpins, which contain specific loop sequences, are unusually stable (*229*).

Hairpin structures in DNA and RNA, consisting of stem and loop regions, occur naturally in single-stranded DNA and RNA. These structures are considered to have important biological functions (*98*). The structures of unusually stable hairpins have been extensively studied, including the tetra-nucleotide loops d(GAAA) (*100*), d(GTTA) (*96*), r(U*N*CG) (*37;185;226*) and r(G*NR*A) (*94*) (where *N* is any nucleotide and *R* is a purine nucleotide). The latter two structures have been identified in ribosomal RNA (*37;94*), RNA coliphages (*3*), hammerhead ribozymes (*95*), group I catalytic introns (*148*), and in bacteriophage T4 mRNA (*226*).

DNA sequences, such as d(CGGAAGC) and d(CGGAAGC) form extraordinarily stable DNA mini-hairpins (*99;101;251*). Hirao *et al.* (*101*) compared a variety of sequence variants of the d(GCGAAGC) and r(GCGAAGC) hairpins. Rapid mobility of these structures in denaturing electrophoretic conditions was observed. The GAAA and GAA loops were found to be more stable than hairpins of other sequences and the different stabilities noted for DNA and RNA mini-hairpins were proposed to arise from

the different conformations of the stem structure in each. The structure of the d(GCGAAGC) oligonucleotide was determined by NMR spectroscopy (*98*). The determined hairpin structure was folded back between the A4 and A5 adenosines, and stabilised by two Watson-Crick base pairs in the stem region, and an additional non-Watson-Crick *wobble* G-A base pair in the GAA loop (Fig.3.2). Extensive base stacking interactions in the stem region were observed to further stabilise the hairpin structure.

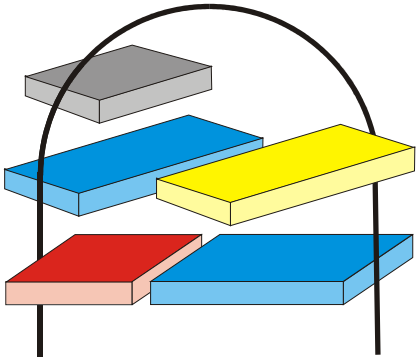


Fig.3.1: Schematic of the DNA mini-hairpin. Cytosines are shown in red, guanines in blue, and adenines in yellow.

The d(CGGAAGC) and d(CGGAAGC) hairpin sequences have been shown to occur frequently in biologically important regions, such as replication origins and promoter regions. The GAAA hairpin loop exists in the replication origins of G4 phage ssDNA (*97*). The GAA loop has been identified in the replication origins of herpes simplex virus (*57*) and bacteriophage φX174 DNA (*6*), and in the promoter region of *E-coli* heat-shock gene (*48*).

In a study by Yoshizawa *et al.* (*251*), all 64 variants of the sequence d(GC*NNN*GC), where *N* is one of the four DNA bases, were synthesised and characterised in terms of their ability to form stable hairpin structures. Mobility studies on polyacrylamide gels were again conducted to determine which DNA fragments formed a more compact structure. They also examined the resistance of each fragment to exonuclease activity using DNA polymerase. NMR studies allowed the elucidation of much structural

information and highlighted the importance of the base-stacking interactions. Overall, these workers found that only hairpins with G*N*A loops formed extraordinarily stable trinucleotide-loop hairpins. The hairpin structures described in this paper have increased thermostability and, again, show high resistance against nucleases. The four fragments containing GAA, GGA, GCA and GTA loop sequences, d(CGG*N*AGC), were shown to have T_m values of 76, 73, 71 and 70 °C respectively. Some fragments, other than GNA hairpins, also formed loops, but these were markedly less stable, with T_m values in the range 47-59 °C. The enhanced stability of G*N*A loops is proposed to arise as a result of the formation of a sheared G-A pair (Fig.3.2).

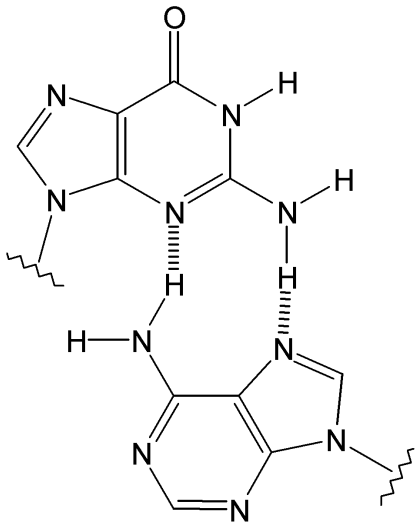


Fig.3.2: Sheared G-A base pair.

Moody & Bevilacqua (*150*) investigated whether the network of interactions involved in G*N*A loop folding and stabilisation communicated with each other to facilitate loop formation. On the basis of base substitutions, they proposed that all interactions were interdependent and that these interactions formed in a highly co-operative manner. They demonstrated that substituting either the guanine or the adenine in the sheared G-A pair for inosine, and thus deleting a single N–H···N hydrogen bond, not only weakens the H-bonding array, but also weakens other interactions. Based on these observations, we can propose

that the energetic contribution of a hydrogen bond in the context of a G-A base pair is artificially elevated by the co-operativity of loop formation.

Other G-A functional group substitutions, which do not affect H-bonding, but do affect the electrostatic surface potential (ESP) of the bases, indicate that the vertical interaction of the G-C closing base pair with the guanine at the first position of the loop accounts for most of the hairpin stability (*151*) . These investigations further suggest that these stacking interactions are mediated in part by the carbonyl and imino functional groups of the loop G. By combining the data from these two sets of mutations, it is possible to deduce that favourable stacking interactions between the closing G-C pair and the loop guanine form the major stabilising factor in G*N*A loops; but that these interactions are facilitated by H-bonding within the loop, which hold the guanine in a complimentary orientation, by means of a sheared base pair.

3.1.2 Summary of Structural Properties

The examination of the structures of numerous nucleic acid hairpins have shown that the exceptional stability observed for such structures is a consequence of several stabilising interactions, the most obvious of which are the Watson-Crick base pair interactions in the stem region of the hairpin. Non-Watson-Crick base pairs, base-phosphate and base-sugar contacts further stabilise the loop section. The loop of the G*N*A mini-hairpin is composed of a single nucleotide, *i.e.*, the central *N*. The first (G) and third (A) nucleotides take part in a non-Watson-Crick base pair and the entire structure is stabilised by base stacking interactions.

3.1.3 Therapeutic Applications of DNA hairpins

The property of nuclease resistance makes these hairpin structures amenable to the stabilisation of oligonucleotides. Poddevin *et al.* (*174*), synthesized a series of 20-mer antisense phosphodiester oligonucleotides constituting of a 5'-dodecameric sequence, complementary to the acceptor splice junction of

herpes simplex virus type 1 (HSV-1) pre-mRNAs, flanked on the 3' end by octameric sequences adopting hairpin-like structures of different stabilities. The presence of the minihairpins on the 3' end of the 20-mer phosphodiester oligonucleotides protected them against serum nuclease degradation, this protection being well correlated to the reported melting temperatures of the minihairpins, and to the gel mobilities of the 20-mer oligonucleotides. The addition onto the 3' terminal of the most stable minihairpin increased more than eightfold the nuclease resistance of the linear antisense dodecamer. This method has been further applied to the protection of antisense DNA oligonucleotides (*4;118;119*) and primers. DNA oligonucleotides harbouring a hairpin sequence at the 3' terminus have been shown to prime the amplification of a wide range of templates ranging from plasmid DNA to plant and animal genomes (*28*).

3.1.4 Bulged Bases in Duplex DNA

Bulged structures in DNA are of general biological significance and are potential targets for therapeutic drugs. They are common, non-helical features of RNA and DNA, and they may vary from single nucleotide bulges to multiple base bulges. Bulges are known to be important in RNA for protein binding and tertiary folding (*31;32*). In DNA, bulged bases play a role in frameshift mutagenesis in sequences with repeating base pairs (*158*). DNA duplex regions often contain bulge loops in which there are one or more consecutive bases on one strand with no complementary bases, with which to form base pairs, on the other strand. The smallest bulge loop has one unpaired base and is termed a single nucleotide bulge. Structural studies have shown that single base bulges can assume two different conformations: They may form stacked-in structures, in which the unpaired base is stacked between adjacent bases, or looped-out structures, in which the unpaired base faces away from the helical axis. There have been relatively few reports involving the characterisation of the interactions of bulged oligonucleotides with ligands.

3.1.5 Adenine Bulges

Patel et al. (*163*) collected proton and phosphorus NMR parameters of the 12-mer duplex d(CGCGAATTCGCG)$_2$ and the 13-mer duplex d(CGC^{A}GAATTCGCG)$_2$ to determine whether the extra non-complementary dA between base pairs 3 and 4 in the 13-mer duplex would disrupt the formation of a stable helix, loop out into solution or stack into the helix without causing a disruption. They observed that the inclusion of an additional dA residue into the helix did not rupture the adjacent Watson-Crick base pairing interactions, and that the adenosine stacked into the 13-mer duplex rather than form a bulge loop. The H8 and H2 protons of the extra adenosine underwent large upfield shifts on formation of the 13-mer duplex from the melted single strand. This unambiguously demonstrates the base stacked into the duplex, with the ring currents from the adjacent dG and dC base pairs shifting the adenosine base protons upfield. A 15 ºC decrease in thermal stability for the duplex state was induced by the presence of the stacked, extra adenosine residue. A similar study (*90*) was carried out on the sequence d(CGC^{A}GAGCTCGCG)$_2$, in which the central AATT segment of Patel's oligomer is replace with an AGCT segment. Again the extrahelical dA stacked in without causing disruption to the flanking base pairs, but the duplex-strand transition midpoint was elevated, consistent with the introduction of two stronger C-G base pairs

Furthermore, it has been demonstrated that an extrahelical adenosine stacks into the helix at all temperatures below the onset of the melting transition in solution, regardless of whether the non-pairing base is flanked by cytidines or guanosines (*108*). For the self-complementary sequence d(CCG^{A}GAATTCCGG)$_2$, NOEs were observed between the H2 proton of the extra dA and the imino protons of the flanking guanosines. The oligomer d(CCGGAATTC^{A}CGG)$_2$, in which the extrahelical base is sandwiched between two cytidine resiues, revealed through-space interactions between the H2 proton of dA and the imino protons of the guanosines that were base paired

with the flanking cytidines in the NOESY spectrum. In each case the extra adenosine bulge site adopts a wedge conformation (Fig.3.3), confirmed by observed NOEs between adjacent base protons in the cytosine-cytosine in the former (GAG) 13-mer, and between protons in the guanosine-guanosine step in its sister (CAC) oligomer. Interestingly, these NMR data are contradicted by X-ray crystal structures published at about the same time (*106;149*), in which the extrahelical adenosine adopts a looped-out formation. This reflects a fundamental difference between solution and solid-state environments.

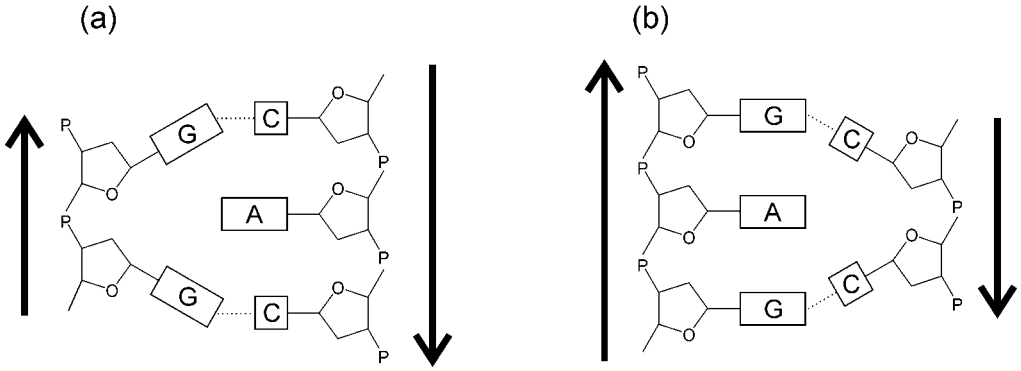


Fig.3.3: Extrahelical adenosine flanked by cytosines (a), and guanosines (b).

3.1.6 Guanine Bulges

Like extrahelical adenosines, unpaired guanosines in B-form DNA duplexes have been shown to stack into the helix in a wedge conformation (*247*). In each case the helix axis was bent by 18-23 ºC, in good agreement with gel mobility data obtained for the sequences (*122*).

3.1.7 Cytosine Bulges

Morden *et al*. (*152*) determined the conformation of the oligonucleotide duplex $d(CA_3{}^{C}A_3G)\cdot(CT_6G)$, which contains an extra, unpaired cytosine residue. The conformation was investigated by using NMR methods to observe both the aromatic base protons and the base-pairing imino protons. The presence of NOE interactions between imino protons on either side of the potential intercalation site for the dC residue indicates that this extra base flips out into solution. If the cytosine were stacked into the helix, the distance

between base pairs 4 and 5 would be almost 7 Å, and no such NOEs would have been observed. The extrahelical cytosine was shown to cause a decrease in the duplex melting temperature of 15 ºC.

The self-complementary duplex d(CCG^{C}GAATTCCGG)$_2$ has an extra cytidine on each strand in its duplex form, and was studied by Kalink *et al*. (*109*). NOE and chemical shift parameters establishes that the predominant conformation at 273 K has the extra cytidine looped out of the helix. Specifically, the H5 and H6 proton chemical shifts correspond closely to the high temperature (343 K) unstacked strand values. The conformational equilibrium shifted towards a structure with a stacked in extrahelical cytidine on raising the temperature to 313 K prior to helix-coil melting transition, as evidenced by the marked upfield shift in H5 and H6 proton resonances.

3.1.8 Thymine Bulges

Studies on the preferred conformation of extrahelical bases have also been carried out on oligomers with unpaired thymidines. Kalnik *et al*. (*107*) used two 13-mers with extra dT residues, but otherwise self-complemenatry, to observe the effect of flanking bases and temperature on thymidine bulges. For the oligonucleotide d(CCG^{T}GAATTCCGG)$_2$, the thymidine methyl protons had a chemical shift very close to its strand value at 278 K. This resonance shifted upfield by 0.2 ppm upon raising the temperature to 308 K. This demonstrates that when the extra thymidine is flanked by guanosines, the non-pairing base is in a conformational equilibrium between looped-out and stacked states. The looped-out state is favoured at lower temperatures, but the equilibrium shifts to the stacked form at elevated temperatures. However, when the thymidine is flanked by cytosines, as was the case for the 13-mer d(CCGGAATTC^{T}CGG)$_2$, it remains looped-out regardless of temperature in the duplex state. This was demonstrated by the observation of a temperature independent chemical shift of the methyl protons between 278K and 338K, at

which temperature the oligomer has exceeded its T_m and is essentially in the strand state.

3.1.9 Relative Stabilities of Bulged Bases

Wartell and co-workers (*112;257*) have studied DNA and RNA fragments containing single base bulges with a view to assessing how their thermal stability is affected by the nature of an included extrahelical base and the identity of its flanking bases. The stabilities of 48 RNA oligomers with bulged bases at one of six positions, and 32 DNA oligomers differing by a single unpaired base were determined by temperature gradient gel electrophoresis (TGGE), using a temperature gradient of 36-41 ºC. Once loaded, the fragments moved into the gel until they reached a temperature at which their first melting domain unwound. This unwinding drastically reduced their mobility, so the further an olionucleotide migrated, the greater its stability. In both DNA and RNA, purine bulges were shown to be generally less destabilising than pyrimidine bulges. Additionally, sequences containing at least one neighbouring base identical to itself were more stable than those where the extrahelical base is different than its nearest neighbours. For B-form DNA in solution, it can be said that extrahelical purines are generally stacked into the helix, while extrahelical pyrimidines are in equilibrium between stacked and looped-out states.

3.1.10 Complexes with Bulged DNA

Bulged or looped out nucleotides have been proposed to be intermediates in frame-shift mutagenesis, arising from the movement of one strand relative to the other in homopolymeric DNA sequences (*218*). It is possible that drug intercalation would increase the lifetime of the mismatched state, increasing the probability of a base addition or deletion. As a result, drug interactions at nucleotide bulges have been intensively studied.

Nelson and Tinoco (*156*) demonstrated, by competition dialysis, that ethidium bromide has a higher affinity for an oligonucleotide containing an extrahelical cytidine at the intercalation site compared to a normal duplex of the same sequence. The drug binds approximately an order of magnitude more tightly to a bulged site than to a regular duplex binding site.

Woodson and Crothers (*246*) examined the binding of the intercalator 9-aminoacridine (9-AA) (Fig.3.4) to guanosine bulges by studying a series of small synthetic oligomers containing bulged bases at fixed and unfixed sites. 9-AA has been shown to bind to DNA at GpC, and to a lesser extent CpG, steps (*252*). Initially pairs of oligomers with perfect Watson-Crick base pairing and with guanosine bulged at fixed sites, such as d(GATGGGCAG)·(CTGCCCATC) and its bulge containing sister d(GATGGGCAG)·(CTGC^{G}CCATC), were studied and characterised. Competition dialysis between regular and bulged helices showed that, like ethidium, 9-AA has a higher affinity for bulged sequences than for normal oligomers of similar sequence and length. Subsequently, similar oligonucleotides differing only by the presence or absence of bulged bases in their homopolymeric tracts, like d(GATGG^{G}GCAG)·(CTGCCCATC) and d(GATGGGCAG)·(CTGCC^{C}CATC), were studied by NMR methods. The NOESY spectra of these sequences proved impossible to fully assign, presumably due to conformation averaging in these regions. It has been suggested that there is no preferred binding site for the drug in these sequences, and that the drug most probably intercalates at a number of positions along the GC run. Again, however, competition dialysis confirms a higher affinity for the bulged sequences. Although the data presented were limited, these experiments confirmed that intercalating drugs will specifically bind and stabilise bulge-containing helices.

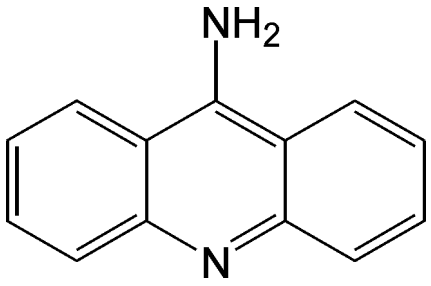


Fig.3.4: The DNA intercalating drug, 9-aminoacridine.

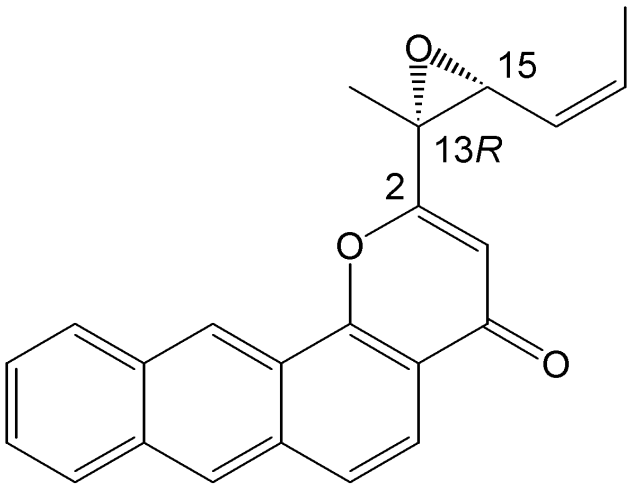


Fig.3.5: Chemical Structure of 13*R*-1.

Nakatanai and coworkers (*154*) designed and synthesized the chiral molecule 13*R*-1 (Fig.3.5) as an aglycone model for the pluaramycin antibiotic, altromycin B. This molecule has been shown to specifically alkylate guanine opposite a thymidine bulge flanked by cytidines. The chemical modification occurred primarily near the bulge as a result of the much more efficient formation of an intercalated complex at the bulged site than at an intrinsic binding site for the drug. The drug specifically alkylates guanine opposite a dT bulge, not only because of the increased stability, but because of the formation of a uniquely oriented complex.

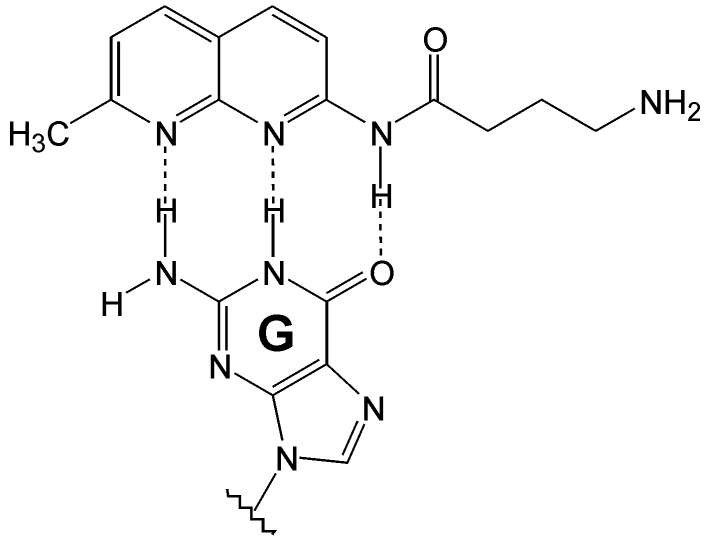


Fig.3.6: 2-acylamino-1,8-naphthyridine bound to guanine.

Other work carried out by the Nakatani group in Kyoto demonstrated that 2-acylamino-1,8-naphthyridine (Fig.3.6), which possesses H-bonding groups fully complementary to guanine, selectively binds to a single G-bulge of duplex DNA (*155*). Upon additon of equimolar drug, the melting temperature of the G-bulged duplex oligonucleotide d(GTTGC^{G}CTGGA)·d(TCCAGGCAAC) was increased by 5 ºC, but no increase in T_m was observed for similar A- and T-bulge sites or for the fully complementary duplex. DNase I footprinting studies indicated a selective binding of the drug to the bulge with an association constant of 3.4×10^{-4} M^{-1}.

As we have previously discussed, bulged dG and dA bases tend to remain intrahelical and cause a kink in the helix, whereas the behaviour of bulged dC or dT is more variable; either intrahelical or extrahelical depending on flanking bases and temperature. Caceres-Cortes and Wang (*27*) have studied the 2:1 complexes of nogalamycin (Ng) and the dT bulge containing, self-complementary heptamers d(C^{T}GTACG)$_2$ and d(CGTAC^{T}G)$_2$ to assess what happens to the extrahelical dT upon binding of Ng. When nogalamycin is mixed with the bulged DNA duplex in a 1:1 ratio, the spectrum of each complex showed resonances from the free DNA, 1:1 drug:DNA complex, and

2:1 complex, indicating that the binding equilibrium was slow on the NMR timescale (since resonances from all three species co-exist). The perfectly self-complementary form of these two oligomers, $d(CGTACG)_2$, had previously been titrated with Ng and the complex characterized by Robinson et. al. (*182*).

NOESY spectra were collected of all four systems; $d(C^TGTACG)_2$, $d(CGTAC^TG)_2$, Ng-$d(C^TGTACG)_2$ and Ng-$d(CGTAC^TG)_2$. Using a standard sequential assignment procedure (*182*) all resonances, except those of the deoxyribose H5' and H5" protons, were assigned. In each case it was expected that the extra dT residue at the CpG intercalation site would loop out into solution (Fig.3.8 (b);(d)), forming a complex almost identical to Robinson's. However, this was not the case. In both cases the DNA preferred to maintain an uninterrupted backbone at the intercalation site, leading to the formation of modified base pairs and causing the terminal bases to overhang (Fig.3.8 (c);(e)). In the Ng-$d(C^TGTACG)_2$ complex the 5'-C nucleotide was shown to overhang into solution while the drug intercalated at the TpG site formed from the T2 and G3 bases. The aglycone system is stacked with a G-T wobble base pair (Fig.3.7). Intercalation occurred for the Ng-$d(CGTAC^TG)_2$ complex at the 5' terminal CpG site, and in this case the overhang occurred at the 3' end of the heptamer. In this instance, a pyrimidine-pyrimidine C-T base pair was formed. This is significantly less stable than a normal Watson-Crick base pair, or even a wobble base pair.

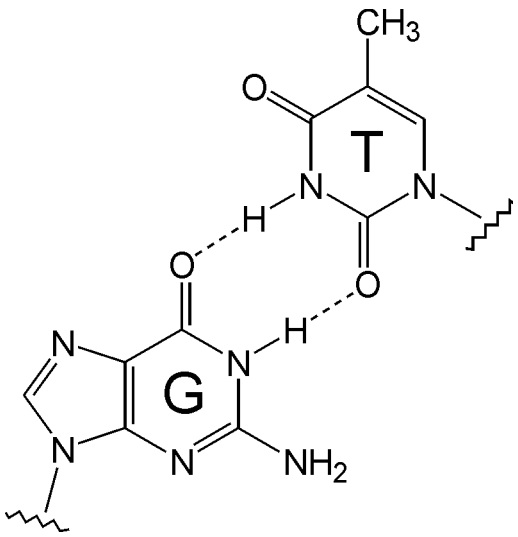


Fig.3.7: G-T wobble base pair.

Binding of nogalamycin may induce the extrahelical base to form a base pair over the aglycon ring, and shift the bulged distortion away from its original location. It is therefore important to consider the possible long-range effect of a conformational change propagated away from the original bulged site when considering the biological function of intercalating anti-tumour drugs.

Most non-covalent, intercalating drugs interact with DNA by inserting their planar chromophores between base steps to form elongated and unwound duplex structures without disrupting the flanking base pairs. Actinomycin D (ActD) (Fig.3.9) is one of the older chemotherapeutic drugs, which has been used in therapy for many years. It inhibits DNA–primed RNA polymerase by complexing with DNA via deoxyguanosine residues. The drug's phenoxazone ring intercalates at a 5'-GC step and the two cyclic pentapeptide lactones anchor the chromophore in the minor groove by lying in both 5'- and 3'-directions along that groove. Because of it's sequence specificity and potency, the ActD-DNA complex has been well characterised (*36;38;110*).

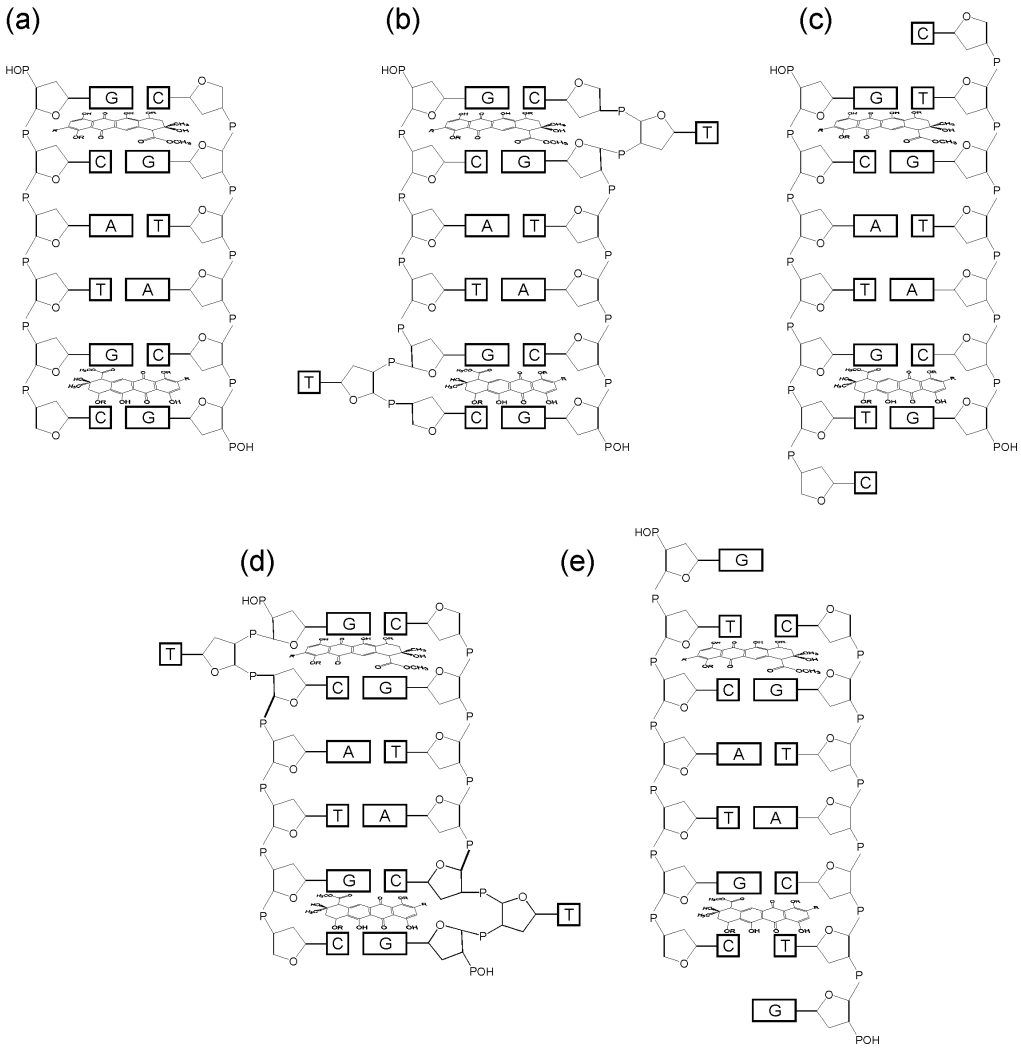


Fig.3.8: Schematic representations of nogalamycin complexed with the DNA oligonucleotides d(CGTACG) (a), d(CTGTACG) (b;c) and d(CGTACTG) (d;e).

Chou and co-workers (*38*) have reported the structures of ActD binding with previously uncharacterised DNA recognition sites 5'-G*X*C/C*Y*G-5', where *X/Y* is an A-T or C-G base pair, in hairpins closed by a mini ACC-loop (*39*). Once the sequence d(GTCACCGAC) was mixed with an equivalent of ActD, the central T-A pair was disrupted, and the bases looped out either side of the drug chromophore. The flanking G-C pairs were both highly buckled due to the extrusion of the unpaired T/A residues. Furthermore, the looped-out bases were shown to be perpendicular to the long axis of the flanking base pairs. These residues were not disordered, but rather involved in some tertiary interactions. As well as a strong H-bond detected between the O4' of the extra thymidine and the quinoid amino group (Fig.3.9), another H-bond was also

formed between the looped-out adenosine amino group and the backbone phosphate oxygen on the adjacent guanosine nucleotide.

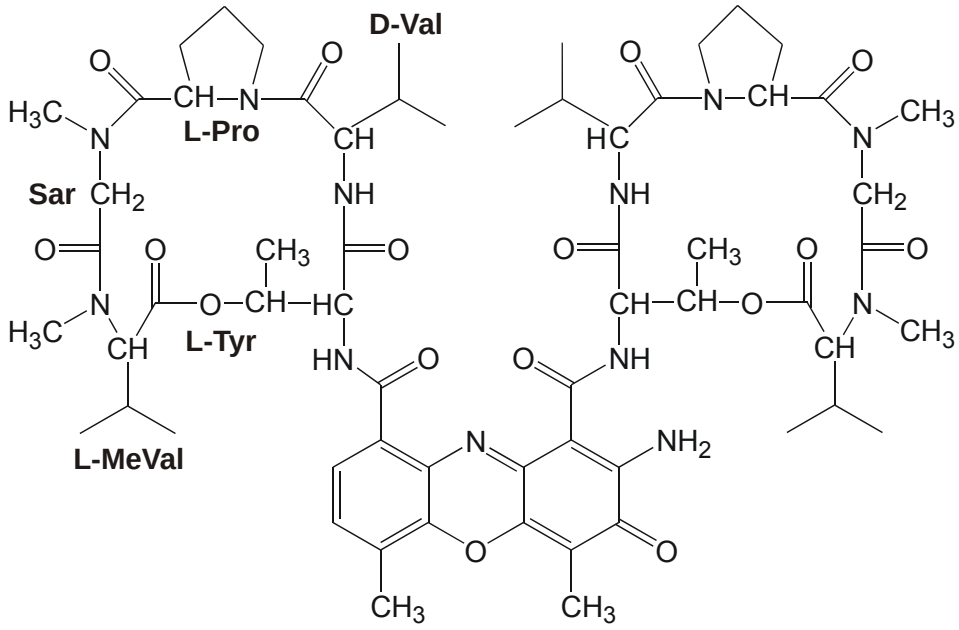


Fig.3.9: Actinomycin D.

The ActD-d(GCCACCGGC) complex structure was shown to be very similar to the T/A structure, with the amino group of the looped-out guanosine residue forming a hydrogen bond with the adjacent guanosine backbone oxygen, and the looped dG forming a perpendicular interaction with the ActD chromophore (*38*). UV melting studies showed that ActD binding caused a considerable stabilisation of 12 ºC for the control DNA hairpin d(GCACCGC), but an even greater stabilisation of more than 20 ºC was observed for complexes containing 5'-G*X*C sites.

3.1.11 Biological Responses to DNA Bulges

The ability of many species to repair mismatches in double-stranded DNA has been well documented. The first critical step in this process is the recognition of the mismatched DNA. In the major mismatch repair pathway in *Escherichia coli*, this is accomplished by the *Mut*S protein. The *Mut*S protein binds to the site of a mismatch in double-stranded DNA and, with the cooperation of the *Mut*L and *Mut*H proteins, targets a section of one of the DNA strands at that

location for removal. Other proteins complete the repair process: the section of DNA that has been targeted is removed and degraded, a patch is synthesized using the complementary strand as a template, and the patch is ligated into place resulting in a repaired section of double-stranded DNA without mismatches. The comparison between eukaryotes and prokaryotes is complicated by the fact that eukaryotes encode multiple *Mut*S homologs. For example, in yeast there are six *Mut*S homologs, MSH 1-6 (*24*). In yeast, as in humans, MSH2, MSH3, and MSH6 are involved in the repair of mismatches in nuclear DNA (*221;225*). MSH2 is required for all such repair and MSH3 and MSH6 provide specificity for the type of mismatch: MSH6 for base:base mismatches and small loops and MSH3 for large loops.

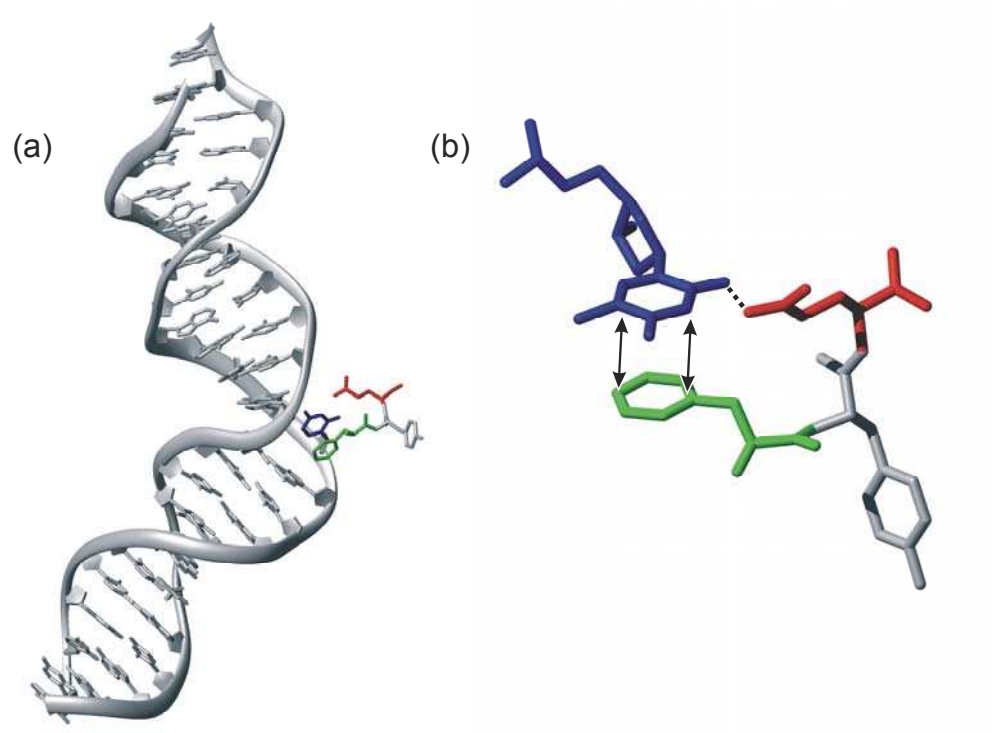


Fig.3.10: The Phe-X-Glu region of *E.coli Mut*S protein bound to a G-T mismatch pair (a). Phe^{36} (green) takes part in a ring stacking interaction, while the amino proton of Glu^{38} (red) is involved in a H-bond with the *O2* of the mispaired thymine base (blue).

Recent crystal structures of the *Mut*S protein from *Thermus aquaticus* and *Escherichia coli* complexed with bulge containing DNA (*126;157*) have revealed that the Glu^{38} residue in a conserved Phe-X-Glu motif participates in a H-bond with either an unpaired thymidine or the thymidine of a mismatched

G-T base pair (Fig.3.10 (b)), while Phe^{36} (*E. coli*) or Phe^{39} (*Taq*) is involved in a stacking interaction. The role of hydrogen bonding in mismatch recognition has been assessed by Schofield *et al*. (*189*). The relative affinities of *Mut*S for DNA duplexes containing non-polar shape mimics of adenine and thymine, 4-methylbenzimidazole (4MB) and difluorotluene (DFT) (Fig.3.11), were determined in gel mobility shift assays (*190*).

When both adenine and thymine bases in an A-T base pair were replaced with non-polar mimics, both *E. coli* and *Taq Mut*S were shown to exhibit a preference for duplexes containing 4MB and DFT compared to those containing normal A-T base pairs. Overall, a 2- to 3.5-fold increase in binding affinity was observed. This is consistent with an induced fit mode of mismatch binding, in which destabilised duplexes are preferred substrates for *Mut*S.

The affinity of *E. coli Mut*S for duplexes with single base bulges was also measured. The single unpaired base was either A or T or, alternatively, 4MB or DFT. A reduced affinity for the DFT substrate compared to dT, consistent with the loss of a H-bond between Glu^{38} and the unpaired or mispaired base/analogue, was recorded. Only a 2-fold difference in relative affinities was observed between adenine and its mimic, indicating that the relative contribution of hydrogen bond formation was not as significant for dA as dT. Qualitatively similar results were also obtained with *Taq Mut*S.

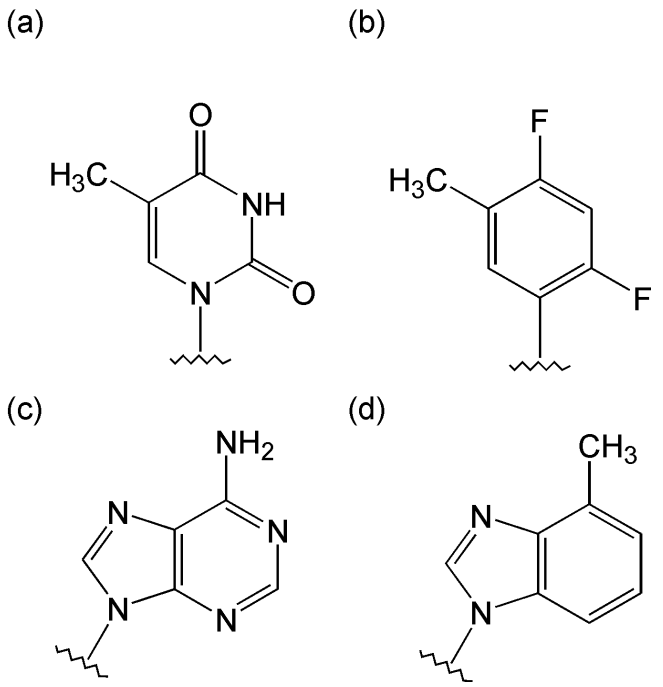


Fig.3.11: Thymine (a) and adenine (c) bases and their mimics, difluorotoluene (b) and 4-methylbenzimidazole (d).

These data suggest that two modest opposing forces are at work; one being the destabilisation of DNA duplexes by 4MB and DFT, that would favour binding by *Mut*S, and opposing this the loss of a H-bond that would destabilise mismatch binding.

The Klenow fragment of DNA polymerase I is one of the most structurally well defined members of the DNA polymerase family. Proofreading by the Klenow fragment is accomplished in a 3'-5' exonuclease site that is separated from the polymerase site by 25-30 Å (*66;159*). Lam *et al*. (*125*) demonstrated that the presence of an extrahelical base near the 3' terminus of the template strand can significantly increase the occupancy of the 3'-5' exonuclease site, suggesting that a DNA substrate will be transferred from the polymerase site to the exonuclease site following a misalignment event. Time resolved fluorescence spectroscopy was used to quantify the partitioning of a series of dimethylaminonaphthalene-1-sulfonyl (dansyl) labeled 17*/28-mer primer-template oligonucleotides between the two sites. The abundance of each of these distinct binding modes of the DNA-protein complex was estimated

under solution conditions by analyzing the decay of the dansyl probe (*86*). The inclusion of a bulged base at the first position from the primer terminus increased this partitioning, in favour of the exonuclease site, ranging from 3- to 7-fold, relative to the perfectly base paired primer, depending on the identity of the extrahelical base (dG>dA≈dT>dC). The largest effects were observed for an extrahelical base at the third or fourth position from the primer terminus, where partitioning increased by 8- to 18-fold.

3.2 Experimental Design

3.2.1 Project Overview

Here, we examine by NMR a series of DNA hairpin sequences containing 5'-TpG (5'-CpA) nogalamycin binding sites incorporating an extra bulged thymine in both the Watson and Crick strands (5'-C^{T}A and 5'-T^{T}G) in non-terminal sites. This work follows on from preliminary NMR studies of bulge-containing DNA hairpin loop sequences by Colgrave (*40*).

This study endeavours to examine in structural detail how specificity for binding to a G-C base pair influences the conformation and dynamics of the bases at non-standard interacalation sites. The extra thymine base can potentially be accommodated through a drug-induced frameshift with base mis-pairing occuring downstream of the drug-binding site, or through local stabilisation of an extrahelical T-bulge preserving the Watson-Crick base pairing of the duplex regions.

The major aim of this investigation is to show that a key feature of nogalamycin binding, namely specific G-C basepair recognition, can induce a number of novel structural features when an extrahelical base is inserted into the sequence, and to provide structural insights into the stabilisation of these non-standard DNA conformations.

3.2.3 Sequence Design

Previous studies have sought to determine whether an unpaired, extrahelical nucleotide in a normal B-form DNA duplex will loop out into solution, or stack into the helix, and how this is affected by drug binding. The question of whether or not this may give rise to a frameshift mutation has only been briefly, and accidentally, addressed by Caceres-Cortes and Wang (*27*).

To begin to address this issue, a family of novel DNA hairpins containing a G*N*A loop sequence (Section **3.1.1**), which has previously been reported to show remarkable thermal stability in short mini-hairpins, was designed. Although the GAA sequence shows the highest stability, it gives the poorest NMR visibility in terms of chemical shift dispersion, due to the fact that it comprises three purine bases, the protons of which come into resonance over a relatively narrow range. In this context, the GTA loop sequence is most amenable to NMR analysis. Both of these loop sequences have been employed in this study. These loops were included in these sequences in order to regulate the direction of any possible frameshift mutation. For example, the sequence d(CCCCC)·d(GGTGG), which contains a mismatched C-T pair, can potentially adopt two conformations in which the dT is looped out, depending on the direction, 5'-3' or 3'-5', of the frameshift (Fig.3.12(a);(b)). However, if one end of this duplex is capped with a hairpin loop, giving rise to the sequence d(CCCCCG*N*AGGTGG), only one conformation, corresponding to the 3'-5' frameshift, is now possible (Fig.3.12(c)). Based on the more stable GAA mini-hairpin sequence, two hairpins with 5'-TG (5'-CA) drug intercalation sites in the complementary double-stranded stem region were synthesised. The binding sites of these sequences, d(GC^{T}ACGAAGTGC) and d(GTGCGAAGC^{T}AC), each possessed a bulged-thymine (5'-C^{T}A) in the 5'-CA sequence, in each of the two possible orientations with respect to the loop (Fig.3.13(a);(b)).

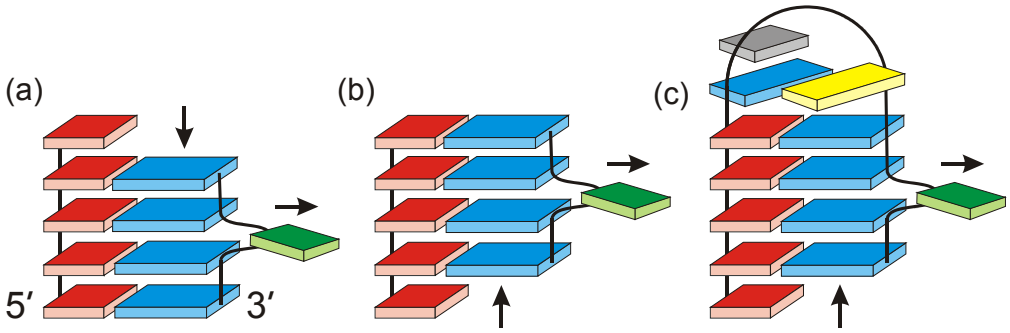


Fig.3.12: Possible conformations of the bulge-containing sequence d(CCCCC)·d(GGTGG), with (c) and without (a;b) d(G*N*A) mini-hairpin. Cytosines are shown in red, guanines in blue, and thymine in green.

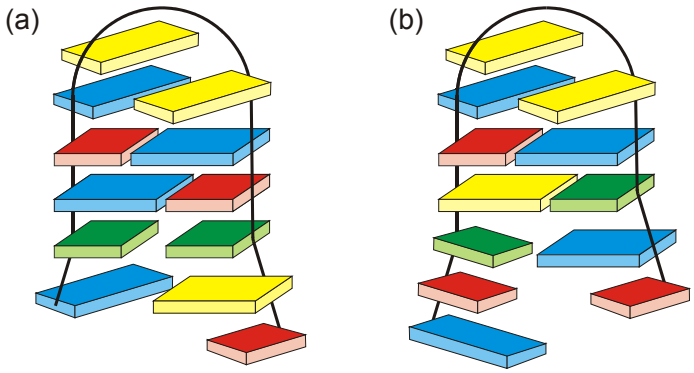


Fig.3.13: Structure of the oligonucleotide sequences d(GTGCGAAGC^{T}AC) (a), and d(GC^{T}ACGAAGTGC) (b), in their proposed folded conformation with frayed ends. Cytosines are shown in red, thymines in green, guanines in blue, and adenines in yellow.

As we have discussed in Section **2.1.2**, nogalamycin is an intercalating anthracycline antibiotic that binds with high affinity and sequence specificity to double stranded DNA through recognition of a 5'-pyr-pur site (CpG or TpG) (*64;65*). The bound orientation of the drug is dictated by the requirement for a guanine on the 3'-side of the intercalation site (CpG or TpG). While a palindromic 5'-CG site permits binding in either orientation, binding to the 5'-TG site results in a unique binding orientation dictated by the presence of a guanine base on only one strand (Section **2.1.2**). The sequences d(GC^{T}ACGAAGTGC), d(GTGCGAAGC^{T}AC) and d(CCCACGTAGT^{T}GG) have been designed to ascertain whether TpG sites, which incorporate an extra bulged thymine on either face, can still bind nogalamycin and, if so, what effect this has on the structure of the surrounding sequence.

3.3 Materials & Methods

3.3.1 Synthesis & Purification

The oligonucleotides were synthesised and purified as described in Section **2.3.1**, and were shown to be > 95% pure and in the hairpin form by ^{1}H NMR spectroscopy. The oligonucleotide sample was quantified spectrophotometrically. The NMR samples were prepared in the single strand concentration range of 4.8-7.6mM in 600 µl of mixed solvent, having a final salt concentration of 100 mM NaCl and 10 mM Na_2HPO_4, for WATERGATE experiments. The hairpins were lyophilised and redissolved in D_2O solution (600 µl) for the observation of non-exchangeable proton resonances.

3.3.2 NMR Analysis

NMR experiments were performed on either a Bruker DRX™500 or Bruker Avance™600 spectrometer, as described in Section **2.3.3**. WATERGATE NOESY spectra at 300 ms, 150 ms, 100 ms and 75 ms mixing times were acquired using samples dissolved in 90% H_2O and 10% D_2O. All TOCSY experiments employed a spin-locking field of 7 kHz. Two-dimensional data were zero-filled to 2048 x 1024 points prior to Fourier transformation, optimised with a shifted sine squared function in both dimensions, and treated with automatic baseline correction.

3.3.3 Distance Restraints

Interproton distances for nogalamycin-d(GC^{T}AC<u>GAA</u>GTGC) were derived from integration of NOE cross peak volumes from NOESY experiments (288 K, D_2O) with mixing times of 150 ms, 100 ms, and 75 ms. Distances were determined by extrapolation to zero mixing time as described in Section 2.3.4. A full description of the criteria for selecting upper and lower error bounds, and for comparing geometrical inconsistencies arising from the application of distance restraints can be found in Section **4.3.3**.

3.3.4 Starting Structures

NMR structure determination of DNA molecules has been shown to be more difficult than for proteins because the network of NOEs in nucleic acid spectra is less dense than in protein spectra (*249*). For this reason, structure calculations based on Cartesian space molecular dynamics often fail to find nucleic acid conformers that satisfy the experimental data (*216*) unless *ad hoc* assumptions about the 3D structure are made. In such instances, torsion angle dynamics have been shown to be successful in finding structures that satisfy the experimental restraints. The program DYANA (*87*) (DYnamics Algorithm for NMR Applications) for efficient calculation of three-dimensional protein and nucleic acid structures from distance constraints and torsion angle constraints collected by NMR experiments, performs simulated annealing by molecular dynamics in torsion angle space and uses a fast recursive algorithm to integrate the equations of motions. The principal benefit of a torsion angle dynamics (TAD) approach is that it works with internal rather than Cartesian coordinates: the covalent structure parameters (bond lengths, bond angles, chiralites & planarities) are kept fixed at their optimal values. The strong potentials required in traditional Cartesian MD to retain the covalent structure, and the high-frequency motions caused by such potentials are absent in torsion angle dynamics. Thus, torsion angle dynamics can be more efficient than molecular dynamics in Cartesian coordinate space because of the reduced number of degrees of freedom and the concomitant absence of high-frequency bond and angle vibrations, which allows for the use of longer time-steps and/or higher temperatures in the structure calculation. Test calculations (*87*) starting from conformers with random torsion angle values have shown that DYANA is capable of efficient calculation of high-quality protein structures with up to 400 amino acid residues, and of nucleic acid structures.

The torsion-angle molecular dynamics algorithm starts from an extended-strand conformation and proceeds in four stages: high-temperature torsion-

angle dynamics, slow-cooling torsion-angle dynamics, Cartesian molecular dynamics, and minimisation.

Candidate structures were generated from NMR-derived distance restraints using simulated annealing combined with torsion angle dynamics from 100 randomly generated structures using DYANA. The sequence d(GCTACGAAGTGC) was created by combining DYANA nucleic acid library files, in the appropriate order. Each file declares atom types, connectivities, dihedral angles and standard geometries. The covalent geometry of the nucleotides is based on the AMBER (*29*) force field.

Calculations with more than one molecule are made possible with DYANA through the use of special linker residues. These "invisible" linkers consist exclusively of pseudoatoms that have zero van der Waals interactions and thus can penetrate the real molecules without causing any steric repulsion, and so allow the program to formally treat a multimolecular system in the same way as a single molecule. Various linkers included in the DYANA library were used to attach the oligomer, d(GCTACGAAGTGC), to nogalamycin: An NL (nucleotide–generic linker) linker was used to link the 3'-dC nucleotide residue to a chain of four generic linkers, LL5, each with 5 Å bond lengths and 90° bond angles. Finally, a modified LN (generic linker–nucleotide) linker attached the terminal generic linker to the C10 position of the drug molecule. Each of these linker residue types in the standard library provides three rotatable bonds. A total of six linkers were used to separate drug from DNA, as a sufficient number of these linker residues must be used between two molecules such that no artificial constraint on the relative positioning of the two molecules with respect to each other is introduced by the finite length and flexibility of the stretch of linker residues. The linkers essentially create a flexible "chain" between two molecules. The drug molecule was modelled as a fixed unit with no free torsion angles, the internal co-ordinates of which were taken from a previous complex (*241*).

Upper bound distance restraints were derived from NOE data obtained at 75, 100 and 125 ms mixing time and corrected for spin diffusion by extrapolating distances to 0 ms mixing time as before (Section **2.3.4**). For all distances, upper error bounds of either 10% or 20 % were applied. The lower error value was applied by default, while the higher value was reserved for those peaks that did not appear on all three low mixing time NOESY spectra. In total 260 distance restraints were applied, of which 221 were intra-DNA and 39 drug-DNA restraints. As the drug was treated as a rigid body, intra-drug distances restrains were excluded from the DYANA calculations. Forty-five torsion angle restraints were applied based on observed $\Sigma J_{H1'}$ values from ^{1}H and DQF-COSY spectra at 288 K (Fig.3.14). Although the number of NOEs in the loop region is low, chemical shift information (Section **3.4.1**) indicates the formation of the GAA loop. Therefore, the loop was further defined by the addition of 40 distance restraints derived from previous G*N*A loop structures (*98;258*). Additionally, eleven H-bonding restraints were added for those residues that could be clearly identified as having imino protons involved in slow exchange processes. The sequence d(GCTACGAAGTGC) and associated restraints were allowed to undergo 10000 steps of TAD simulation, 3000 of which used the elevated temperature function. The resulting output structures were scored based on the target function and the lowest energy structures were selected for further simulation. Two structures were selected from a total of 20 on each of two independent dynamics runs, and two structures from a reduced annealing time simulation (600 high temperature steps, 6000 total steps), giving rise to a total of six structures. DYANA outputs conformations in Cartesian coordinate format, which can subsequently be read into AMBER for minimisation and traditional Cartesian rMD simulations.

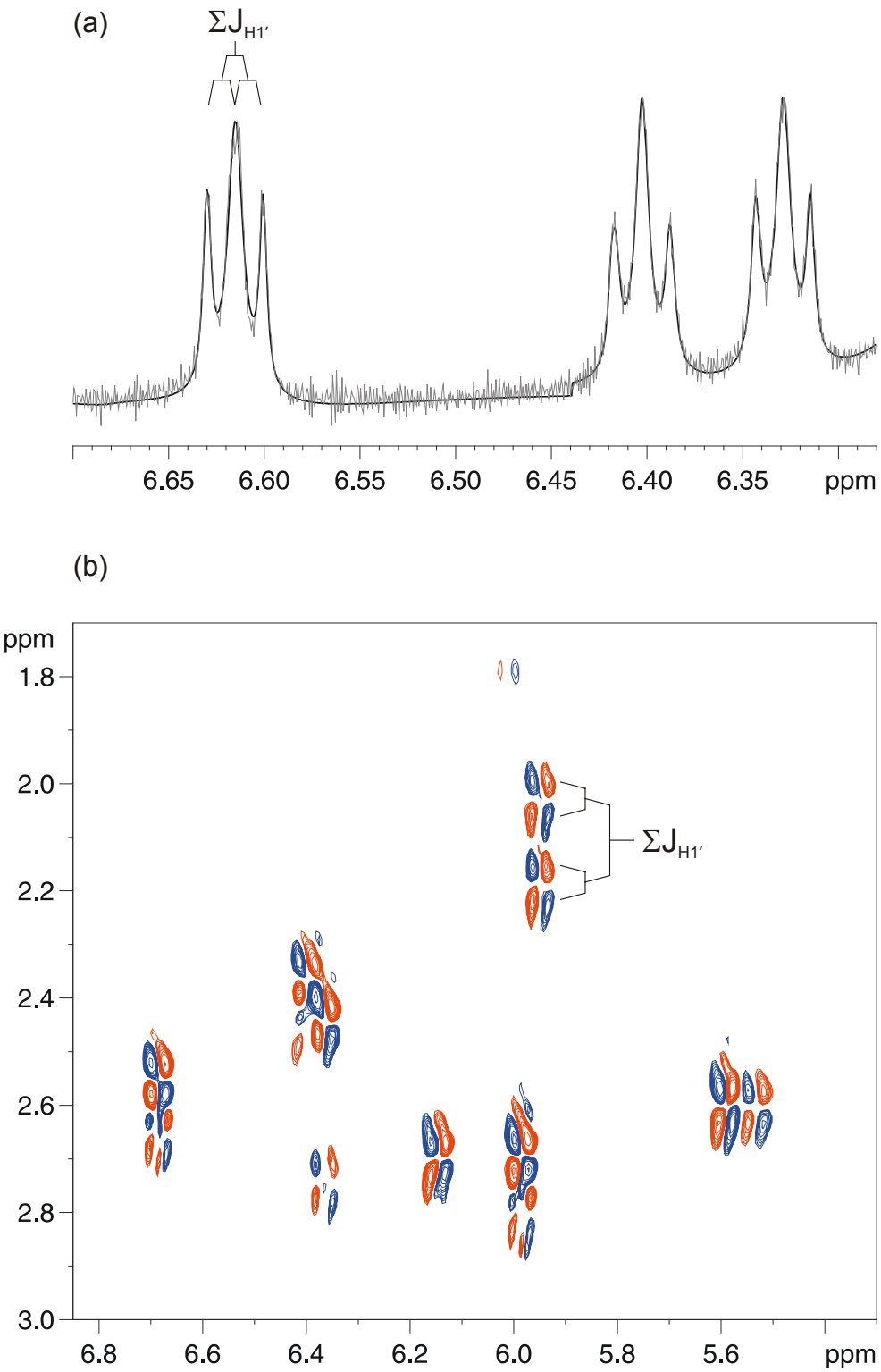


Fig.3.14: Sugar pucker restraints for nogalamycin-d(GC^{T}ACGAAGTGC) were fixed by reference to coupling constants obtained from deconvoluted ^{1}H NMR spectra broadened using a gaussian window function (a), or from crosspeaks on the DQF-COSY spectrum (b), each at 288 K.

Minimisation was followed by ten iterations of molecular dynamic simulation. Each of the first nine rounds of MD comprised 10 ps of simulation, split into 5000 steps of 0.002 ps duration. During the opening three iterations, the restraining field of 100 kcal $Å^{-2}$ mol was applied only to the DNA and counter-ions, while the temperature of the system was increased from 100K to 300K. At this temperature, the sodium ions were released from the restraining field and another round of simulation was run. Over the next five iterations the restraining field was reduced to 50, 20, 10, 5 and 2 kcal $Å^{-2}$ mol. The final round of MD was made up of 50000 steps, totalling 100 ps, during which the entire system was unrestrained. Following minimisation, all six structures were found to be superimposable with an RMSD of 1.15 (±0.1) Å.

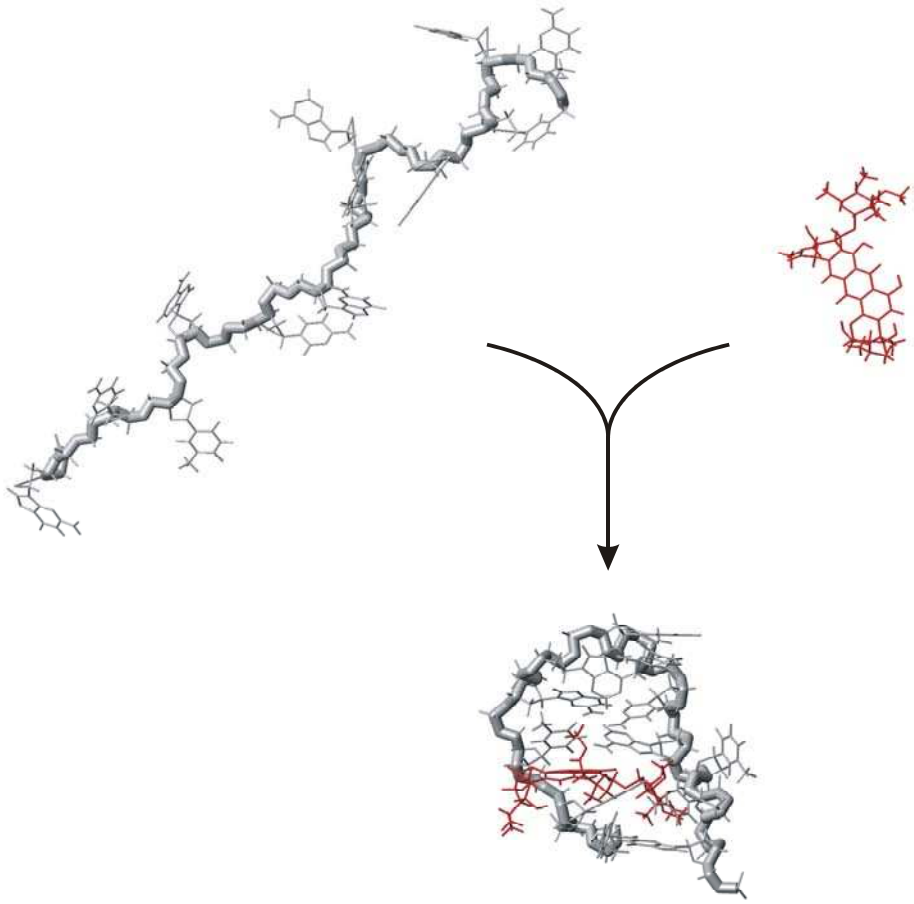


Fig.3.15: 3D representations of nogalamycin-d(GC^{T}ACGAAGTGC) before and after TAD simulation with DYANA.

3.3.5 Hairpin Structures

Following minimisation and unrestrained MD simulations, NMR restraints were introduced to the equilibrated structure and the system allowed to undergo 1 ns of restrained MD to satisfy the NMR restraints and search adequate conformational space under the influence of water molecules and cations. A total of 347 NOE-derived distance restraints were applied to nogalamcyin-d(GC^{T}ACGAAGTGC), and the system was subjected to rMD for 1 ns with all restraints active. 230 DNA-DNA distance restraints, 68 drug-drug and 39 drug-DNA restraints were applied together with 13 H-bonding restraints corresponding to the four Watson-Crick pairs and single G-A base pair. Pseudrotational phase angles (PRPAs) were fixed for each of the twelve DNA bases, based on relative intensities of their (H1'↔H2') and (H1'↔H2") NOE crosspeaks at 288 K. rMD was followed by energy minimisation. Snapshots were extracted throughout the rMD run at picosecond intervals. In the final averaged energy minimised structure, only one of the applied distance restraints were violated by more than 0.5 Å, while five deviated by more than 0.3 Å. All of the violated restraints corresponded to weak NOEs between nogalamycin and the terminal residues of the DNA stem region. The mean pairwise RMSD, calculated for all heavy atoms, over the final 100 acquired snapshots was 0.73 (±0.1) Å. This figure is reduced to 0.65 (±0.1) Å if only DNA atoms are considered, and further to 0.65 (±0.1) Å if the looped-out T3 residue is excluded from the calculation. The NMR restraints for nogalamycin-d(GC^{T}ACGAAGTGC) were satisfied well within the nanosecond of simulation.

3.4 Results & Discussion

3.4.1 Overview of NMR Analysis of Hairpin Loops

The investigations of these unusual loop-containing structures were undertaken by first characterising the free oligonucleotides followed by examination of the adducts formed upon reaction with nogalamycin. The NMR spectra of d(GC^{T}ACGAAGTGC) and d(GTGCGAAGC^{T}AC) were assigned based on observations from previous studies (*40*). The

conformations of uncomplexed d(GC^{T}ACGAAGTGC), d(GTGCGAAGC^{T}AC) and d(CCCACGTAGT^{T}GG) were investigated in the absence of drug using 1D and 2D NMR in H_2O and D_2O solutions (Section **2.4.2**).

It has been well documented that the H4' proton of the bridging nucleotide in a G*N*A loop undergoes a large upfield shift as a consequence of characteristic stacking interactions in the loop region (*98*). When a G*N*A loop is formed, the H4', H5' and H5" protons are subject to the ring current of the adenine in the sheared G-A base pair (Fig.3.16) and the H4' resonances are observed to be shifted upfield by ~2 ppm relative to the other H4' signals from nucleotides in the stem region (*98*). The H5' and H5" protons, although still affected by the adenine ring current, are not in such close proximity to the electrostatic surface of the ring, and are therefore shifted upfield to a lesser extent, typically 0.5-1.5 ppm.

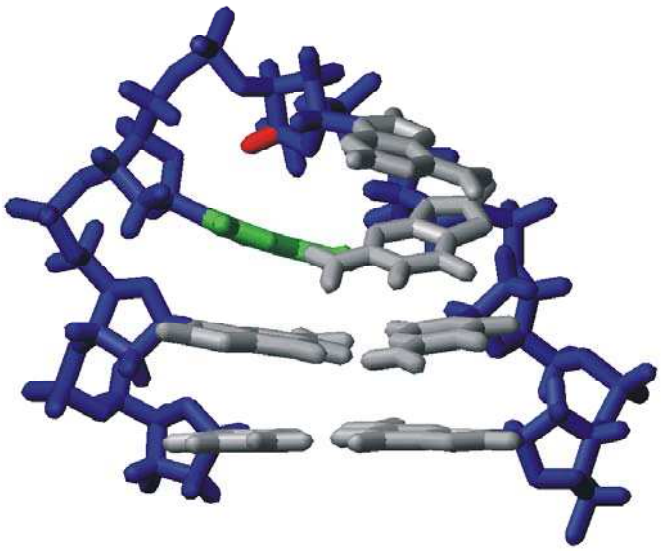


Fig.3.16: Stick model of the d(GCGAAGC) hairpin loop, viewed from the major groove and illustrating the proximity of the A6H4' (red) to the A5 ring system (green). Structure solved by Hirao *et al.* (*98*).

Stacking of the loop base occurs largely with the guanine of the G-A base pair. This is reflected in very weak sequential NOE connectivities between base H6/H8 and deoxyribose H1' and H2'/H2" protons for both

d(GC^{T}ACGAAGTGC) and d(GTGCGAAGC^{T}AC) across the ApA steps of their GAA loops (*41;240*). This stacking interaction is unaffected by the nature of the base, as evidenced by a similar weak interaction across the TpA step of d(CCCACGTAGT^{T}GG) (Fig.3.30) (Section **3.4.4.5**).

Further evidence for the formation of a sheared G(*anti*)-A(*anti*) conformation in these sequences is provided by the fact that the GH1'-GH8 and AH1'-AH8 crosspeaks are both weak in comparison with the reference NOEs observed between the H5 and H6 of cytosine, consistent with *anti* glycosidic torsion angles for the G and A bases at the mismatch site (Section **4.4.1**). Additionally, the G*N1*–H proton invariably appears as a broad peak at approximately 10 ppm, indicative of a non-H-bonded imino proton, such as would be involved in a sheared G-A base pair (Fig.3.2).

3.4.2 Drug Titrations

1D ^{1}H NMR titration studies of d(GTGCGAAGC^{T}AC), d(GC^{T}ACGAAGTGC) and d(CCCACGTAGT^{T}GG) show that nogalamycin intercalates with a 1:1 stoichiometry with the free and bound species at intermediate drug:DNA ratios in slow exchange on the chemical shift timescale.

3.4.3 NMR Analysis of d(GC^{T}ACGAAGTGC)

3.4.3.1 Loop Formation

Fig.3.17(a) illustrates the H4'-H2'/H2" region, expanded from the NOESY spectrum (D_2O, 288 K, 300 ms) (Fig.3.18) of d(GC^{T}ACGAAGTGC). The NOEs between the H2' & H2", H4' & H2', and H4' & H2" protons of the bridging adenine, A7, are indicated. The loops in this sequence, and in d(GTGCGAAGC^{T}AC) and d(CCCACGTAGT^{T}GG), were determined to be formed by the observation of the upfield shifted H4' and H5'/H5" sugar resonances and the pattern and intensity of NOEs across the G*N*A loop. In the NOESY spectrum of bulge sequence d(GC^{T}ACGAAGTGC) the A7 sugar H4',

H5' and H5" protons observed at 2.116, 2.241 and 2.361 ppm respectively. These were again >2 ppm upfield of the stem region A4 sugar H4', H5' and H5" protons, which came into resonance at 4.414, 4.024 and 4.122 ppm correspondingly.

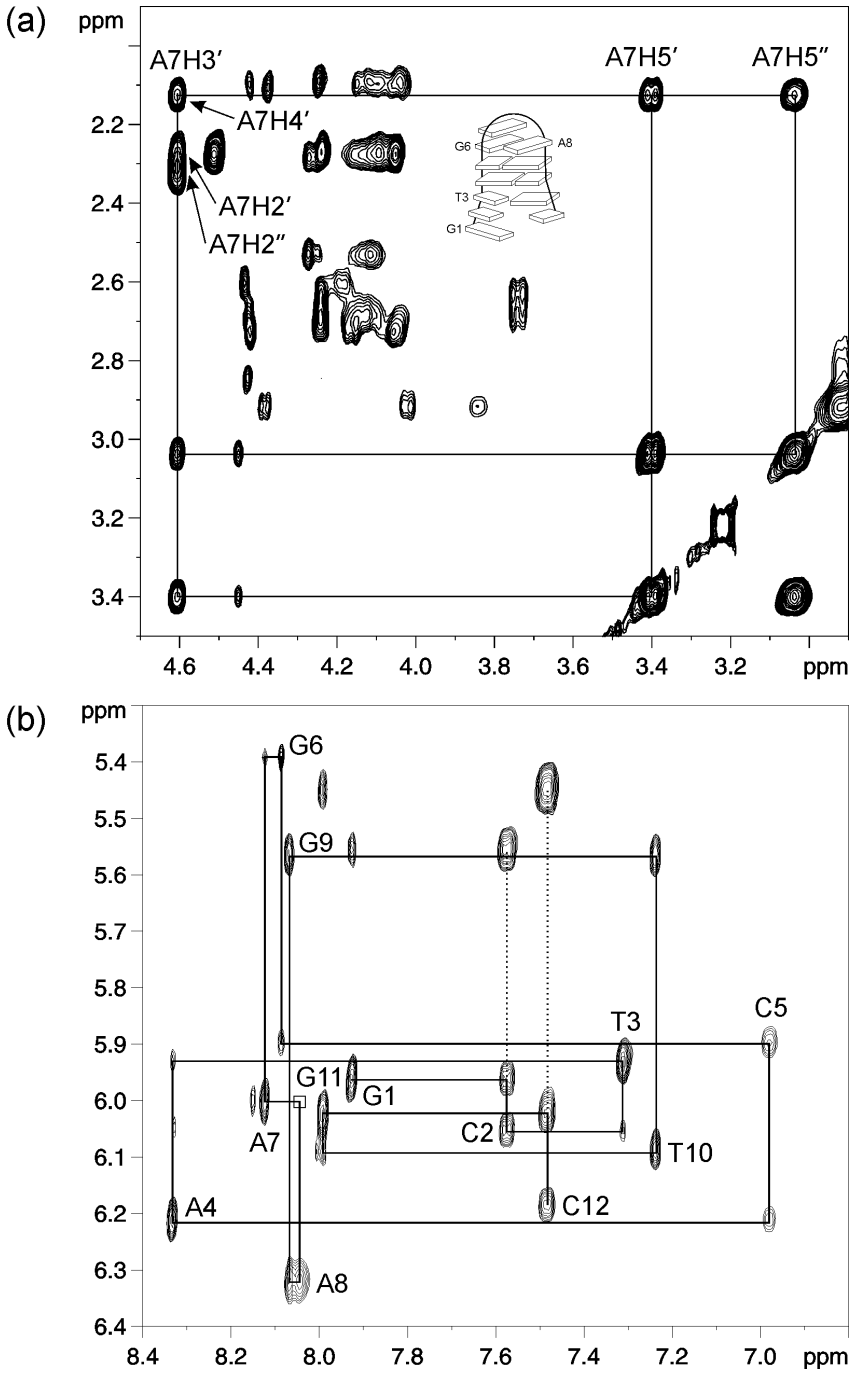


Fig.3.17: Expanded region of the 500 MHz NOESY spectrum of d(GC^{T}ACGAAGTGC) (288 K, D_2O, 300 ms) showing (a) the H4'-H2'/H2" interactions of the loop nucleotide, and (b) sequential H1'-H6/H8 connectivity, typical of B-like DNA, along the oligonucleotide chain.

Additionally, the C5H1'-G6H8 NOE crosspeak in the sequential H1'-H6/H8 connectivity pathway of d(GC^{T}ACGAAGTGC) (Fig.3.17(b)) is relatively weak. Weakness in, or often absence of, this interaction between the stem and loop regions has been shown to be another good indicator of loop formation (*41*).

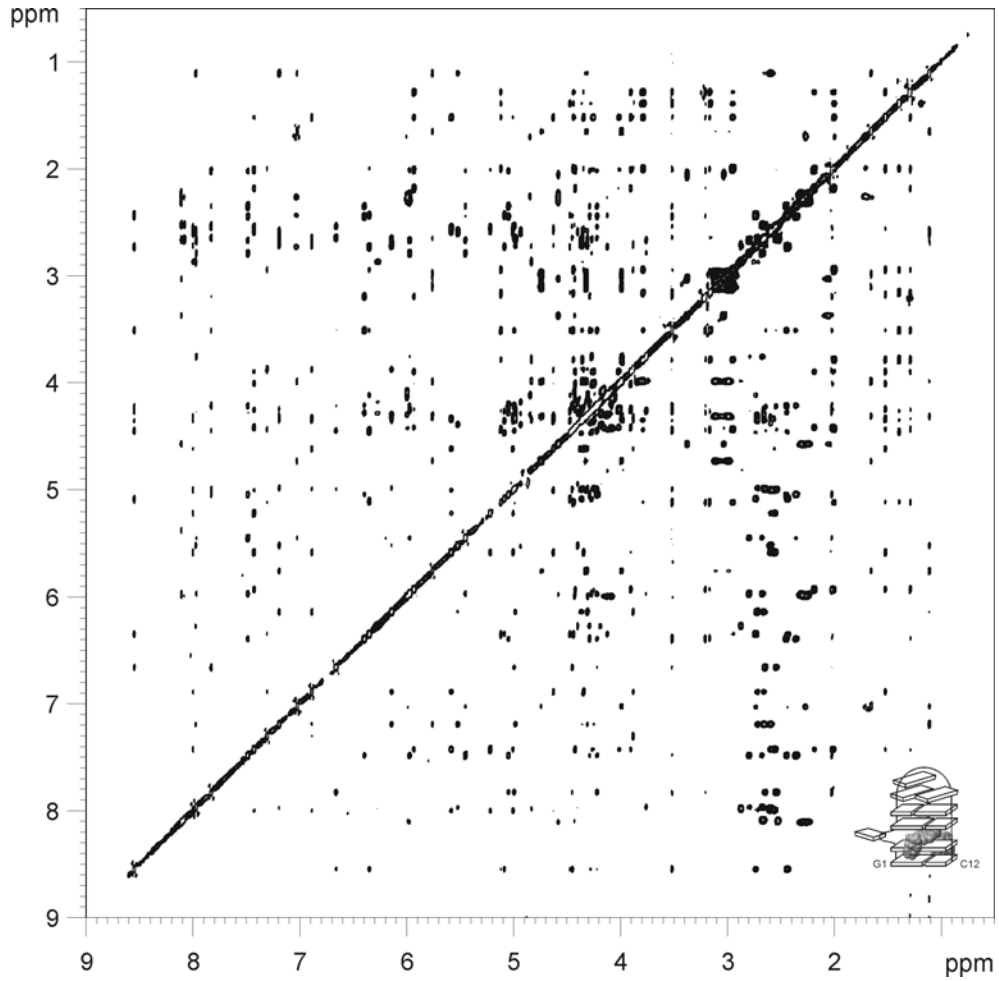


Fig.3.18: NOESY spectrum of nogalamycin-d(GC^{T}ACGAAGTGC) in D_2O solution, recorded at 288 K with a mixing time of 300 ms.

3.4.3.2 Imino Protons of Free DNA

For this sequence, only two imino resonances, at 12.80 and 13.85 ppm, which are indicative of stabilised Watson-Crick pairs, can be detected on the ^{1}H spectrum. (Fig.3.21(b)) These correspond to the G9 and T10 protons, with the latter (T10) showing significant exchange broadening. This demonstrates that while the stem region close to the loop is well formed, the ends of this hairpin,

and of d(GTGCGAAGC^{T}AC) (Section **3.4.4.1**) and d(CCCACGTAGT^{T}GG) (Section **3.4.5.2**) are destabilised by the introduction of the extra dT residue and, as a result, are poorly structured. Although, the absence of signals from the imino protons for the terminal residues of these hairpins is indicative of fraying at the stem ends, the oligonucleotides still appears to be in the *B*-like conformation. NOEs allowing the sequential H1'-H6/H8, and H2' & H2"-H6/H8, connectivity pathways to be traced along the DNA strand were observed, indicating internucleotide stacking consistent with helical *B*-form DNA. If the strands were adopting a random conformation, there would be no discernable pattern of backbone NOEs to follow, which is not the case (Fig.3.17). Many of the sequential NOEs are, however, quite weak suggesting some degree of conformational averaging.

3.4.3.3 Titration

Fig.3.20 shows the aromatic region of the 1D ^{1}H NMR spectra resulting from the addition of aliquots of nogalamycin to d(GC^{T}ACGAAGTGC). As with d(GTGCGAAGC^{T}AC), titration studies showed that nogalamycin formed a clean 1:1 complex with d(GC^{T}ACGAAGTGC), with resonances from the free DNA replaced by those of a complex containing a single-bound ligand. In this case, the most pronounced changes to chemical shifts upon ligand binding occurred at the intercalation site. In particular, the A4H8 proton was shifted downfield by 0.18 ppm to 8.55 ppm (Fig.3.20).

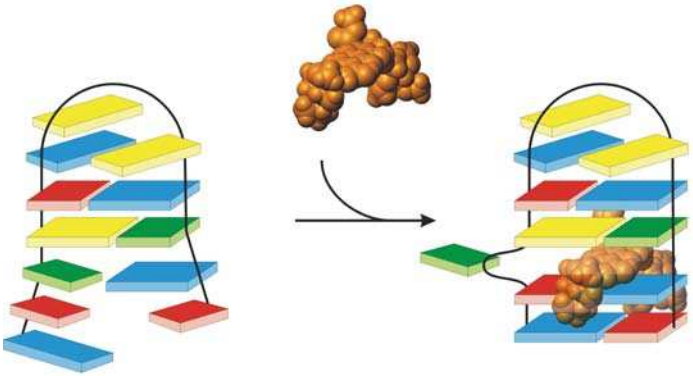


Fig.3.19: Schematic representation of the titration of d(GC^{T}ACGAAGTGC) with nogalamycin.

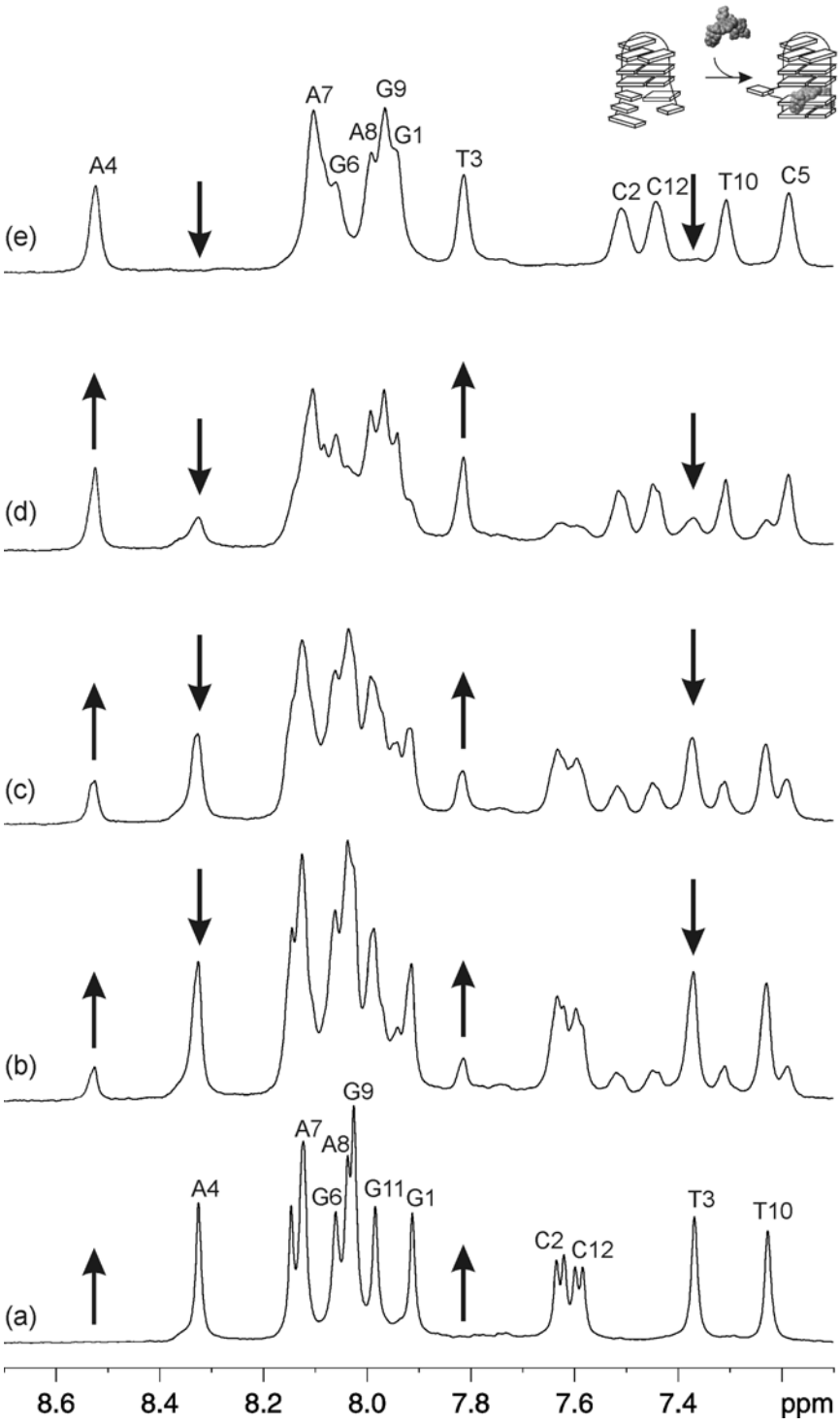


Fig.3.20: 1D ^{1}H NMR spectra of the titration of d(GC^{T}ACGAAGTGC) with nogalamycin ($D_2$0, 288K), highlighting the aromatic region (7.0-8.5 ppm), at drug:DNA ratios of (a) 0:1, (b) 1:6, (c) 1:3, (d) 3:4 and (e) 1:1. The downward facing arrows highlight clearly defined signals from the free DNA that decrease upon addition of the nogalamycin. Upward facing arrows indicate new signals that increase owing to complex formation.

Nogalamycin intercalation with d(GC^{T}ACGAAGTGC) produces a highly stable 1:1 complex. Six exchangeable resonances can be detected in the 1D spectrum of nogalamycin-d(GC^{T}ACGAAGTGC) peaks in the 12-15 ppm range. Two of these peaks correspond to the H4 and H6 phenolic protons of nogalamycin (Fig.3.21(a)) at 11.50 and 12.10 ppm respectively, while the remaining signals correspond to the DNA base imino protons. Four imino proton resonances can be identified on the ^{1}H spectrum (Fig.3.21(a)), which are indicative of the involvement of G1, G9, T10 and G11 in stable Watson-Crick base pairs. Although the G1 and T10 (13.30 ppm) protons are overlapped on the 1D spectrum, their mutual presence was verified by integration of the peaks in the imino region (11.0-14.0 ppm), and their identities confirmed by connectivites between imino protons and aromatic base protons on the 2D WATERGATE spectrum (288 K, 300 ms) of the complex, in the same manner as described for nogalamycin-d(GTGCGAAGC^{T}AC) (Fig.3.28).

3.4.3.4 Imino Protons of Drug-DNA Complex

The peak observed at 13.30 ppm relating to the T10 imino proton is shifted upfield compared with the T10 in the free oligonucleotide ($\Delta\delta$ = 0.65 ppm) (Fig.3.21(b)), consistent with a thymine at a 5'-TG intercalation site (*41*). The G1 imino proton also comes into resonance at ≈ 13.20 ppm. This peak, however, has a much greater linewidth in its complexed form, which can be attributed to the overlap of the G1 and T10 imino protons. The peak at 13.20 ppm, attributed to the G9 imino proton assigned based on imino-aromatic proton connectivities, is also shifted upfield ($\Delta\delta$ = 0.40 ppm). Using similar methodology, the peak observed at 12.71 ppm can be attributed to the G11 imino proton. These peaks were further resolved at elevated temperature (318 K) (data not shown). Only the terminal guanine base shows any evidence for exchange broadening at temperatures above 308K. The absence of a signal relating to the T3NH proton indicates that it is not involved in base-pairing,

and not protected from NH/ND exchange. It can, therefore, be deduced that the T3 nucleotide must be extrahelical.

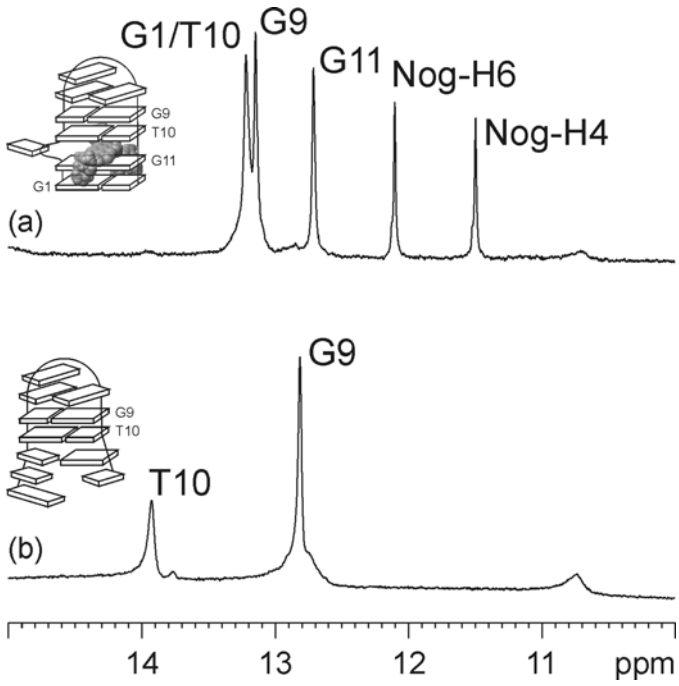


Fig.3.21: Imino-proton region of the 500 MHz ^{1}H NMR spectrum (288 K) of (a) nogalamycin- d(GC^{T}ACGAAGTGC)and (b) d(GC^{T}ACGAAGTGC).

A detailed analysis of chemical shift perturbations and NOE data is strongly indicative of the drug stabilising an extrahelical conformation of T3 in which the pyrimidine base is flipped-out of the helical stack into a more solvent-accessible environment. This allows complementary pairing downstream of the intercalation site. Chemical shift data show that both the base and sugar ^{1}H resonances of T3 are significantly downfield of those of the uncomplexed DNA ($\Delta\delta$ H1' = 0.75 ppm; $\Delta\delta$ CH_3 = 0.48 ppm; $\Delta\delta$ H6 = 0.53 ppm), which is indicative of diminished ring current effects that would only arise from a de-stacking of the thymine base. This is confirmed by the pattern and intensity of NOEs, which show very weak connectivities between C2H2'/H2" and T3H6, and an absence of the sequential NOE between C2H1' and T3H6 (Fig.3.23(b)). There are only very weak NOE crosspeaks evident between the protons on T3 and A4. These data are consistent with T3 not being accommodated within the helical stack but flipped-out into the major groove.

3.4.3.5 Drug-DNA Interactions

The orientation of T3 in the major groove is confirmed by the observation of weak NOEs from $T3CH_3$ to C2H6 and C2H3', but only very weak NOEs, attributable to spin-diffusion, from the extrahelical base protons to C2H1' and C2H2'/2" (Fig.3.22).

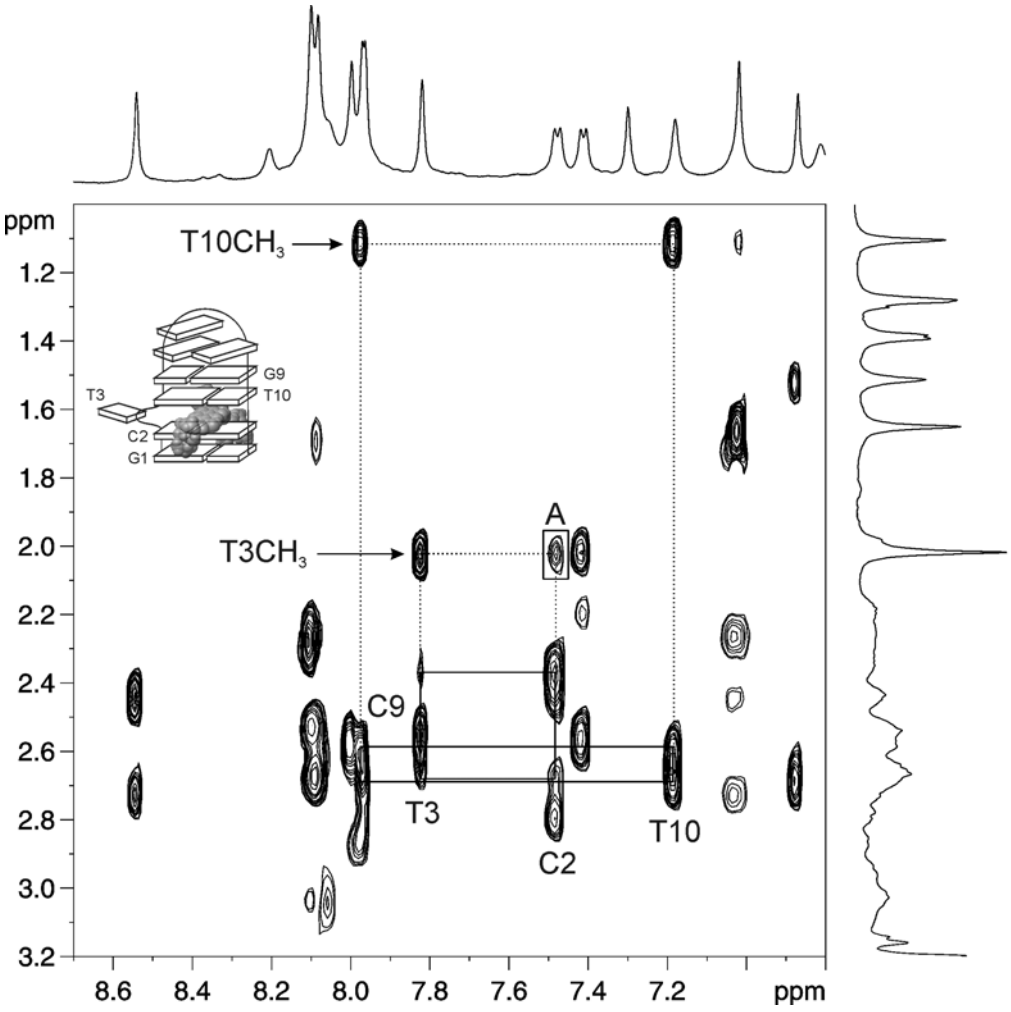


Fig.3.22: Expanded regions of the ^1H and NOESY spectra (90% H_2O, 10% D_2O, 288K, 300 ms) of nogalamycin-d(GC^{T}AC<u>GAA</u>GTGC), showing NOE connectivity pathways between aromatic and thymine methyl protons. The crosspeak labelled [A] corresponds to the $T3CH_3$-C2H6 interaction, and defines the orientation of T3 in the major groove.

A large number of drug-DNA NOEs that provides details of the drug orientation and stabilising interactions in the grooves can be observed on the NOESY spectrum of nogalamycin-d(GC^{T}AC<u>GAA</u>GTGC). NOEs from Nog-H11 to T10H1' and G11H1' (Fig.3.23(b)) unambiguously establish

intercalation at the 5'-TpG step, with NOEs from Nog-H1" and Nog-5"-CH_3 to T10CH_3 confirming that the bicycloaminoglucose is located in the major groove with the 2"-OH and 4"-OH groups positioned to hydrogen bond with the C2-G11 base pair. These interactions determine the orientation of the bound antibiotic. The bulky hydrophobic nogalose sugar is located in the minor groove with numerous NOEs from drug methyl and methoxy groups to deoxyribose H1' and H4' of the C2-G11 and G1-C12 base pairs demonstrating that a significant hydrophobic surface area is buried in the minor groove.

The H8/H6-H2'/H2" connectivity along the DNA backbone is conserved in the NOESY spectrum of nogalamycin-d(GC^{T}AC<u>GAA</u>GTGC), further indicating the location of the intercalation site (Fig.3.23(a)). The connectivity can be followed from G1 to C2 to T3, although the latter C2-T3 step is significantly weaker than the G1-C2 step. There is a break in the pathway between C2 and A4, which when combined with the weak C2-T3 interaction, favours a model in which the thymine is extrahelical. Connectivity through A4-G5-A6-A7 can also be observed. A loss of sequential connectivity is noted in the loop region between A7 and A8, again indicating that the loop A7 is stacked primarily on the G6 ring system. The connectivity along the 3'-strand can then be traced from A8 to G9, G9 to T10, from T10 to Nog-H11 and Nog-H11 to G11, and from G11 to C12. As was the case with nogalamycin-d(GTGC<u>GAA</u>GC^{T}AC), the connectivity along the DNA backbone can be followed with the drug bridging the gap between the bases at the intercalation site, namely T10 and G11. As for nogalamycin-d(GTGC<u>GAA</u>GC^{T}AC), the H8/H6-H1' sequential connectivity can be followed along the DNA backbone through the intercalation site, with the ligand bridging the gap between the T2 and G3 resonances. In this case, the C2-T3 NOE is absent, while the T3-A4 interaction is weak compared with others of the same type. These data clearly establish that the site of nogalamycin intercalation is in the expected TpG step, and suggests that the unpaired dT is not stacked into the stem helix. Furthermore, interactions between the non-exchangable base protons, T3H6 &

$T3CH_3$, and C2 protons in the major groove suggest that the extrahelical base is oriented in the major groove and cannot interact freely with the solvent. Strong crosspeaks can be detected on the NOESY spectrum (288 K, D_2O, 300 ms) corresponding to interactions between T3-CH3 and C2H6 & C2H3', while an albeit weaker interaction is evident between T3H6 and C2H2". Since it has already been unambiguously shown that the unapired thymine is not incorporated into the helix, these intractions can only be explained in terms of the extrahelical base resides in the major groove, and oriented towards the 5' end of the oligomer.

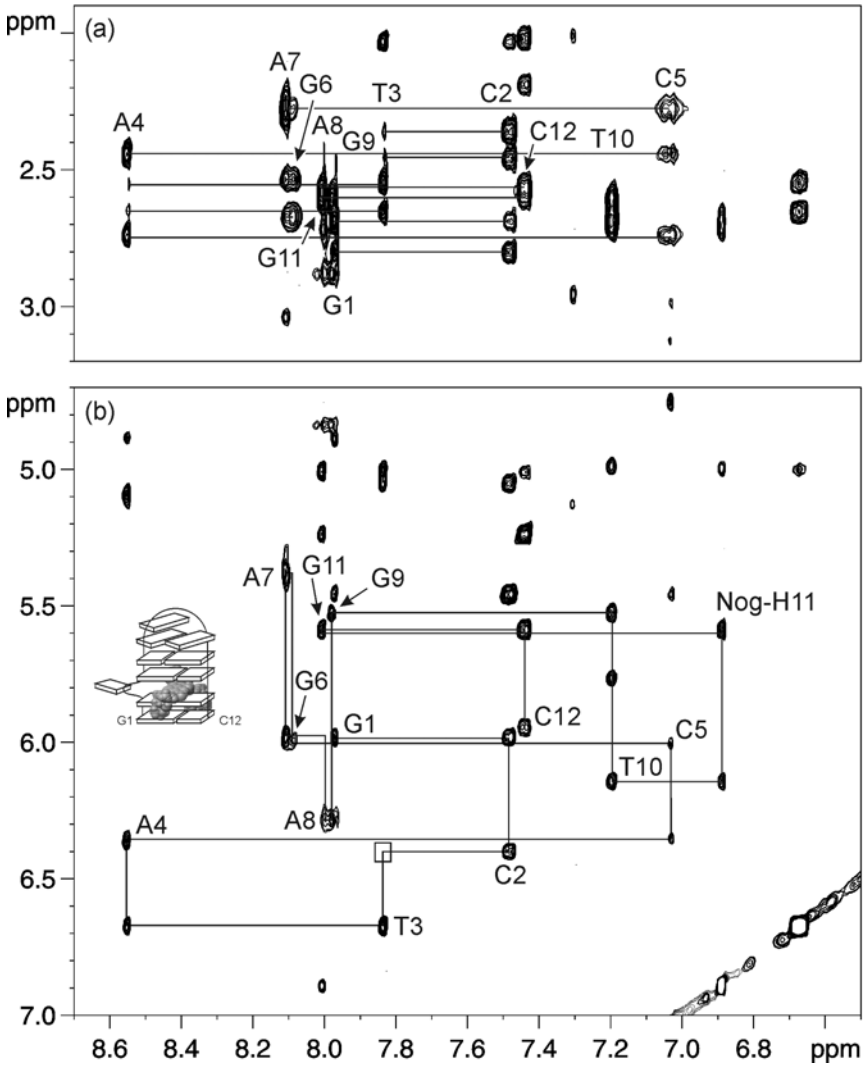


Fig.3.23: Expanded regions of the NOESY spectrum of nogalamycin-d(GC^{T}ACGAAGTGC), showing NOE connectivity pathways between (a) H6/H8 & H2'/H2" protons, and (b) H6/H8 & H1' protons.

3.4.4 NMR Analysis of d(GTGCGAAGC^{T}AC)

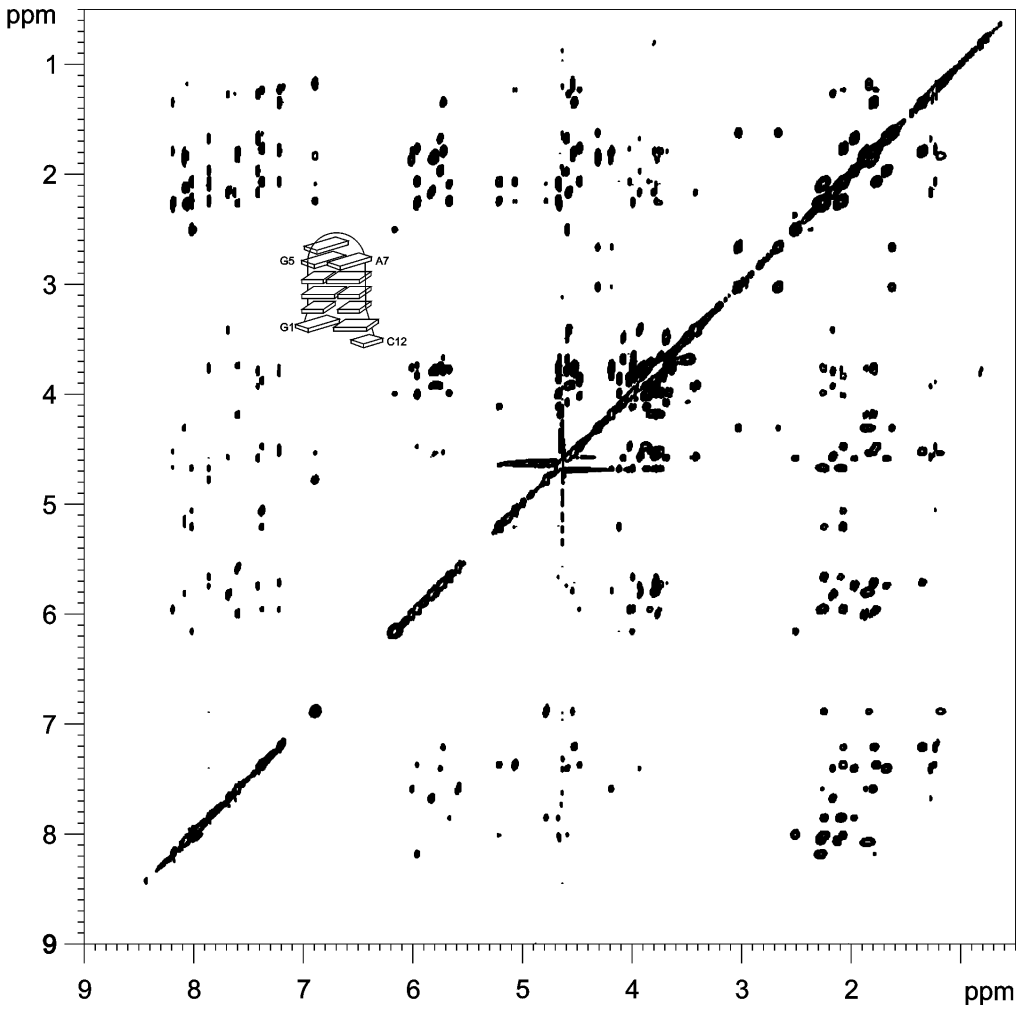


Fig.3.24: NOESY spectrum of d(GTGCGAAGC^{T}AC) in D_2O solution, recorded at 288 K with a mixing time of 300 ms.

3.4.4.1 Loop Formation

For the bulge sequence d(GTGCGAAGC^{T}AC), the A6H4', A6H5'/H5" protons of the bridgehead adenosine (data not shown) were observed at 2.094 and 2.299 ppm whilst the A11H4' and A11H5'/H5" protons from the stem region were observed at 4.163 and 4.073 ppm respectively. The difference in chemical shift between the stem and loop adenine H4' and H5'/H5" sugar protons is >2 ppm in this case, consistent with the observations of Hirao *et al.* (*98*). (N.B. The H5' and H5" resonances were overlapped for both A6 & A11; the average value of each is reported).

3.4.4.2 Imino Protons of Free DNA

The imino proton region of the spectrum at 278K (Fig.3.25(b)) reveals only two resonances for stabilised Watson-Crick pairs in each hairpin structure. The G3 and G8 imino resonances, at 13.15 and 13.05 ppm, are evident for d(GTGCGAAGC^{T}AC),

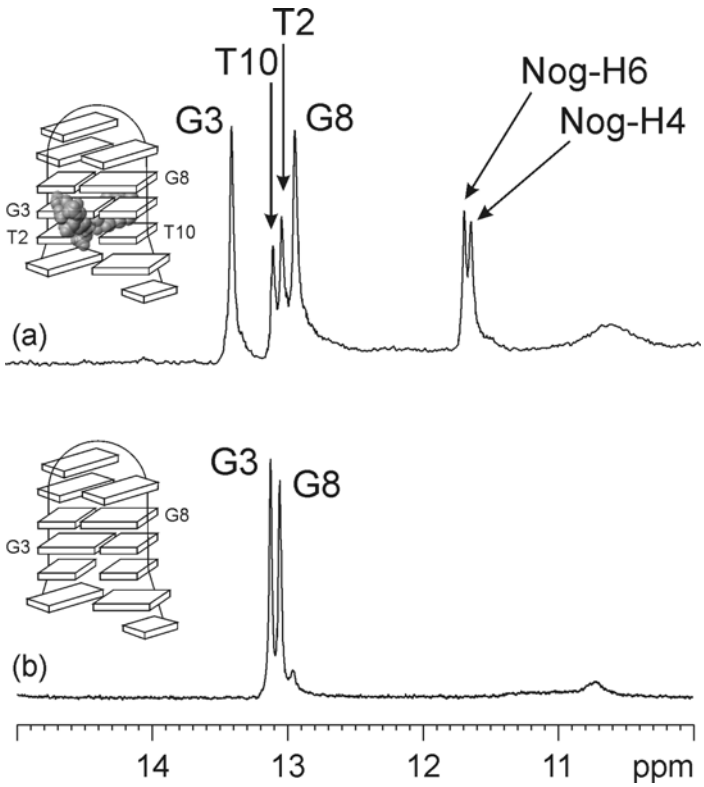


Fig.3.25: Imino-proton region of the 500 MHz ^{1}H NMR spectrum (278 K) of (a) nogalamycin-d(GTGCGAAGC^{T}AC) and (b) d(GTGCGAAGC^{T}AC).

3.4.4.3 Titration

In the 1D ^{1}H NMR of d(GTGCGAAGC^{T}AC), for example, the resonance at 8.24 ppm, corresponding to the A11H8 of the free oligonucleotide, decreases in intensity upon addition of nogalamycin, whilst a resonance appears at 8.43 ppm, corresponding to the A11H8 of the bound species (Fig.3.26). The downfield shift can be rationalised in terms of nogalamycin intercalating between the base pairs. This is consistent with other studies, in which such a shift has been found to be typical of an adenine involved in a 5'-CA intercalation site (*240;241*).

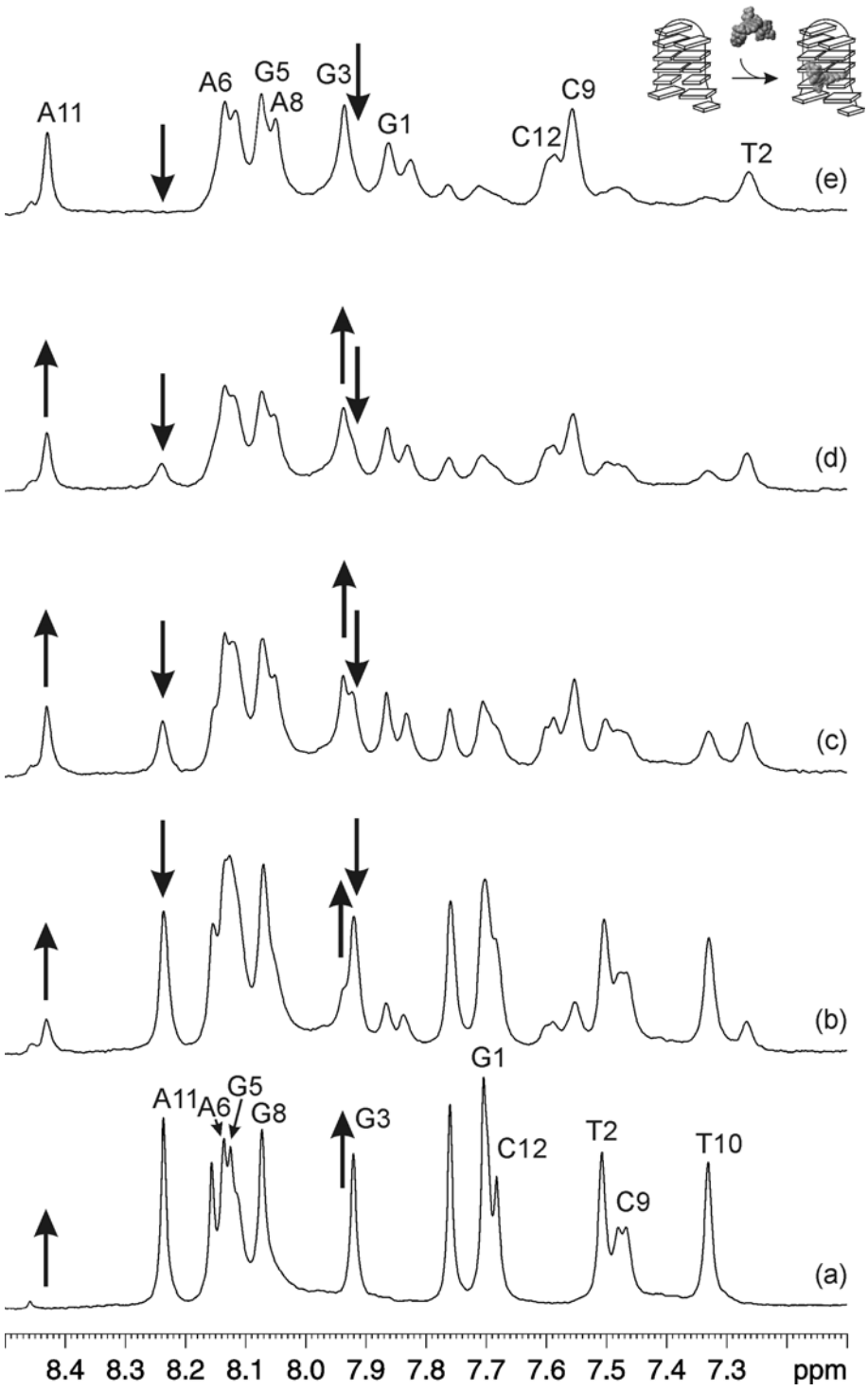


Fig.3.26: 1D ^{1}H NMR spectra of the titration of d(GTGCGAAGC^{T}AC) with nogalamycin ($D_2$0, 288K), highlighting the imino region (7.0-8.5 ppm), at drug:DNA ratios of (a) 0:1, (b) 1:4, (c) 1:2, (d) 3:4 and (e) 1:1. The downward pointing arrows highlight clearly defined signals from the free DNA that decrease upon addition of the nogalamycin. Upward pointing arrows indicate new signals that increase owing to complex formation.

3.4.4.4 Imino Protons for Drug-DNA Complex

The nogalamycin-d(GTGCGAAGC^{T}AC) complex reveals two sharp imino proton resonances corresponding to G3 and G8, as evident for uncomplexed d(GTGCGAAGC^{T}AC), plus two additional weaker resonances of similar intensity at 13.01 and 13.08 ppm (Fig.3.25(a)), which simultaneously broaden and disappear above 288 K. Theses signals can be unambiguously assigned based on the 2D WATERGATE NOESY spectrum (90 % H_2O, 10% D_2O, 288 K, 300 ms): strong NOEs can be observed between the resonance at 13.01 ppm and the A11 amino proton at 8.44 ppm, and between the signal at 13.08 ppm and both G3NH_2 protons at 6.71 and 7.90 ppm, indicating that these are the T2 and T10 protons respectively (Fig.3.28). The relative broadness of these peaks is indicative of some form of conformational averaging. Imino-amino NOE crosspeaks can be similarly used to identify the G3NH (13.46 ppm), which interacts strongly with the G3NH_2 protons at 6.68 and 7.95 ppm, and the G8NH (12.99 ppm), which gives strong crosspeaks to the G8NH_2 protons at 6.30 and 8.06 ppm.

Examination of the imino region of the WG-NOESY spectrum provides an insight into the structure of the drug-DNA complex. NOE crosspeaks are observed between the adjacent base pairs in the loop-closing region of the complex: strong interactions are evident between the G8 imino proton of the C4-G8 base pair and the G3 imino proton of the adjacent G3-C9 base pair (Fig.3.29(a)). This through-space interaction is typical of stacked C-G base pairs on B-form DNA (*106*).

The broad NOE crosspeak between the G3 and T10 imino protons would initially appear to indicate that the T10 base is stacked onto the flanking G3-C9 base pair. However, since these bases are separated by more than 4.5 Å, which is the maximum interproton distance over which NOEs can normally be detected, by the presence of the drug chromophore, it would seem unlikely that this NOE would be observed. This observation can, however, be

rationalised in terms of spin-diffusion: as a consequence of spin diffusion, NOE crosspeaks between pairs of protons that are far apart can gain intensity from magnetisation that has been transferred via an intervening spin. In this case, the H3 proton of nogalamycin at 7.06 ppm, which gives strong NOEs to each of the T2 and T10 iminos (data not shown), is the proton involved in the magnetisation transfer. This is supported by the absence of NOEs between G3NH and T10NH at lower mixing times of 125 and 100 ms. Cross-strand interactions between G3 and T10, together with the presence of a strong NOE between the T2 and T10 imino protons, rule out the formation of an extrahelical T10 bulge. When combined with the absence of a G1 imino resonance, which would indicate the formation of the G1-C12 base pair, this forms compelling evidence that both thymines are involved in a T-T *wobble* base pair (Fig.3.29(b)) on the upper face on the nogalamycin chromophore. This would induce a 5'-3' frameshift at the 3'-terminal of the hairpin, precluding the formation of the terminal G-C base pair and resulting in a single 3'-dC overhang (Fig.3.27).

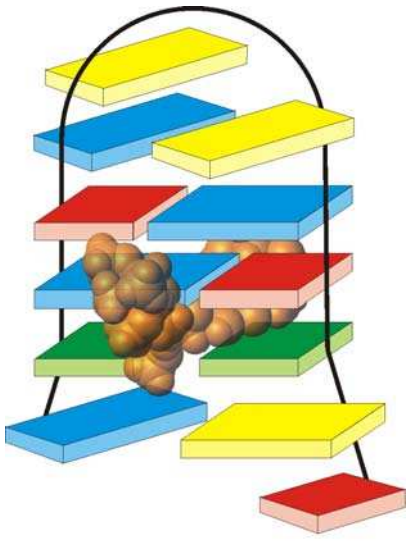


Fig.3.27: Schematic of the proposed conformation of nogalamycin-d(GTGCGAAGC^{T}AC). Cytosines are shown in red, thymines in green, guanines in blue, and adenines in yellow.

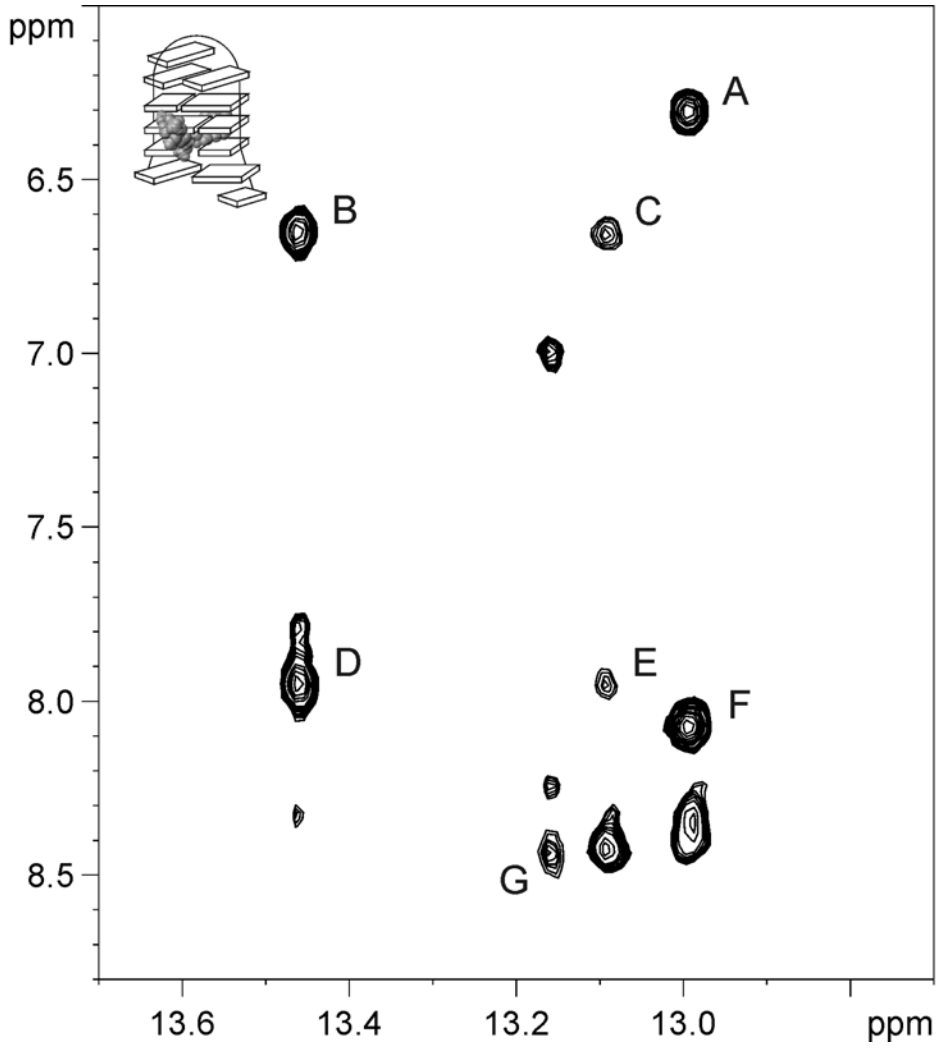


Fig.3.28: A section of the NOESY contour plot (90% H_2O, 10% D_2O, 288 K, 300 ms) of nogalamycin-d(GTGCGAAGC^{T}AC), expanded to show the NOEs between the exchangeable base protons in the stem region. NOEs, labelled (A)-(G) are assigned as follows: (A) G8NH_2*-G8NH, (B) G3NH_2*-G3NH, (C) G3NH_2*-T2NH, (D) G3NH_2-G3NH, (E) G3NH_2-T2NH (F) G8NH_2-G8NH, and (G) A11NH_2-T10NH.

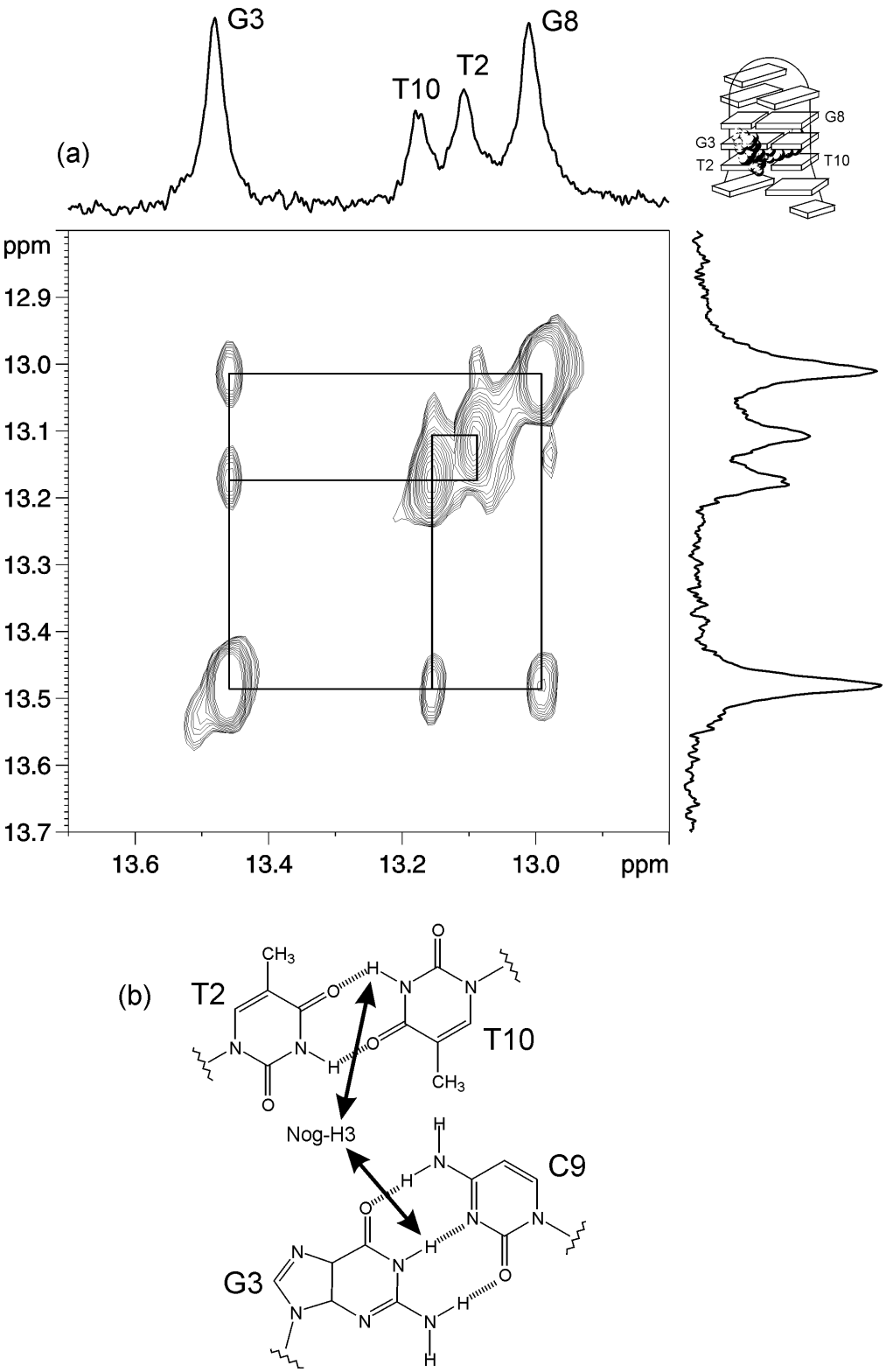


Fig.3.29: (a) ^{1}H NMR and NOESY spectra of nogalamycin-d(GTGC<u>GAA</u>GC^{T}AC) at 288K (300 ms, 90% H_2O, 10% D_2O), expanded to highlight the imino region (12.8-13.7 ppm). Note the crosspeaks between T2NH (13.01 ppm) and T10NH (13.08 ppm) are partially obscured due to their closeness to the diagonal. (b) Schematic representation of spin-diffused NOE interactions between G3 and T10 imino protons.

3.4.4.5 Drug-DNA Interactions

The location and orientation of the bound drug molecule is readily established on the basis of drug-DNA NOEs. The aromatic proton Nog-H11 presents clear interactions with T2 and G3, indicating that the drug is intercalating at the 5'-TG step. A large number (34) of drug-DNA contacts were observed in the NOESY spectrum (D_2O, 288 K, 300ms) of nogalamycin-d(GTGCGAAGC^{T}AC).

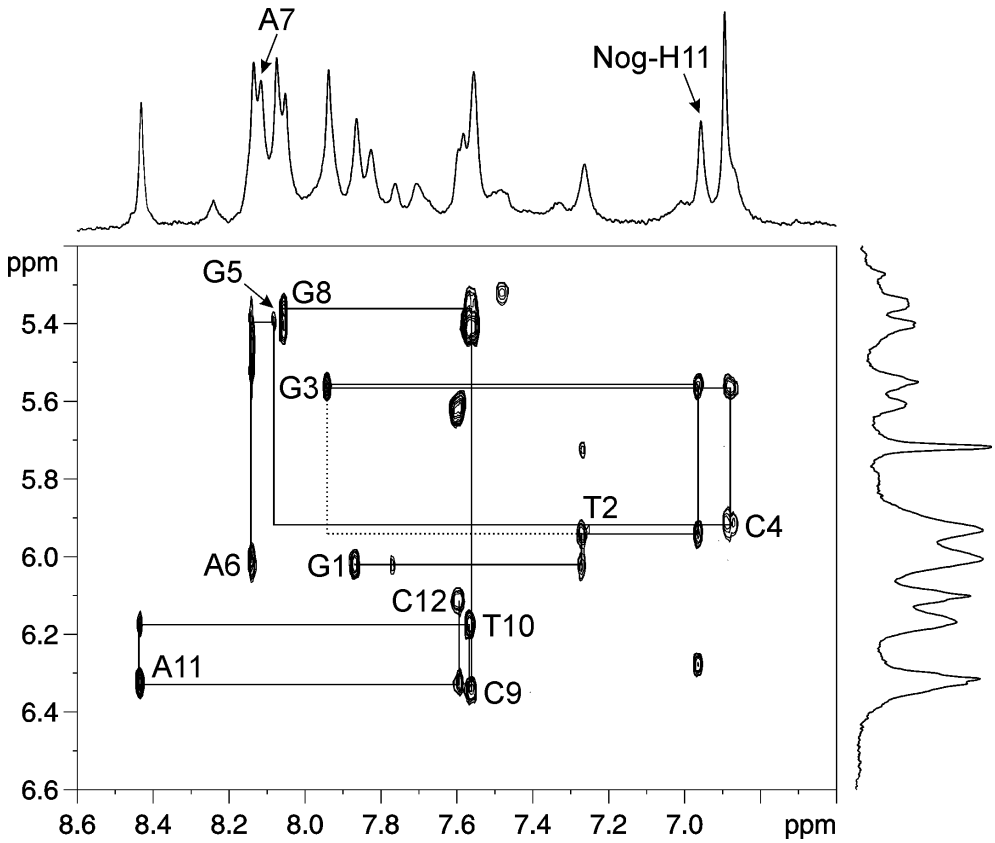


Fig.3.30: Expanded region of the ^{1}H and NOESY spectra (D_2O, 288 K, 300ms) of nogalamycin-d(GTGCGAAGC^{T}AC), showing NOE connectivity pathways between H6/H8 & H1' protons.

The H6/H8-H1' sequential connectivity can be followed along the DNA backbone through the intercalation site, with the ligand bridging the gap between the T2 and G3 resonances, and on through a break in the pathway at the C4-G5 step consistent with loop formation (Fig.3.30). The T2H1' proton interacts with its own H6 proton and also to the H11 proton of nogalamycin. A

crosspeak can also be observed corresponding to the NOE between the H11 proton of nogalamycin and the H1' sugar proton of G3. The H11 proton also gives weaker crosspeaks to the G3H1' and C4H6 resonances. These data clearly define the intercalation site to be the TpG step as was expected.

There were extensive drug-DNA NOEs observed in nogalamycin-d(GTGCGAAGC^{T}AC) between nogalamycin methyl resonances and the T2 sugar, as would be expected as the thymine was positioned within the intercalation site. There were, however, NOEs noted for the nogalose sugar protons (H2", H4' and H5') to T10, suggesting that this thymine was not in a bulged conformation and had taken the place of the A11 as would occur for a frame shift event.

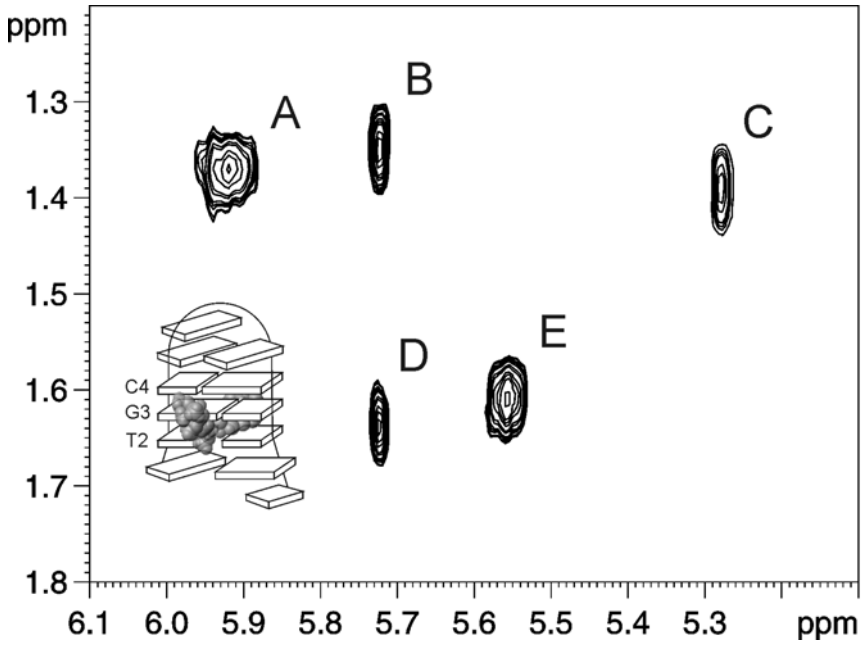


Fig.3.31: Expanded region of the NOESY spectrum of nogalamycin-d(GTGCGAAGC^{T}AC), highlighting key drug-DNA interactions at the intercalation site. NOEs, labelled (A)-(E), are assigned as follows: (A) Nog-3'-CH_3–C4H1', (B) T2CH_3–NogH1", (C) Nog-3'-CH_3–NogH1', (D) Nog-5"-CH_3–NogH1" and (E) Nog-9-CH_3–G3H1'.

Among the drug-DNA NOEs at the intercalation site are those between Nog-H1" and T2CH_3 (Fig.3.31), Nog-H7 and T2H4', Nog-H11 and all T2 sugar protons H1'-H4', and between the nogalamycim methyl ester (H21) and

T2H1'/H2'. There is no evidence of any NOEs between these protons on the drug chromophore and the A11 adenosine. This suggests that the A11 base is not adjacent to the drug. The presence of weak crosspeaks between T10 base protons and the drug chromophore further favours a model in which both T2 and T10 are stacked over the drug.

Numerous NOEs from methyl and methoxy groups on the nogalose sugar identify hydrophobic interactions within the minor groove consistent with previous studies (*194;241;256*). NOEs from NogH1" and Nog-5"-CH_3 on the bicycloamino glucose sugar to T2CH_3 confirm its orientation in the major groove with the drug hydroxyl groups positioned to interact with the G3-C9 base pair (data not shown).

The resonances for T10 show evidence for significant exchange broadening. However, weak NOEs from Nog-2'-OCH_3 to T10H2"/H4', together with the observation of the two broadened imino proton resonances (Section 3.4.3), verify that the drug stacks with a weakly stabilised T2-T10 mismatched pair with the 3'- and 5'-termini poorly structured, suggesting considerable conformational averaging in which the T2-T10 may be only one possible contributor. Thus, the precise nature of the terminal base pairs remains poorly defined in this complex.

3.4.5 NMR Analysis of d(CCCAC<u>GTA</u>GT^{T}GG)

The effects of a bulged thymine on the complementary strand were examined in the context of the hairpin d(CCCAC<u>GTA</u>GT^{T}GG), where an additional thymine is inserted at the 5'-TG step (5'-T^{T}G). In this case, the marginally less stable, but more easily characterised, GTA loop sequence was incorporated. In order to facilitate a frameshift, a homopolymeric run of three dC residues was included at the 5'-terminal of this oligomer. Two of the three cytosines were matched, by normal Watson-Crick base pairing, to guanine bases at the 3'-terminal end. The remaining residue, which was included to form part of

the symmetry-related CpA step associated with a normal TpG nogalamycin binding site, was mispaired with a dT residue. The most likely conformation for this sequence to adopt is shown in Fig.3.32.

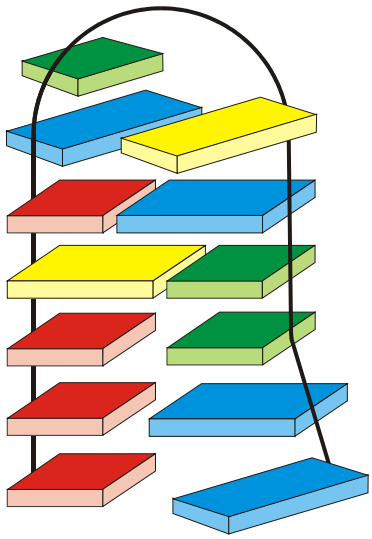


Fig.3.32: Structure of the oligonucleotide sequence d(CCCAC<u>GTA</u>GT^{T}GG) in its proposed folded conformation with frayed ends. Colour scheme as for Fig.3.13.

3.4.5.1 Imino protons of Free DNA

As stated, only two imino resonances can be detected on the ^{1}H spectra of d(GTGC<u>GAA</u>GC^{T}AC) and d(GC^{T}AC<u>GAA</u>GTGC). Similarly, the 1D spectrum of d(CCCAC<u>GTA</u>GT^{T}GG) reveals only two imino resonances; the G9 imino at 13.24 ppm, corresponding to the loop-closing C-G base pair, and an exchange-broadened T10 imino at 13.53 ppm from the A-T base-pair.

3.4.5.2 Loop Formation

G*N*A hairpin loop formations was verified for d(CCCAC<u>GTA</u>GT^{T}GG) by reference to the H4'-H2'/H2" region of the NOESY spectrum (D_2O, 298 K, 300 ms) (Fig.3.33) as described for d(GC^{T}AC<u>GAA</u>GTGC) (Section **3.4.3.1**).

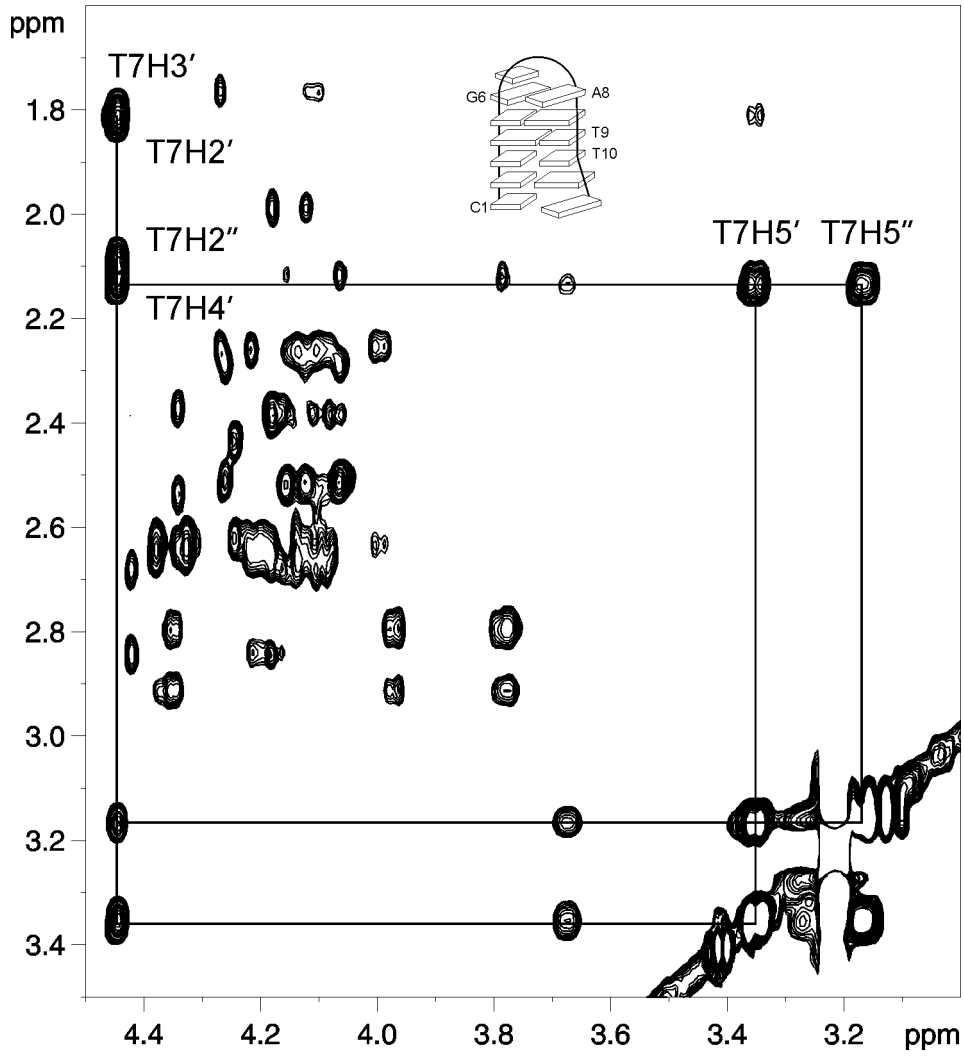


Fig.3.33: Expanded region of the NOESY spectrum of d(CCCAC<u>GTA</u>GT^{T}GG) showing the H4'-H2'/H2" interactions of the loop nucleotide.

3.4.5.3 Titration

When d(CCCAC<u>GTA</u>GT^{T}GG) is titrated with nogalamycin, the resonances of the free DNA are replaced not by a single set of new resonances, as for d(GTGC<u>GAA</u>GC^{T}AC) and d(GC^{T}AC<u>GAA</u>GTGC), but rather by two sets

corresponding to two species that are approximately equally populated at 278 K (data not shown).

3.4.5.4 Drug-DNA interactions

The hydroxyl region (10.2-11.8 ppm) of the ^{1}H spectrum (278 K, 90% H_2O, 10% D_2O, 300 ms) shows matched pairs of drug protons (Fig.3.35). This strongly indicates the presence of two, non-equivalent species in equilibrium. No assumptions were made about the nature of these two equilibrium states. Given the requirement at the binding site for a 5'-TG step, two possible drug-induced DNA conformations that are consistent with the NMR data can be envisaged (Fig.3.34). The first has the extra dT flipped out of the stack resulting in a one-base frameshift, preserving Watson-Crick base pairing down stream (C3-G12 and C2-G13), but resulting in a single dC overhang (Fig.3.34(a)). Alternatively, the drug stabilises a mismatch C-T base pair at the drug intercalation site with complementary base pairing elsewhere in the structure (Fig.3.34(b)).

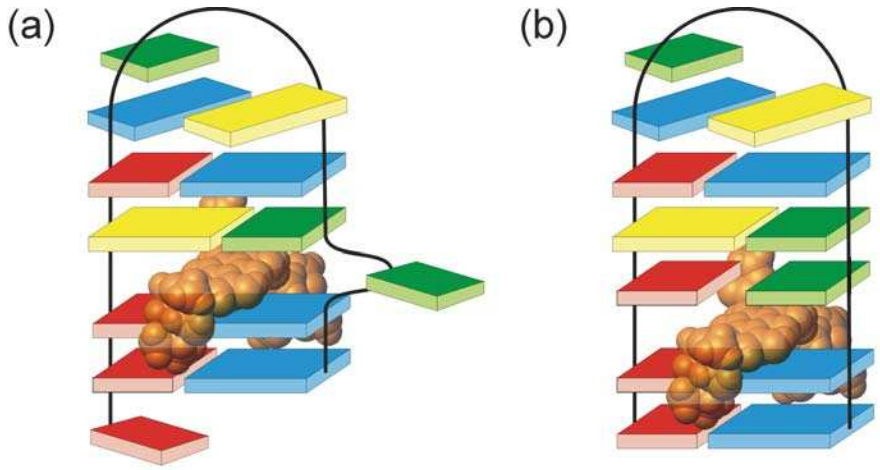


Fig.3.34: Schematic representations of the two proposed conformations of nogalamycin-d(CCCAC<u>GTA</u>GT^{T}GG), based on the requirement for a 5'-TG intercalation step.

As discussed, the ^{1}H WATERGATE spectrum of nogalamycin-d(CCCAC<u>GTA</u>GT^{T}GG) shows matched pairs of drug protons at 278 K (Fig.3.35(b)), indicating the presence of two non-equivalent species in

equilibrium. The sample was heated from 273K to 308K in 5K steps, and a ^{1}H WATERGATE spectrum collected at each temperature. Upon heating, these exchanging peaks broaden and coalesce until, at 303K, only the loop and terminal stem protons, which are relatively unaffected by the exchange process, can still be detected and assigned. To determine the nature of the various species present, further studies were carried out at 278K and 303K.

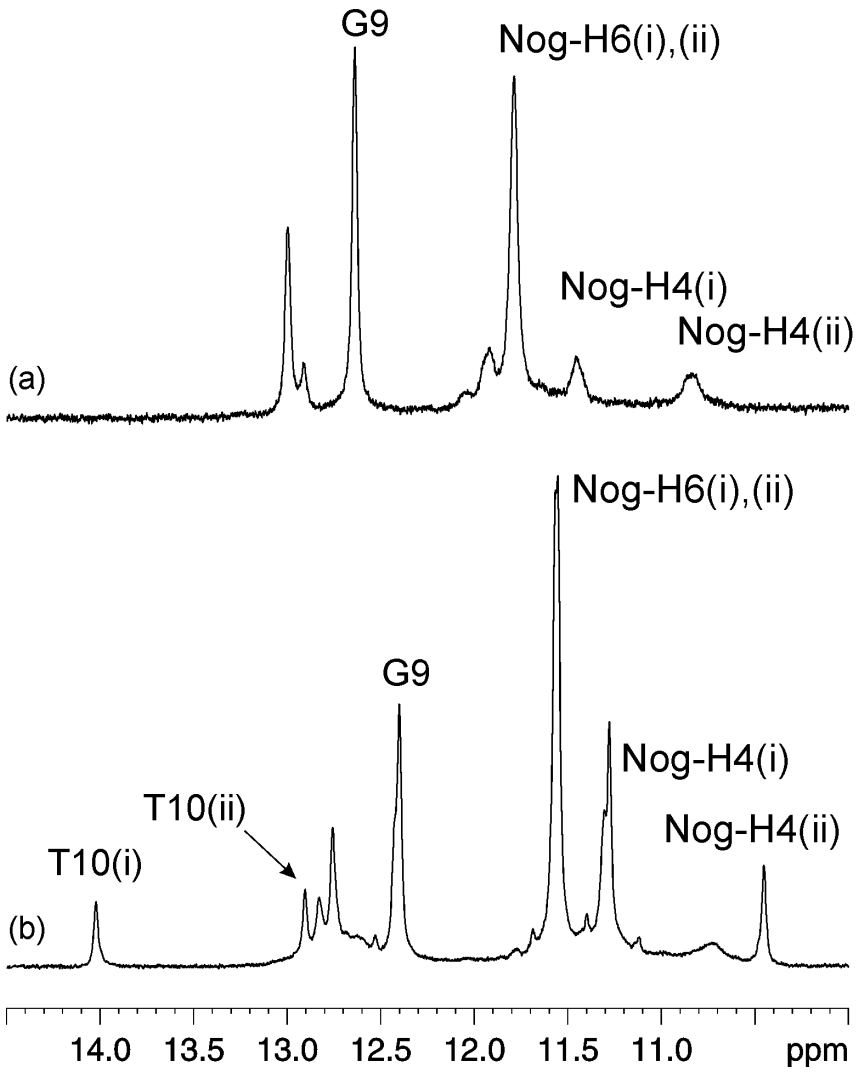


Fig.3.35: Imino and hydroxyl regions of 1D ^{1}H WATERGATE spectra of nogalamycin-d(CCCAC<u>GTA</u>GT^{T}GG) complex at (a) 303K, and (b) 278K, illustrating matched pairs of drug protons at the lower temperature.

The imino region (12.0-14.3 ppm) also provides an insight into the nature of the exchanging species: The stacking of an A-T base pair directly on the drug

has previously been shown to result in a characteristic upfield shift of the thymine N3 proton by ≈ 1 ppm to ≈ 13 ppm (*240*). At 278 K, a resolved peak at 13.95 ppm in a position characteristic of Watson-Crick A-T base pair stacked onto nogalamycin can be seen (Fig.3.35). Since nogalamycin-d(CCCAC<u>GTA</u>GT^{T}GG) has only a single A-T basepair, this resonance can be attributed to the C-T mismatched structure (Fig.3.34(b)), in which the A4-T10 base pair is isolated from the shielding effect of the drug chromophore by the mismatch base pair. At the lower temperature, this proton is in slow exchange with another thymine *N3* proton, which comes into resonance at 12.87 ppm (Fig.3.35(a)), indicating that it is directly adjacent to the intercalated drug. This can be attributed to the alternative conformation of the complex, in which T11 is flipped out and T10 forms part of the interaclation site.

Despite the fact that a full assignment of d(CCCAC<u>GTA</u>GT^{T}GG) could not be obtained, a sufficient number of signals were unambiguously resolved to allow the orientation of the nogalamycin to be verified by reference to drug-DNA NOEs on the NOESY spectrum at 303 K. Since deoxyribose H1' protons form part of the floor of the minor groove in B-form DNA, the presence of crosspeaks between these protons on residues C2, C3, G12, & G13, and Nog-2'-OCH_3, Nog-3'-OCH_3, Nog-4'-OCH_3 & Nog-5'-CH_3 on the nogalose sugar (Fig.3.36) verify that the nogalose resides in the minor groove and is oriented away from the loop sequence towards the ends on the oligomer. The absence, as well as the presence, of NOESY crosspeaks can also tell us about the structure of the complex. The absence of a second downfield shifted H6 proton, corresponding to C3 cytosine, provides sufficient evidence that this residue remains stacked into the helix, and does not flip out as in previous studies (Section **3.1.7**).

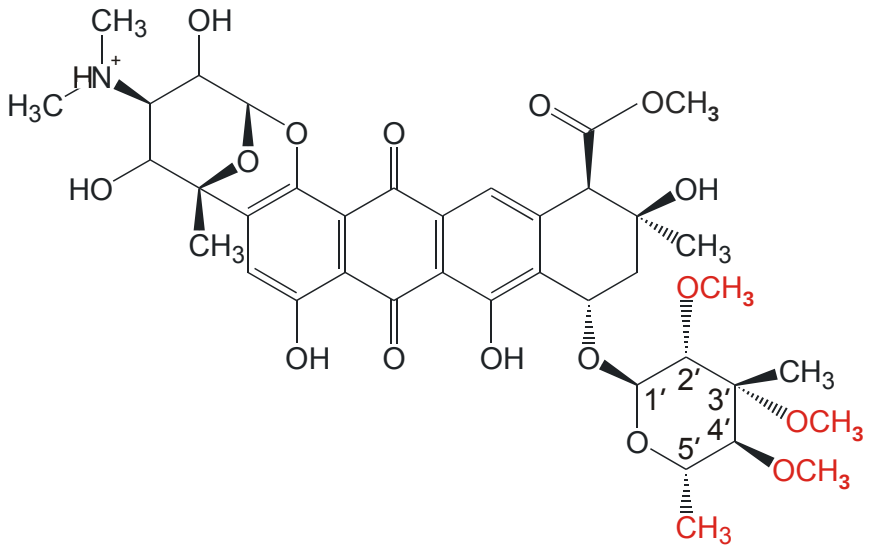


Fig.3.36: Nogalamycin, highlighted to emphasise Nog-2'-OCH_3, Nog-3'-OCH_3, Nog-4'-OCH_3 & Nog-5'-CH_3 protons on the nogalose sugar.

The presence of two species in equilibrium makes analysis of these data extremely difficult. Even at 278 K, the NOESY spectrum is complex and broad, making it impossible to fully assign. However, sufficient information can be gleaned from key regions of the spectrum to deduce the conformations of these co-existing species. Examination of the H6/H8-H1' region of the NOESY spectrum collected at 278 K also points to the existence of two species in equilibrium (Fig.3.37). The assignment pathway is common for both species for the loop sequence d(ACGTAG), but bifurcates for the 3'-terminal residues. The first of these two deviating pathways (red) involves a downfield shifted T11 thymine H6 proton indicating a looped-out dT residue, and a four-step H6/H8-H1' assignment pathway between T10 and G13; T10-(Nog-H11)-G12-G13. The second species (blue) lacks a flipped-out base, and with a five-step assignment pathway; T10-T11-(Nog-H11)-G12-G13. The latter data are consistent with the model in which nogalamycin intercalates in the T11-G12 pyrimidine-purine step.

The structural differences between a TpG nogalamycin binding site containing a mismatched C-T pair and one with a flipped-out thymidine are sufficient to significantly change the magnetic environment that the drug experiences. The

structure of nogalamycin in each of these distinct environments, with specific reference to the aglycone hydroxyl protons is shown in Fig.3.38. The Nog-4-OH and Nog-6-OH protons of nogalamycin, when intercalated into a mismatch containing TpG site, are subject to ring currents from the base-pairs adjacent to both faces of its aglycone chromophore (Fig.3.38(a;c)). However, the localised unwinding of the DNA helix associated with the presence of a looped-out base effectively translates the A-T pair flanking the top face of the drug from the centre of the chromophore towards the amino end of the drug (Fig.3.38(b;d)), and shifting the ring currents away from the Nog-4-OH, and even more so the Nog-6-OH, protons.

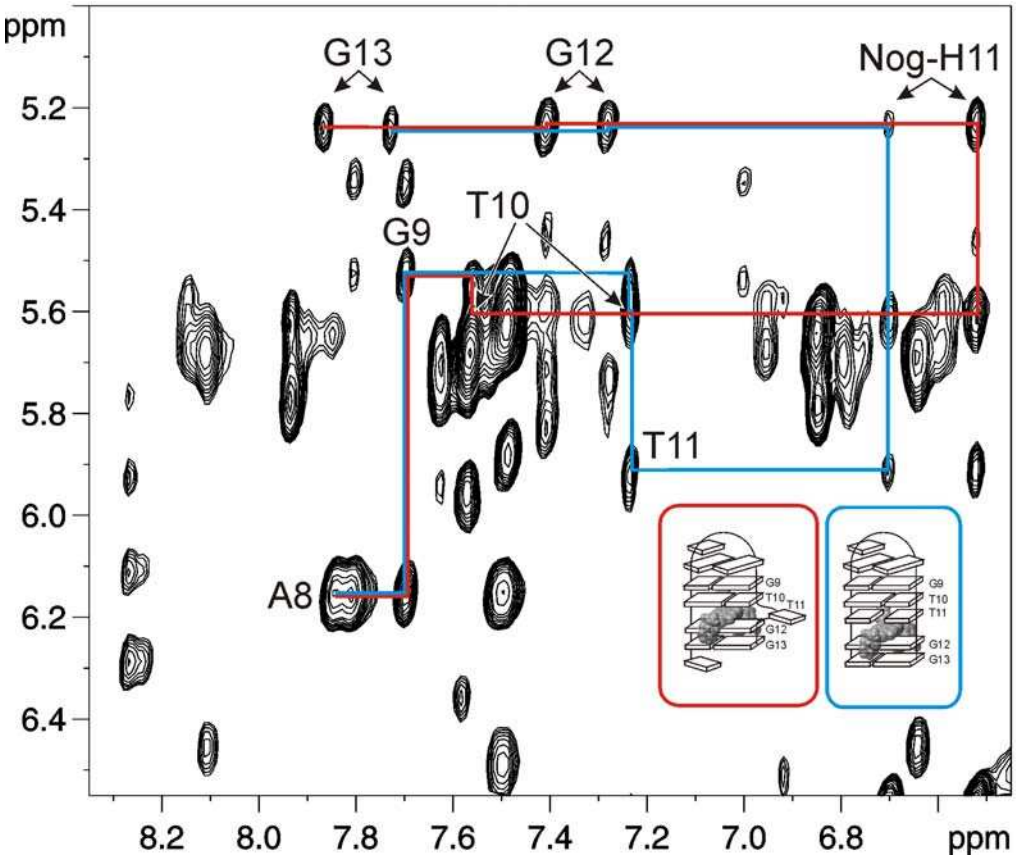


Fig.3.37: Expanded region of the 600 MHz NOESY (278 K, 90% H_2O, 10% D_2O, 300 ms) of nogalamycin-d(CCCAC<u>GTA</u>GT^{T}GG), showing two distinct sequential H1'-H6/H8 connectivity pathways at the 3'-terminal end.

This observation is consistent with the data from the 1-D spectra of the complexes (Fig.3.35). Since the environmental change is greater for the Nog-4-OH protons, it is in slow exchange between the two equilibrium states on the

NMR timescale. The Nog-6-OH protons, in contrast, are less affected by the movement of the flanking bases relative to the drug, and subsequently are in fast exchange, appearing as one broadened peak on the spectrum.

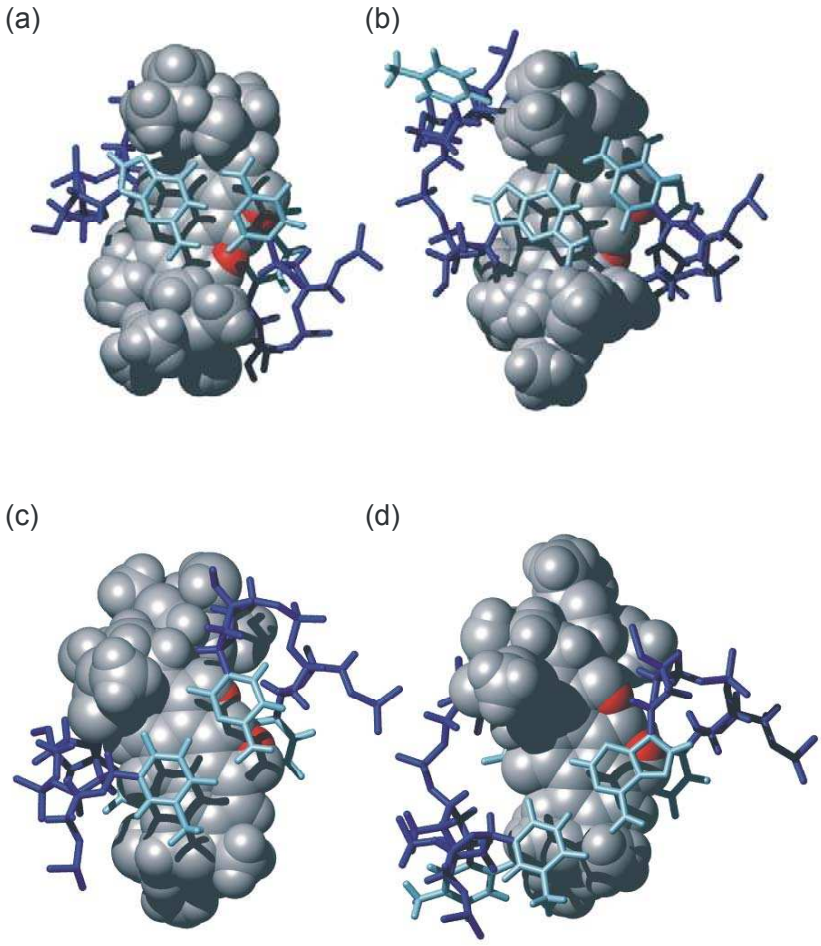


Fig.3.38: Structure of the TpG intercalation site of nogalamycin containing a C-T pair (a;c), and a looped out dT (b;d) after 100ps of MD simulation. Drug hyroxyl protons, 4-OH and 6-OH, are shown in red.

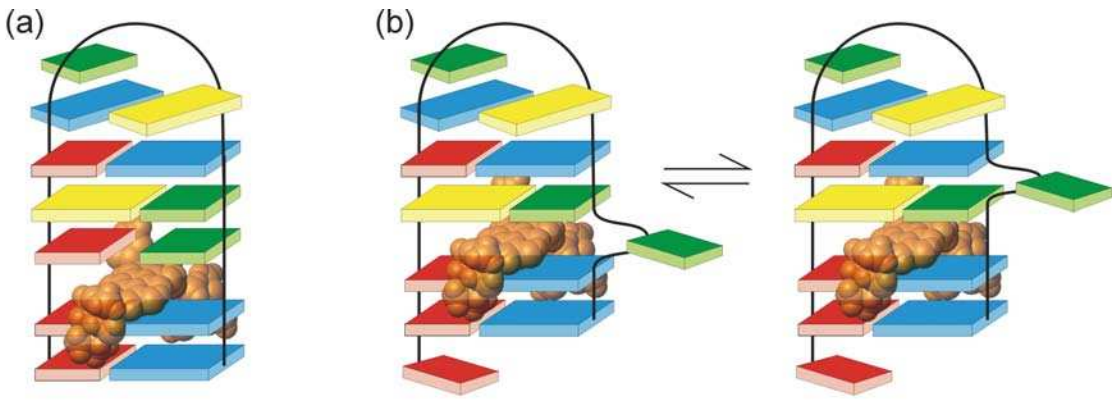


Fig.3.39: Alternate flipped-out orientations of nogalamycin-d(CCCAC<u>GTA</u>GT^{T}GG).

In these variable-temperature studies, the signal at 13.95 ppm and its partner at 12.87 ppm (Fig.3.35(a;b)) become exchange broadened above 298K as the two conformations begin to rapidly interconvert. Although the data show evidence for T11 in the extrahelical conformation at both temperatures, the possibility cannot be excluded that T10 and T11 are also interconverting between alternate flipped-out structures; one with an extrahelical T11 bridging the T10-G12 intercalation site, and another in which a looped out T10 precedes a T11-G12 binding site (Fig.3.39(b)). These structures could not be differentiated by the methods employed. Thus, the data are considerably more ambiguous than described for nogalamycin-d(GC^{T}AC<u>GAA</u>GTGC). Despite extensive efforts to optimise the spectra of nogalamycin-d(CCCAC<u>GTA</u>GT^{T}GG), none of the interconverting species could be isolated to such a degree that a full assignment could be completed and distance restraints could be derived.

3.5 Redesign of C-T Mismatch Sequence

To test the effect of nogalamycin on a mismatch base pair adjacent to its binding site, the C-T mismatch observed in nogalamycin-d(CCCAC<u>GTA</u>GT^{T}GG) was incorporated into a new sequence. This new sequence was designed to eliminate the possibility of forming multiple species in solution. Again a GTA mini-hairpin loop was employed to stabilise the structure and to eliminate the need to synthesise multiple DNA strands. One C-G Watson-Crick base pair and the hairpin loop flanked the mismatch base

pair on one face, while two C-G pairs were incorporated on the other. These base pairs were aligned so as to have alternating purines and pyrimidines and so eliminate the risk of the frameshift that occurred in nogalamycin-d(CCCACGTAGT^{T}GG). In order for nogalmycin to bind, the 5'-TG step must be conserved, and so the potential for the helical thymine to flip out as before is annulled. Thus, d(GCCCGTAGTGC) should give rise to only one possible complex containing a high affinity nogalamycin 5'-TG binding site (Fig.3.40), and this conformation contains the desired C-T base pair.

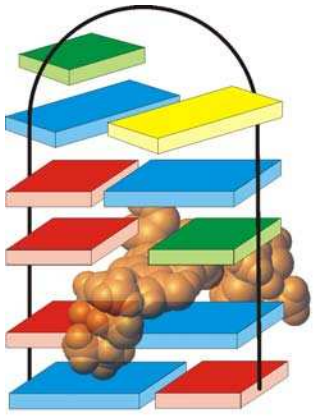


Fig.3.40: Schematic representation of the proposed conformation of nogalamycin-d(GCCCGTAGTGC).

3.6 Materials & Methods

3.6.1 Distance Restraints

Interproton distances for nogalamycin-d(GCCCGTAGTGC) were derived from integration of NOE cross peak volumes from NOESY experiments (288 K, D_2O) with mixing times of 150 ms, 100 ms, and 75 ms as describe in Section **3.3.3**.

3.6.2 Starting Structure

The mismatched conformation of nogalamycin-d(GCCCGTAGTGC) was generated by simulated annealing with DYANA, using a broadly similar procedure to that used for the looped-out structure, nogalamycin-d($GC^TACGAAGTGC$) (Section **3.3.4**). In this case, 232 distance restraints were applied, of which 214 were intra-DNA and 18 drug-DNA restraints. The loop was further defined by the addition of 32 distance restraints as before. H-bonding restraints were added for the sheared G-A base pair and for all Watson Crick base pairs. The output structures were scored based on the target function and the lowest energy structure was selected for further simulation. Minimisation protocol is as described in Section **3.3.4**.

In order to test whether the assumptions that must be made when constructing a hairpin loop from Cartesian coordinates are valid in this context, a second starting structure was generated by these methods for comparison with the DYANA model: A Cartesian coordinate model for DNA structure d(GCCC)·d(GTGC), containing a mismatched C-T pair, was generated using NUCGEN (*29*). NUCGEN requires specification of two strands, and though it only generates reasonable geometries for Watson-Crick base pairs, a single mismatch in a run of complementary pairs is generally well tolerated. The Cartesian coordinates were used to generate an output file in PDB format.

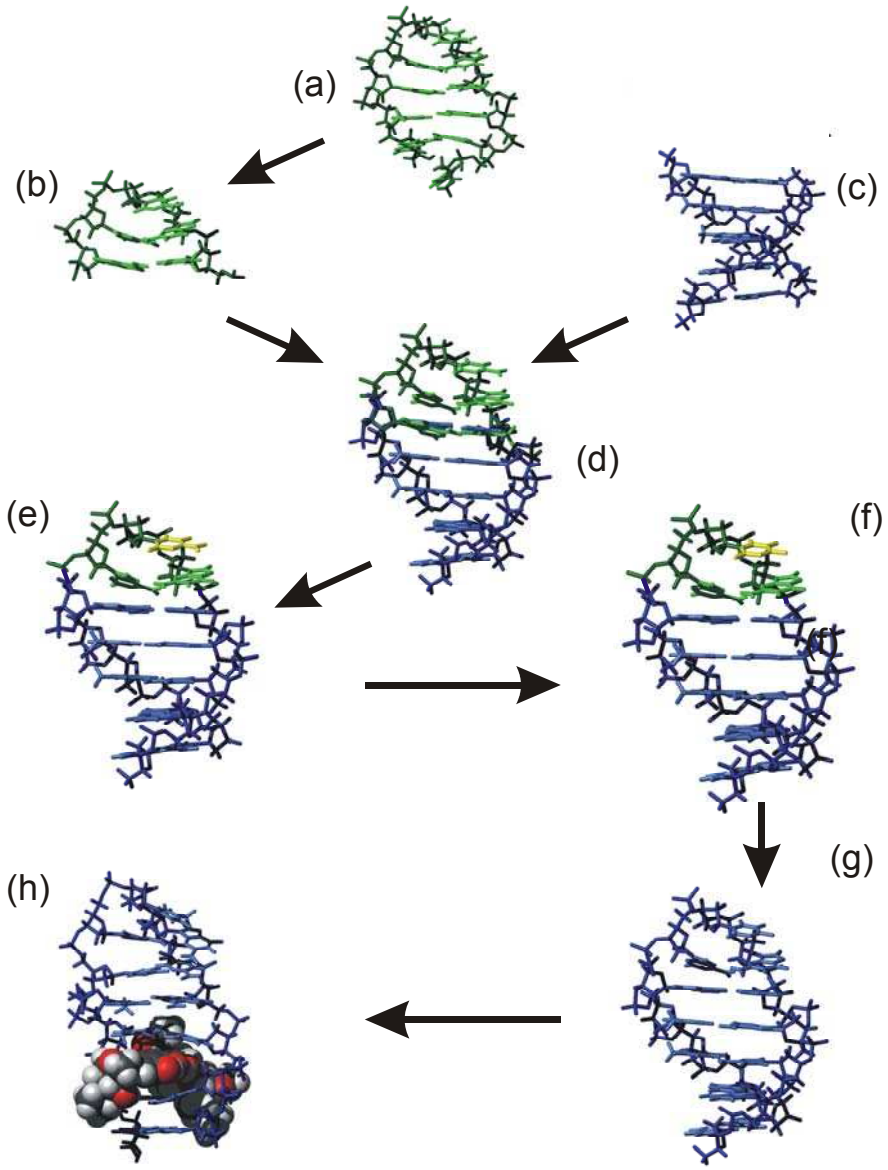


Fig.3.41: Steps involved in generating a model for the sequence d(GCCCGTAGTGC) containing a mismatched C-T base pair and bound to nogalamycin.

A PDB coordinate file of the structure of d(CAATGCAATG) (Fig.3.41(a)), determined by Zhu et. al. (*258*) was downloaded from the archives of the Protein Data Bank. This structure contained a GCA hairpin loop consisting of a single cytidine residue closed by a sheared G-A pair. The PDB file was edited so as to remove the terminal residues, leaving only the GCA turn and the flanking T-A pair (Fig.3.41(b)). The coordinates of this shortened

sequence, d(TGCAA), were loaded into MolMol (*123*), together with those for d(GCCC)·d(GTGC) (Fig.3.41(c)).

The deoxyribose heavy atoms, C1'-C5' and O4', of the A-T pair of d(TGCAA), and the 3'-terminal C-G pair of d(GCCC)·d(GTGC) were superimposed using MolMol (Fig.3.41(d)). The RMSD to the mean averaged over both molecules for the selected atoms was 1.02 Å. A PDB file of the new, overlaid structure was generated and edited to remove the A-T pair from the former structure. The residue numbers of the resulting file were updated, giving rise to the 11-mer d(GCCCGCAGTGC) (Fig.3.41(e)).

In order to create a GTA turn from the pre-existing GCA hairpin, further editing of the PDB was required. The residue name of the pyrimidine in the loop was changed from CYT (cytidine) to THY (thymidine). The N4 (amino) nitrogen was changed to oxygen, and the amino protons deleted. The proton on C5 was substituted with a carbon, which was named C7, consistent with the standard numbering system for a thymine base. The structure of the resulting, transitional, base is shown in (Fig.3.42(b)). The revised PDB file was read into the LEaP module of AMBER 6.0, a program used to prepare input for the AMBER molecular mechanics programs. LEaP attempts to match atoms from the PDB file by atom on a residue-by-residue basis. If atoms are missing relative to its own library file, as is often the case with hydrogens in structures generated by x-ray diffraction methods, LEaP will generate their coordinates using the internal coordinates in its library. Thus, the transitional base is recognised as a thymine by LEaP, which then generates the missing atoms. Once this was completed a PDB file of the new oligomer, d(GCCCGTAGTGC) was generated (Fig.3.41(g)).

As before (Section **2.3.6**), the separation between the stacked bases of the TpG intercalation site was increased, and nogalamycin manually docked. The orientation of the drug in the binding site was based on NOE data from the 2D

NOESY spectrum of nogalamycin- d(GC^{T}ACGAAGTGC) (Section **3.4.4**), and by reference to previously published structures (*211;240;241*). The nogalose was placed in the minor groove, and the aminoglucose inserted into the major groove.

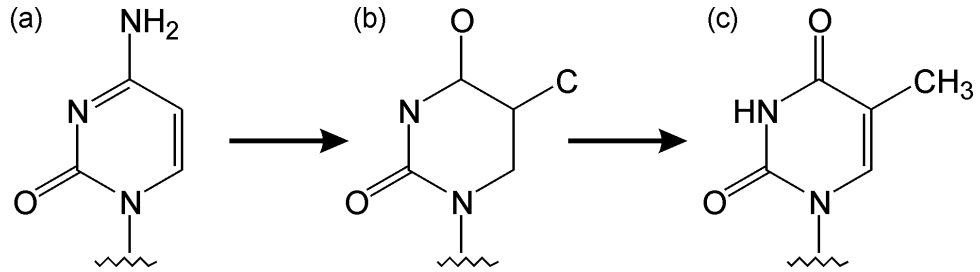


Fig.3.42: Steps involved in generating a thymine base from cytosine using LEaP.

Sodium counter-ions were again used to negate the negative charge on the backbone, and the model was solvated to within a distance of 5 Å using boxes of 216 TIP3P H_2O molecules. The entire system was then energy minimised (Fig.3.41(h)). Following minimisation, the structure was compared to the DYANA starting structure, and these structures shown to have an RMSD of 1.21 (±0.1) Å. In this case, the inherent assumptions associated with creating a starting structure by combining and editing existing structures are reasonable, and produce starting structures of similar quality to those generated by simulated annealing using the torsion angle dynamics protocol of DYANA.

3.6.3 Structure Calculations

The rMD protocol for nogalamycin-d(GCCCGTAGTGC) was identical to that used for for nogalamycin-d(GC^{T}ACGAAGTGC) (Section **3.3.5**). In this case, the system was subject to 268 NOE-derived distance restraints, of which 214 corresponed to DNA-DNA interactions, 36 to drug-drug interactions, and 18 to drug-DNA interactions. PRPAs were fixed as before, and 11 additional H-bonding restraints were applied. The final averaged energy minimised structure, none of the applied distance restraints were violated by more than

0.5 Å, and the mean pairwise RMSD, calculated for all heavy atoms, over the final 100 acquired snapshots was 0.53 (±0.1) Å. Again for nogalamycin-d(GCCCGTAGTGC), the NMR restraints were satisfied well within the nanosecond of simulation.

3.7 Results & Discussion

3.7.1 NMR Spectra

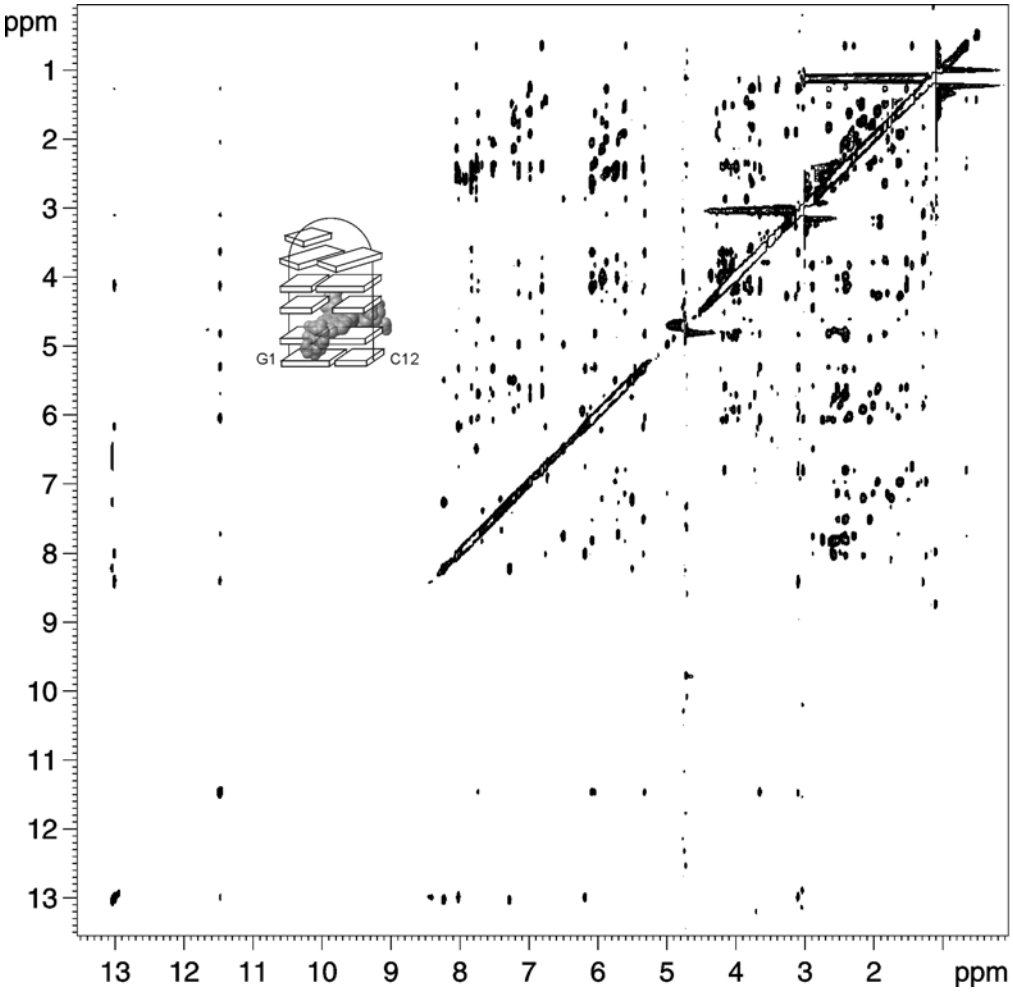


Fig.3.43: NOESY spectrum of nogalamycin-d(GCCCGTAGTGC)in 90% H_2O solution, recorded at 288 K with a mixing time of 300 ms.

NMR studies of the free hairpin show that the loop of d(GCCCGTAGTGC) is formed along with the flanking C4-G8 basepair. The T6 sugar protons, H4', H5' & H5", are shifted upfield to below 3.5 ppm, indicating that the GTA loop

is well formed (Fig.3.45). Furthermore, examination of the imino proton region (12.0-13.5 ppm) of the free oligomer at 288 K (Fig.3.44(b)) reveals that the stem region is well formed, even in the absence of drug, as evidenced by the presence of relatively sharp peaks corresponding to the G8 & G10, as well as the G1 imino protons.

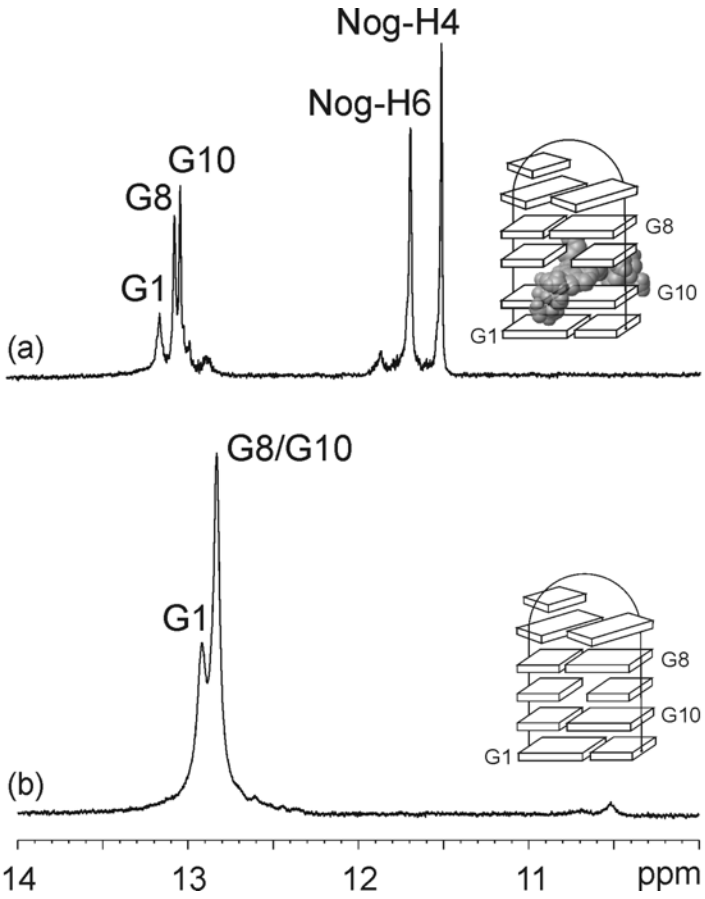


Fig.3.44: Imino and hydroxyl regions of the 600 MHz ^{1}H NMR spectrum of (a) nogalamycin- d(GCCCGTAGTGC) and (b) d(GCCCGTAGTGC) at 288 K, indicating that the stem regions are well formed in both cases.

Drug binding to d(GCCCGTAGTGC) produces a clean 1:1 complex with only a single species present in solution (Fig.3.46), which proves amenable to a detailed NMR analysis. The T9H6 peak at 7.12 ppm is replaced by a clearly-resolved signal at 6.82 ppm corresponding to the same proton in the nogalamycin-d(GCCCGTAGTGC) complex.

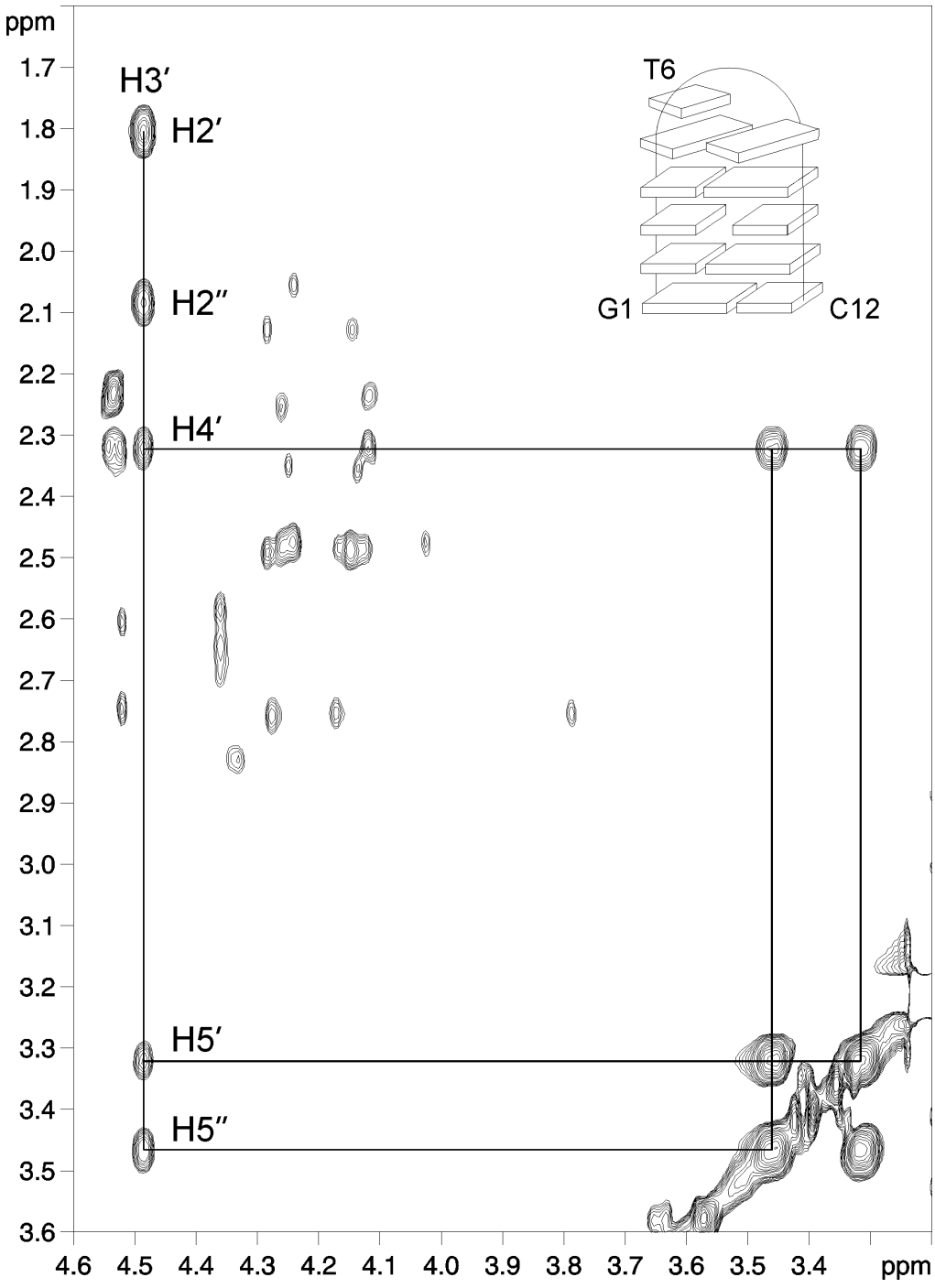


Fig.3.45: Expanded region of the 600 MHz NOESY spectrum of d(GCCCGTAGTGC) (288 K, D_2O) showing interactions between H2'/H2", H3', H4' and H5'/H5" sugar protons on the loop (T6) nucleotide.

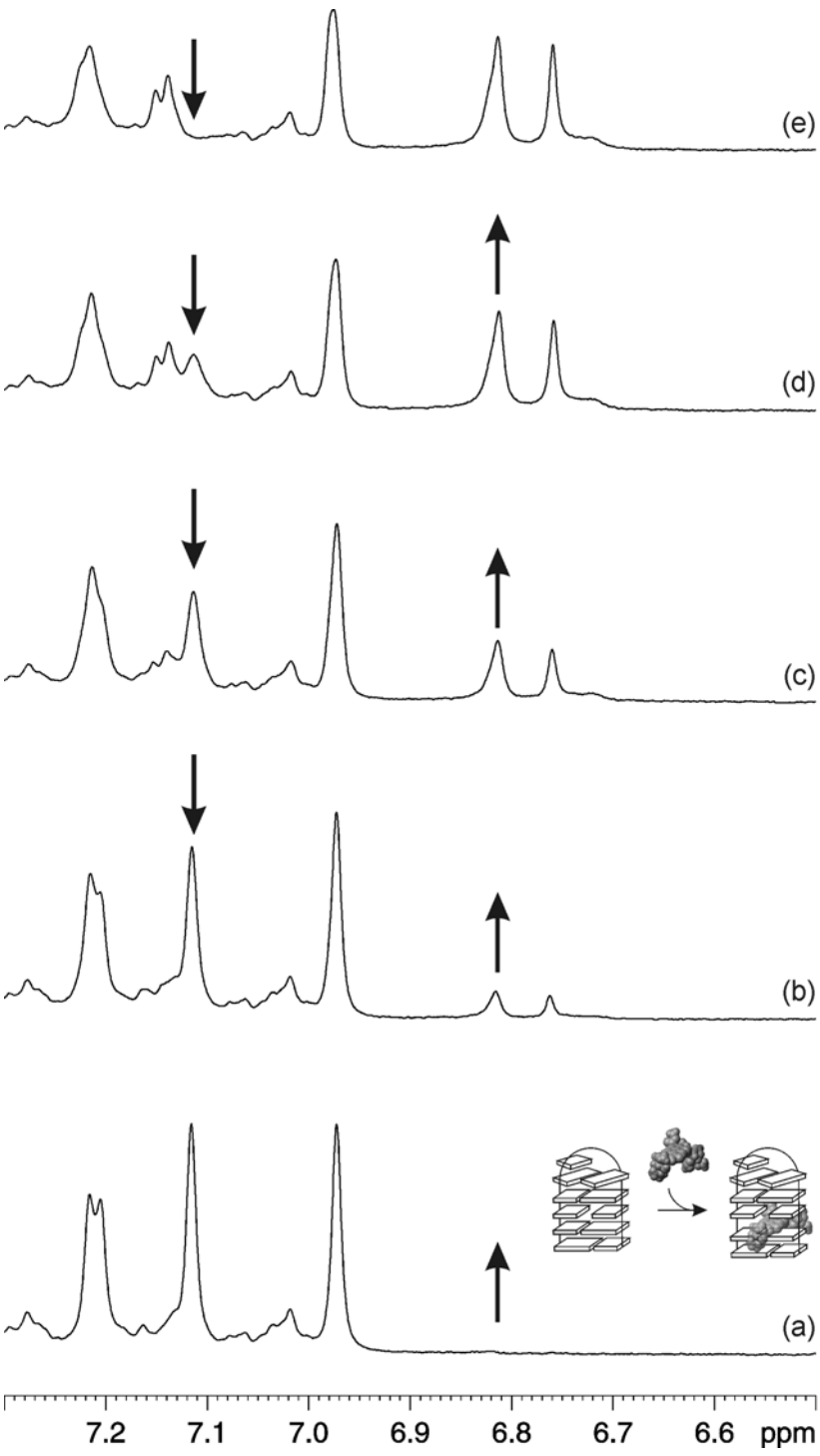


Fig.3.46: 1D ^{1}H NMR spectra of the titration of d(GCCCGTAGTGC) with nogalamycin (D_20, 298K), highlighting the imino region (7.0-8.5 ppm), at drug:DNA ratios of (a) 0:1, (b) 1:4, (c) 1:2, (d) 3:4 and (e) 1:1. The downward facing arrows highlight clearly defined signals from the free DNA that decrease upon addition of the nogalamycin. Upward facing arrows indicate new signals that increase owing to complex formation.

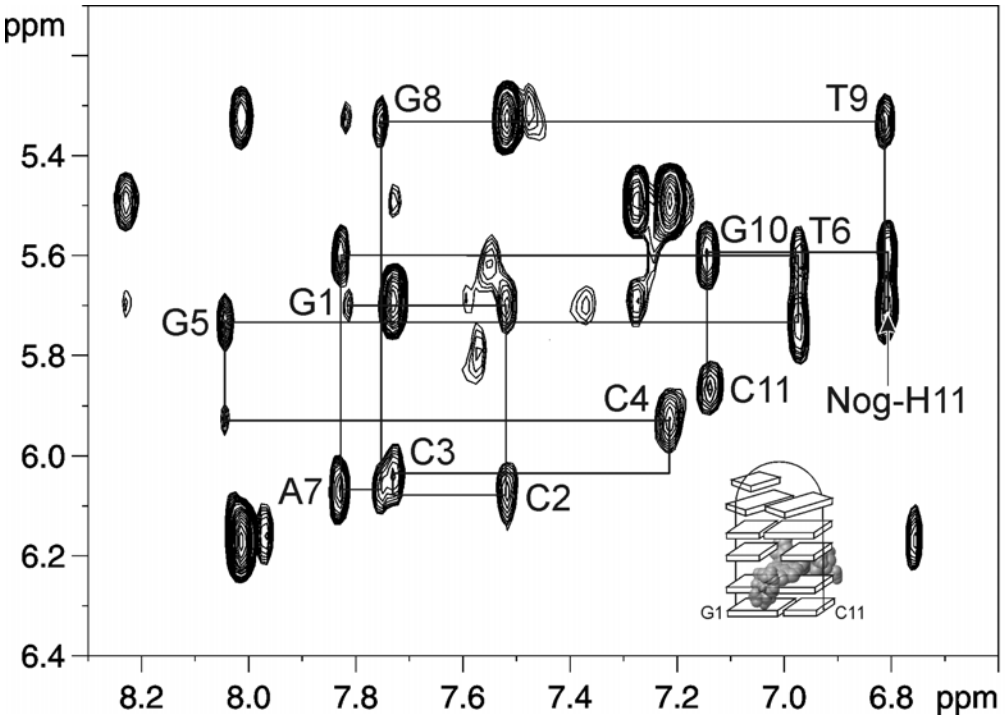


Fig.3.47: Expanded regions of the 600 MHz NOESY spectrum (288 K, 90% H_2O, 10% D_2O) of nogalamycin-d(GCCCGTAGTGC), showing NOE connectivity pathways between H6/H8 and H1' protons, and revealing the location of the intercalation event.

2D NMR (NOESY, TOCSY, DQF-COSY) analysis at 278 K and 288 K enabled a complete sequential assignment to be made, and intermolecular NOEs and chemical shift changes consistent with the proposed structure of nogalamycin-d(GCCCGTAGTGC) to be determined: the sequential connectivities between base H6/H8 and deoxyribose H1' are traced out in Fig.3.47, and show that the NOEs are contiguous except where interrupted across the 5'-TG step due to structural distortions associated with drug intercalation. Loop formation is again verified by sugar H4', H5' and H5" proton shifts (Fig.3.45). The diagnostic NOEs from Nog-H11 to the deoxyribose H1' of T9 and G10 confirm binding at the 5'-TG step. The stacking of the T9 base results in an unusual upfield shift of the T9CH_3 to 0.64 ppm, which is clearly indicative of direct stacking with the bound drug (Section **3.4.3**). Strong NOEs from the T9CH_3 to Nog 5"-CH_3 and Nog- H1" confirm this interaction (data not shown). A detailed analysis of the NOE data identifies many of the drug-DNA interactions previously described (Section **3.4.4**). Although intermolecular NOEs confirm the stacking of the C-T pair

with the intercalated drug, the imino proton of T9 is exchange broadened, suggesting that the C-T mismatch is not significantly stabilised by strong base pair hydrogen bonding.

3.7.2 Molecular Dynamics

Further simulations without NOE-restraints were used to examine the dynamics of the C-T base pair over 1 ns. Snapshots taken over the final 500 ps show that the DNA hairpin conformation has many of the structural features previously described for nogalamycin-d(CCCAC<u>GTA</u>GT^{T}GG), with well-formed loop- and stem-regions. However, the C-T mismatch base pair has an average N3-N3 hydrogen bond distance, calculated over the entire nanosecond of the simulation, which is unusually long (3.4(±0.4) Å). The C-T pair samples two major conformations, both of which have a single H-bond that is significantly populated for a portion of the rMD simulation (Fig.3.48(a);(b)). In the first comformer, a single H-bond if formed between the cytosine amino proton and the oxygen at position 4 of thymine (Fig.3.48(a);(c)). The average N4-O4 distance in this case was measured at 2.9(±0.1) Å. The formation of a this form of the C-T base pair requires a significant narrowing of the minor groove to a C1'-C1' distance of less than 8 Å that, in this case, is partly precluded by helix unwinding and steric interactions with the bound drug. The second of these two conformers (Fig.3.48(b);(d)), which becomes more populated as the simulation progresses, also has a stabilising H-bond between the cytosine amino proton and the oxygen at position 2 of thymine, but does not require any groove narrowing in order to form, as evidenced by an average C1'-C1' distance of 8.6(±0.2) Å, which is typical of B-DNA. The H-bond formed is of similar length to that of the first conformer, at 3.0(±0.2) Å.

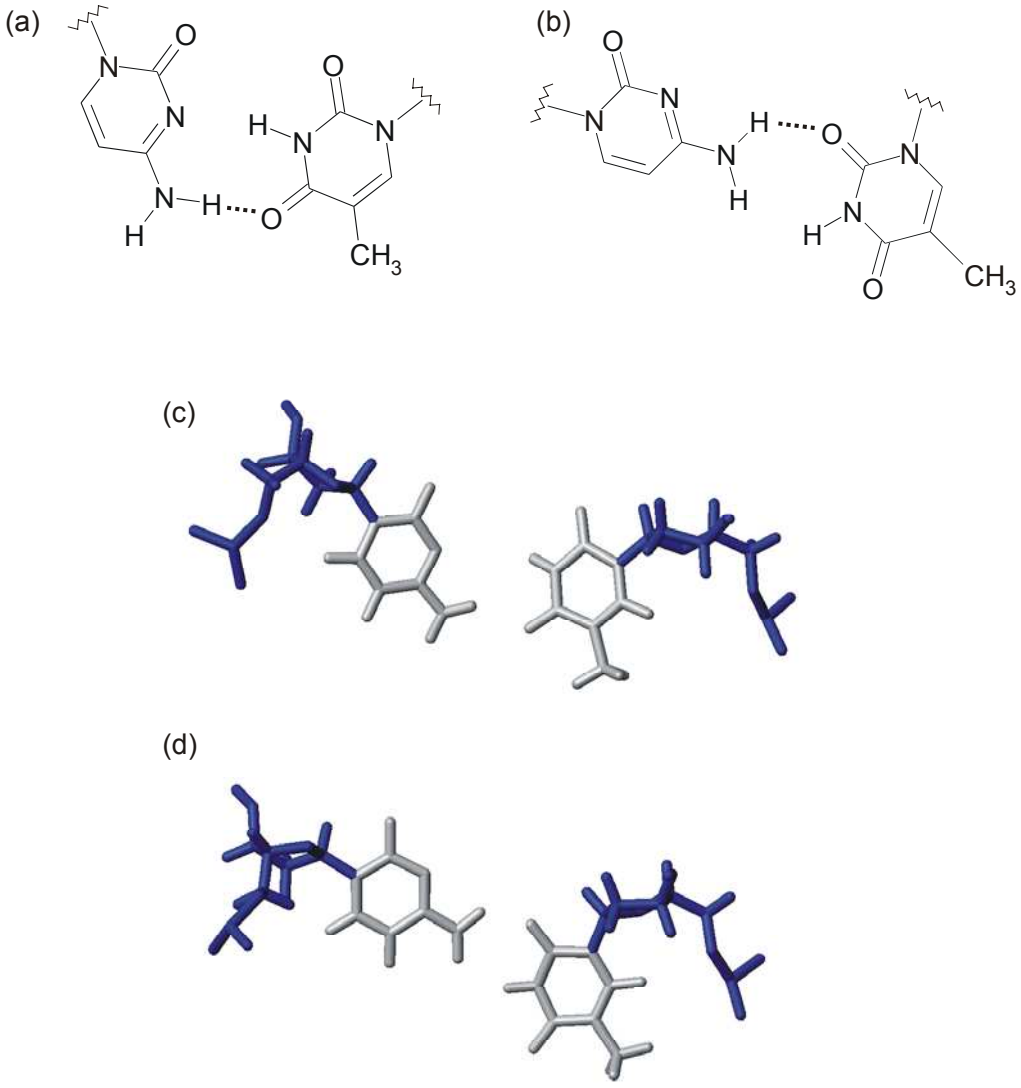


Fig.3.48: C-T mismatch base pair (a);(b), and the two C-T base pair conformations sampled by nogalamycin-d(GCCCGTAGTGC) (c);(d) during a 1 ns rMD simulation.

The autocorrelation plot showed that the structure of nogalamycin-d(GCCCGTAGTGC) had become completely independent of the starting structure after approximately 350 ps of the rMD simulation. Of the 500 structures sampled from the end of the run, the final 400 were shown to be in the second, more open, conformation (Fig.3.48(b)), following a transition from the first conformer during the 599th step of the simulation. Prior to this, step the former (a) form of the C-T base pair was exclusively occupied. Analysis of base-pair opening (Fig.3.49) before, and after, the transition shows an average difference of 11.2º, reflecting the movement of the bases on the XY

plane required to flip between the two conformations. The N4–H···O2 hydrogen bond occupancy, measured using the CARNAL module of the AMBER 6.0 suite, was shown to be 100% for the final 400 ps of rMD. Although both forms of the C-T pair possess a single hydrogen bond, the latter conformation maximises overlap with the flanking base pair and nogalamycin (Fig.3.50). Thus, stacking interactions, rather than hydrogen bonding, appear to be the major stabilising factor for the C-T base pair of nogalamycin-d(GCCCGTAGTGC).

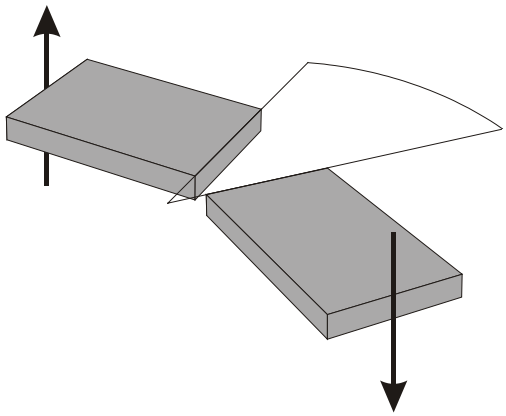


Fig.3.49: Schematic representation of base pair opening (σ).

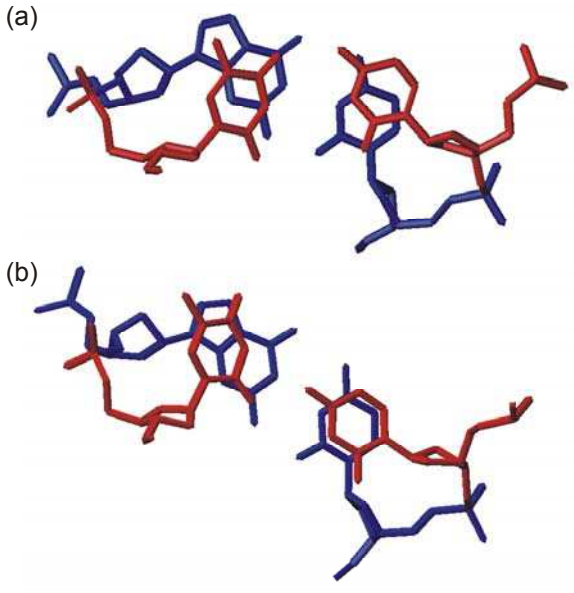


Fig.3.50: Stacking patterns of the C-T mismatch base pair (red) on the flanking C-G base pair (blue), with the mismatch in its (a) hydrogen bonded and (b) non-hydrogen bonded forms, as determined by 1 ns of unrestrained molecular dynamics simulation.

3.8 Conclusions

For nogalamycin-d(GCCCGTAGTGC), a G*N*A hairpin loop structure with B-DNA stem regions, the inherent assumptions associated with creating a starting structure by combining & editing existing structures are reasonable, and produce a starting structure with remarkably similar topology to that generated by simulated annealing using the torsion angle dynamics protocol of DYANA (*87*).

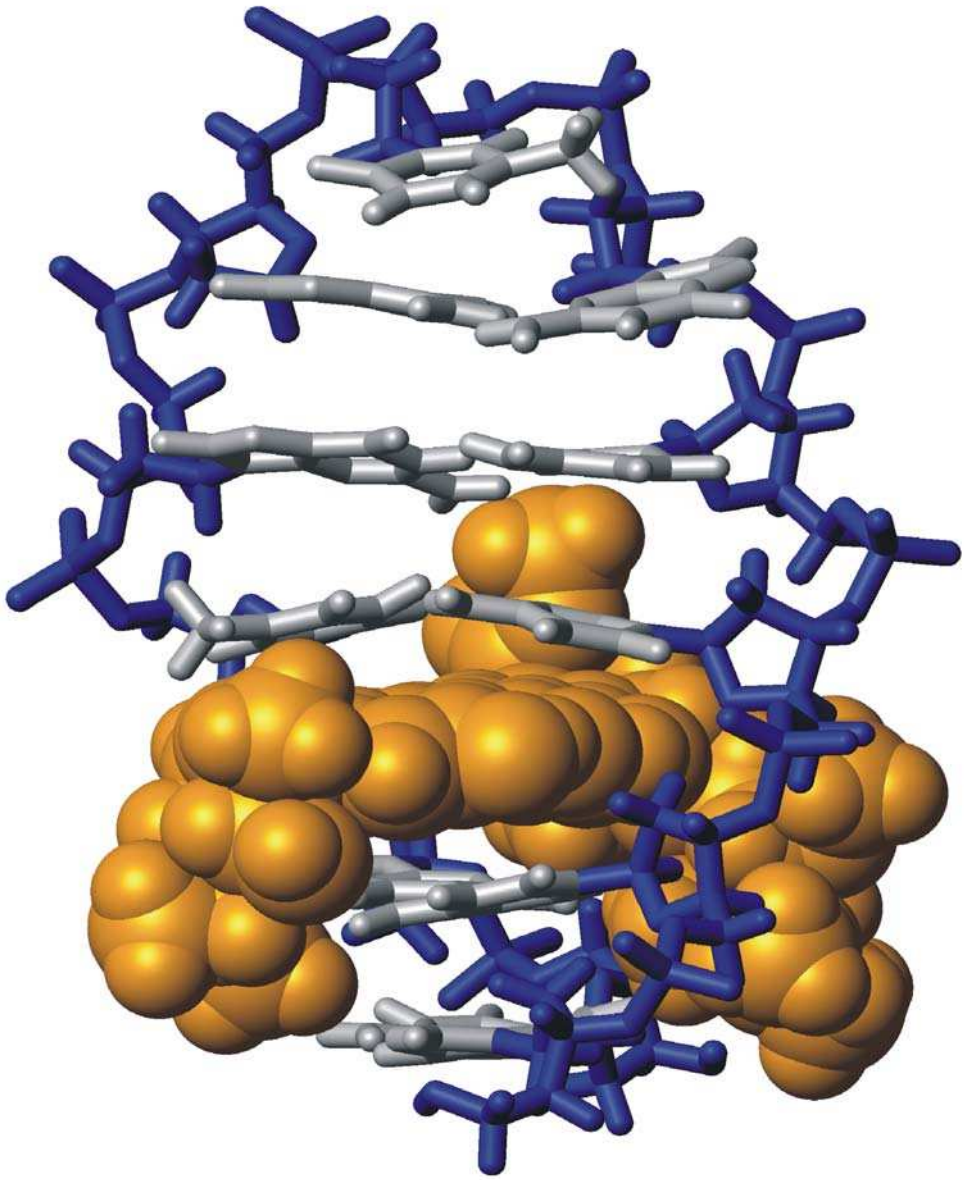


Fig.3.51: Average minimised structure of nogalamycin-d(GCCCGTAGTGC) from the last 100 ps of a 1 ns rMD simulation.

The terminal residues of nogalamycin-d(GTGCGAAGC^{T}AC) are not significantly stabilised by the intercalation of nogalamycin. The groove-lining aminoglucose and nogalose sugars of the drug are oriented towards the already highly-stabilised GAA loop and, as such, have little effect on the frayed terminal region. The desired 3'-5' frameshift at the 5'-terminal of the hairpin, and associated looping-out of the T10 thymine do not occur. Spectroscopic evidence verifies the formation of a T-T *wobble* base pair adjacent to the upper face of the drug chromophore. The T-T pair is estimated to be reasonably stable, as evidenced by sharp T2NH and T10NH peaks on the ^{1}H WATERGATE spectrum at 288 K. The lack of any structure at the terminal ends of nogalamycin-d(GTGCGAAGC^{T}AC) make this complex unsuitable for rMD simulation.

Nogalamycin-d(CCCACGTAGT^{T}GG) was shown to exchange rapidly between at least two conformers at temperatures ranging between 278 K and 318 K. The effects of a bulged thymine at the TpG step proved impossible to characterise in great structural detail, due mainly to the rapid interconversion between two species; one with an extrahelical thymine at the interacalation site, and another with a C-T mismatch base pair adjacent to it. Additionally, the presence of two consecutive thymines, allowed for the formation of both 5'-T^{T}G and 5'-^{T}TG interacalation sites, which have very similar NMR characteristics, and could not be distinguished by the methods employed in this study. Frameshifting of looped-out bases, even when occuring over a single base step, as in this case, can significantly complicate the analysis of otherwise straightforward DNA structures, and must be carefully considered during sequence design, particularly when the structures contain homopolymeric runs of bases. In this case, interconversion between these species precluded any further study into extrahelical thymines on the complementary (5'-TG) face of the nogalamycin binding site, but did intimate the design of a sequence to study a C-T base pair adjacent to the drug.

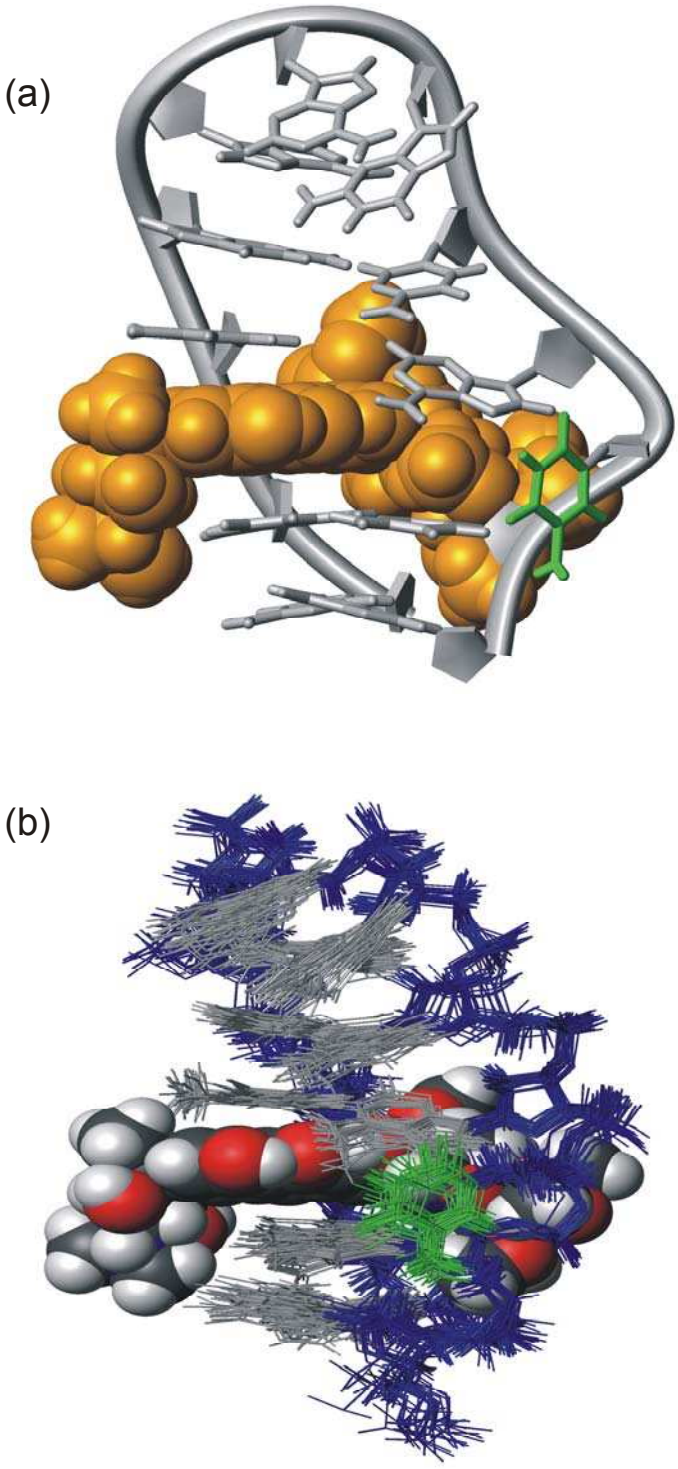


Fig.3.52: (a) Schematic representation of the average minimised structure of nogalamycin-d(GC^{T}ACGAAGTGC) from the last 100 ps of a 1 ns rMD simulation, showing the extrahelical thymine (green) in the major groove. (b) 25 randomly chosen snapshots from the last 250 ps of a 1 ns rMD simulation of nogalamycin-d(GC^{T}ACGAAGTGC).

In this study, the high-resolution solution structures of drug-stabilised hairpin loops have been determined by NMR restrained simulated annealing and Cartesian molecular dynamics methods. The structure of nogalamycin-d(GC^{T}ACGAAGTGC) has been shown to incorporate an extrahelical thymidine, which resides in the major groove and is oriented towards the 5' end of the DNA oligomer (Fig.3.52). In contrast, nogalamycin-d(GCCCGTAGTGC) contains a thymine that is forced to form a mis-match base pair with cytosine. rMD of this structure has shown that the C-T pair can form an *N4*–H···*O2* H-bond, which maximises overlap with ring systems of the flanking bases. Due to the requirement for significant groove narrowing in the formation of a pyrimidine-pyrimidine base pair (*21*) (Fig.3.51), the *N4*–H···*O4* H-bond does not form.

		nogalamycin-d(GC^{T}ACGAAGTGC)	nogalamycin-d(GCCCGTAGTGC)
DYANA			
	drug-DNA	39	18
	DNA-DNA	221	214
	loop	40	32
	H-bonds	11	20
	torsion angles	45	N/A
AMBER			
	drug-DNA	39	18
	DNA-DNA	230	214
	drug-drug	68	36
	H-bonds	22	20
	PRPA	12	11
Restraint Violations			
	>0.3Å	5	3
	>0.5Å	1	0
RMSD			
	heavy atoms	0.73 (±0.1) Å	0.54 (±0.1) Å

Table 3.1: Summary of modelling statistics for the final 100 ps of 1 ns rMD trajectories of nogalamycin-d(GC^{T}ACGAAGTGC) and nogalamycin-d(GCCCGTAGTGC).

4. Structural Studies of an A-Tetrad in Quadruplex DNA

4.1 Introduction

4.1.1 G-tetrads

The phenomenon of unusual, non-Watson-Crick base pairing in guanine-rich oligonucleotides was identified over 40 years ago by Gellert *et al.* (*77*). Guanine residues were found to self-associate into a quadruplex structure, held together by Hoogsteen pairing between the bases (Fig.4.1(a)). The basic unit of the quadruplex was shown to be the G-tetrad, which is formed by a H-bonded network of square, planar aligned guanines. Adjacent guanine bases are hydrogen-bonded on their Watson-Crick and Hoogsteen edges, with their carbonyl groups oriented towards the centre of the quadruplex. One amino proton is involved in a H-bond with the N7 nitrogen of the adjacent base, while the other is solvent exposed. G-tetrad formation requires the presence of monovalent cations (*244*). The location of the stabilising cations has been well defined by a high resolution (0.95Å) crystallographic study of the structure of $d(TG_4T)_4$ (*170*). Sodium ions have been shown to reside between adjacent tetrad planes and to be involved in octahedral coordination with the O6 oxygen atoms of the flanking guanines (Fig.4.1(b)).

It was not until 1988, when Sen and Gilbert (*196*) discovered the self-association of guanine-rich DNA motifs under physiologically feasible ion concentrations, that any biological significance for DNA quadruplexes was hypothesised. Guanine-rich motifs can be found within numerous regions of chromosomal DNA including immunoglobulin switch regions (*20*) and telomeres (*198*), and sequences from these regions have been shown to form quadruplexes. In this review DNA quadruplexes will be discussed with particular attention to their biological significance and methods of study.

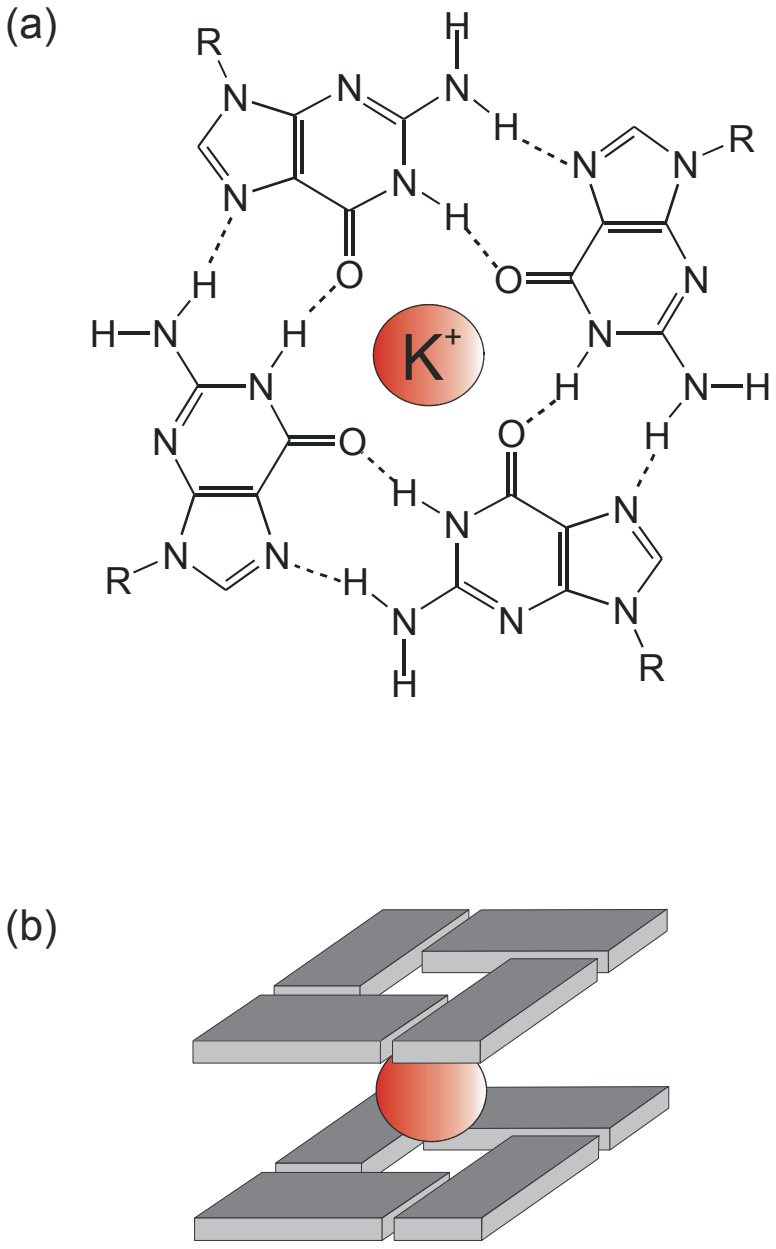


Fig.4.1: (a) G-tetrad structure. The ring is planar; Hoogsteen bonding is indicated by dashed lines. (b) Coordination of monovalent cations midway between tetrad planes.

4.1.2 Quadruplex Structure

DNA quadruplexes can assume a wide array of different structures, usually dependent on the sequence of the component oligomers. In many cases, however, a change in conditions can cause a nucleotide to switch from one quadruplex structure to another (*111;213*). Quadruplexes can be composed of one, two or four strands containing dG_n repeats, in either a parallel or

antiparallel orientation. Quadruplexes containing tetrads of bases other than guanine have also been documented (*59;73;200*). Most recently, a solution structure of $d(TGGTGGC)_4$, demonstrated a stable T-tetrad in the centre of a quadruplex (*166*). Thymine bases at the 5' terminal of parallel quadruplex structures have also been shown to form stable tetraplexes, but not at physiological pH (*76*). The subject of non-guanine tetrad formation, with specific reference to stacking interactions with G-tetrads has been further explored in a comparative NMR study of the solution structures of $d(AGGGT)_4$ and $d(TAGGGT)_4$ in the presence of K^+ (*166*). The thymine residues of the latter sequence were shown to sample a huge range of conformations leading to poor stacking. Only the adenine residues were shown to exhibit good stacking interactions with the adjacent G-tetrad and form stable A-tetrads. The formation of tetrad alignment by residues other than guanine is highly dependent on the overall sequence and their position within it. Where non-guanine or mixed base tetrads have been shown to form, they are invariably stabilised by stacking interactions with flanking G-tetrads. The features of G-quadruplexes will be further explored in the following section, and formation of stable non-guanine tetrads will be discussed presently in Section **4.1.7**.

4.1.3 Four-stranded Quadruplexes

G-quadruplexes composed of four individual strands, each containing one dG_n repeat, are among the best understood and well documented quadruplexes; complex intramolecular quadruplexes found in biological systems are often studied in simplified, intermolecular forms. In all known cases the four strands lie parallel to form a complex of general structure $d(G_n)_4$, as in the case of the hexanucleotide $d(TG_4T)_4$ (Fig.4.2) (*170*). Several four-stranded, parallel quadruplexes have been solved by NMR (*88;136;166*) or X-ray crystallographic methods (*170*), including short sequences derived from telomeric DNA (*166*). Structurally, the backbone parameters of parallel four-stranded DNA quadruplexes are analogous to B-DNA duplexes (*128*) in that

glycosidic torsion angles are all *anti* and sugar puckers are C2'-*endo*. The complementary base-pair parameters of buckle, opening, propellor twist, shear, stagger and stretch are not amenable to comparison with non-Watson-Crick DNA structures (*160*). However, the base step parameters (rise, roll, shift, slide, tilt and twist) of single-stranded quadruplex structures are equivalent to those of B-DNA structures.

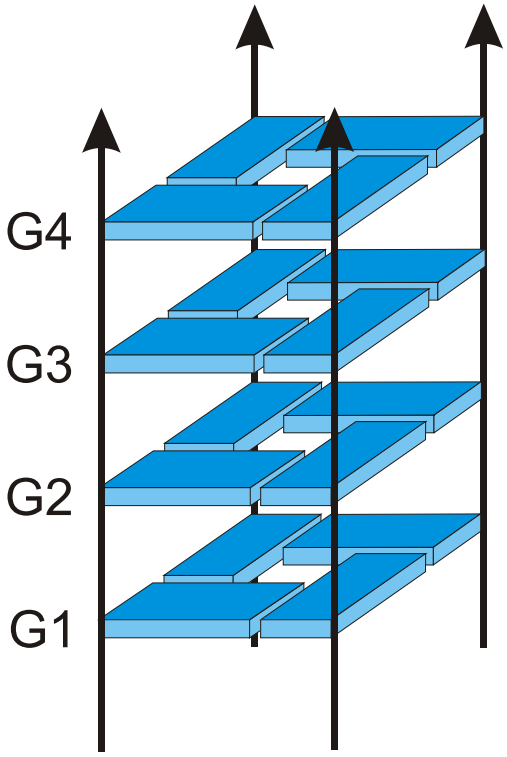


Fig.4.2: Schematic of the X-ray crystal structure of G-tetrads in $d(TG_4T)_4$ (*170*).

4.1.4 Hairpin Dimeric Quadruplexes

Each strand in a two-stranded G-quadruplex generally contains two dG_n repeats connected by a short, usually T-rich, linker sequence to give structures of the type $d(G_nT_nG_n)_2$. The first two-stranded quadruplex to be identified was $d(G_4T_4G_4)_2$, a motif corresponding to the repeat array found in *Oxytrichia* telomeres (*111*). Kang and coworkers studied this structure extensively by X-ray crystallography. The study of this sequence, and the *in vitro* structures it forms, illustrates the profound polymorphism that is characteristic of two-stranded DNA quadruplexes. The quadruplex was found to consist of four tetrads composed of two lateral bound dimeric hairpin loop structures

(Fig.4.3(a)), so that the two T_4 loops are at opposite ends of the molecule and connect strands along the same edge of the complex. All parallel strands have two neighbouring antiparallel strands such that each strands runs in the same direction as that diagonally opposite. Since each base within a tetrad must be similarly oriented, the glycosidic conformation also alternates around the ring within each tetrad, and along the backbone within each strand. Any *anti* conformation on a parallel strand must be matched with a *syn* conformation on an antiparallel strand to facilitate tetrad formation.

However, when Smith et al (*213*) solved the same sequence using NMR the following year, a quite different structure was observed. Their study demonstrated that the quadruplex had a T_4 loop crossing each end of the structure diagonally, so that strands on diagonally opposite vertices share the same phosphodiester backbone (Fig.4.3(b)). Each strand has one parallel and one antiparallel neighbouring strand, and glycosidic conformation follows the pattern *syn-syn-anti-anti* around each tetrad. In this case, the conformation alternates along the backbone of each strand.

These two studies of $d(G_4T_4G_4)_2$ exemplify the two major topologies of two-stranded quadruplexes; structures where the loops run along one face of the complex are commonly known as chair dimers, and those where loops cross over the structure to connect opposite vertices are known as basket dimers. The fact that different structures are observed for the complexes also demonstrates the differing results that NMR and X-ray crystallography studies can produce. Another point of interest, which we will return to later, is the use of different monovalent cations; in the NMR study Na^+ was used as opposed to K^+, which was used to solve the x-ray structure.

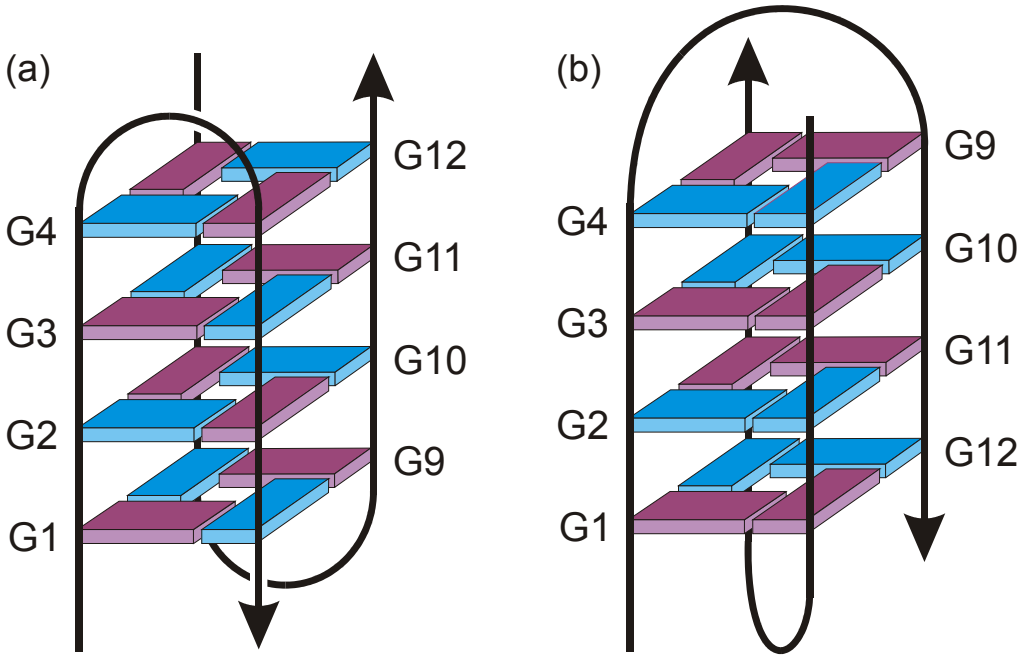


Fig.4.3: Quadruplex structures formed by $[d(G_4T_4G_4)]_2$. (a): crystal structure solved by x-ray crystallography (*111*); (b): solution structure solved by NMR (*213*).

Further polymorphism of two-stranded G-quadruplexes is demonstrated by the solution structure of the complex $[d(G_3T_4G_3)]_2$ (*214*) (Fig.4.4(a)). This complex forms three G-quartets and has a diagonal hairpin loop structure and strand orientation analogous to that found in the solution form of $[d(G_4T_4G_4)]_2$. The complex also shares the *syn-syn-anti-anti* pattern of glycosidic conformation within quartets. However, the pattern along the backbone of strands is quite different. One strand has the sequence *syn-syn-anti*-(T_4)-*syn-anti-anti* whilst the other runs *syn-anti-anti*-(T_4)-*syn-syn-anti*. Thus, although the sequence within the strand is asymmetric, it gives rise to a symmetrical pattern within the folded back structure of the complex. Despite the differences between $d(G_3T_4G_3)_2$ and $d(G_4T_4G_4)_2$, the shared diagonal loop structure suggests a common folding intermediate in the formation of the two quadruplexes (*114*).

Diagonal-looped quadruplexes are favoured at higher strand concentrations when association of the strands occurs first, followed by folding of the loops

(Fig.4.4(b)). Since guanine is the most hydrophobic of the bases, the conformation in which the guanine at the 3' end of the G_4 repeat is stacked over the following thymine is favoured. Due to the requirement for strand reversal, some of the bases must rearrange to the *syn* conformation in order to form the quartet, but the quartet formation is accomplished so that there are fewer 5'-G(*anti*)-G(*syn*)-3' sequential base pairs than either 5'-G(*syn*)-G(*anti*)-3' or 5'-G(*syn*)-G(*syn*)-3' base pairs. It appears difficult for the 5'-G(*anti*)-G(*syn*)-3' stack to achieve stable π-π interactions, and hence it is energetically less stable (*217*). This explains the different pattern of glysosidic conformations between $[d(G_3T_4G_3)]_2$ and $[d(G_4T_4G_4)]_2$. The loop structure is further stabilised by hydrogen bonding between the third thymine and the first thymine in the loop (*115*).

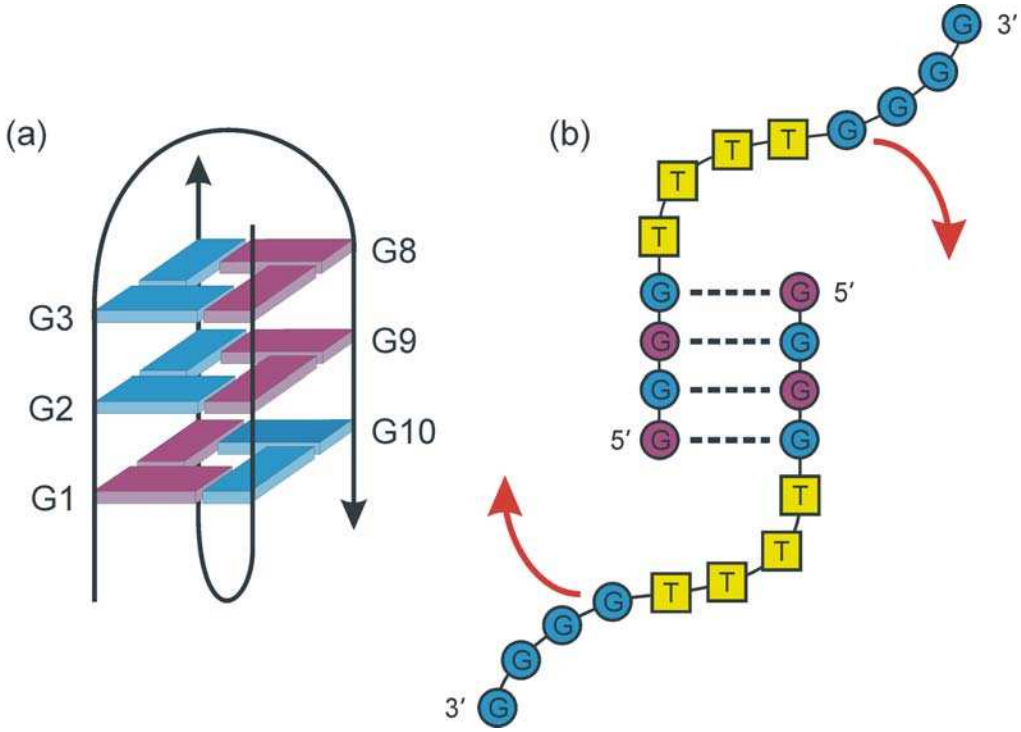


Fig.4.4: (a) Solution structure of $[d(G_3T_4G_3]_2$, solved by NMR (*214*). Blue rectangles indicate guanines in the *anti* conformation; Purple rectangles indicate guanines in the *syn* conformation. Image adapted from (*114*). (b) Association of two strands followed by folding of the two strands about the T_4 segment as exemplified by the diagonal-looped quadruplex, $[d(G_4T_4G_4)]_2$. Image adapted from (*114*).

4.1.5 Single-stranded Quadruplexes

Single-stranded, intramolecular quadruplexes consist of a folded DNA molecule containing four dG_n repeats. The four repeats are connected by three loops, which most commonly form either a chair structure, where all three loops are laterally orientated, or a basket structure, in which the second loop crosses diagonally from one vertex to the other. As an example of a chair structure let us consider the thrombin binding aptamer $d(G_2T_2G_2TGTG_2T_2G_2)$, the solution structure of which was solved by Schultze *et al.* in 1994 (*192*). The folded structure of thrombin binding aptamer contains two guanine quartets, each with a regular *syn-anti-syn-anti* conformation pattern (Fig.4.5). Two T_2 loops at one end of the complex and a TGT loop at the other end link the four strands, which alternate between parallel and antiparallel around the quartet, consistent with the conformation of the individual tetrads. As we can see, the orientation of loop connectivities in quadruplexes is dependent on the sequence and can affect the general folding topology of the structure. Intramolecular G-quadruplexes have been extensively studied in recent years. As we will discuss presently, they are more likely to form at the G-rich 3' overhang of telomeric DNA. Intermolecular two- and four-stranded structures are less biologically relevant, though they are significant *in vivo* in phenomena such as end to end pairing (*143;203*). The human telomeric repeat motif $d(T_2AG_3)_3$ forms a single-stranded quadruplex. It was first studied in solution with Na^+ by NMR techniques (*234*), and found to form a basket structure (Fig.4.6(a)). The quadruplex consists of 3 guanine tetrads joined by three TTA loops; two lateral loops at one end of the structure and a diagonal loop at the opposite end. Nucleotides are arranged in a *syn-syn-anti-anti* conformation pattern within the quartets, and conformation alternates within G_3 repeats, such that each *anti* base stacks onto a *syn* base and vice versa. Each strand lies adjacent to one parallel and one antiparallel strand.

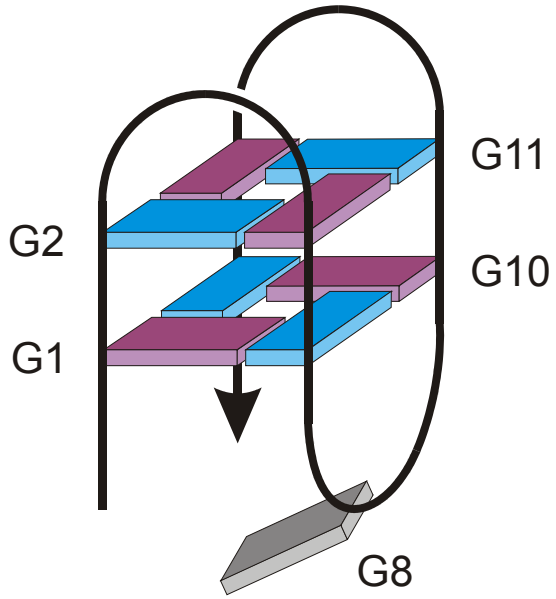


Fig.4.5: Solution structure of d($G_2T_2G_2TGTG_2T_2G_2$) solved by NMR (*192*).

More recent work using x-ray crystallography has identified another structural form of the human telomeric repeat. The crystal structure (*162*) is completely novel, unlike any previously documented quadruplex. All four repeats are parallel, with all glycosidic angles in the *anti* conformation. Repeats are connected by TTA loops extending out from the sides of the complex to form a propeller-like structure (Fig.4.6(b)). The simpler, more regular nature of the structure appears more rational than more complex fold-back structures for the rapid folding and unfolding of telomeric DNA *in vivo*. However, the question of the biological relevance of crystalline structures arises again, despite the use of physiological K^+ concentrations.

Further illustrating the structural diversity of single-stranded G-quadruplexes is the structure formed by the *Tetrahymena* telomeric sequence d$(T_2G_4)_4$. The quadruplex formed contains three G-quartets; strands are linked by two lateral loops: a GTTG loop links the first and second G3 repeat, and at the opposite end of the molecule a TTG loop links the second and third repeat. Finally, a TT loop traverses one of the grooves within the DNA to connect the third and fourth repeat (*234*).

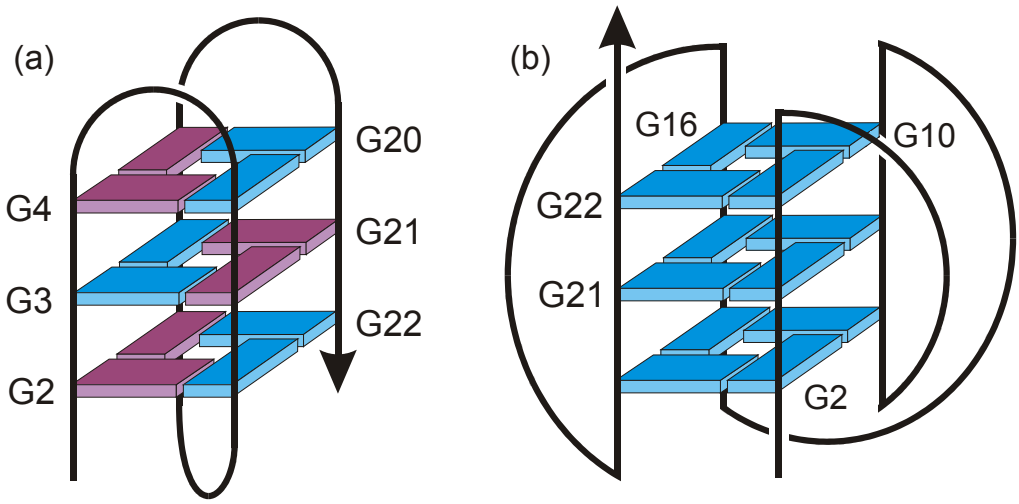


Fig.4.6: Quadruplex structures formed by d[$AG_3(T_2AG_3)_3$]. (a) solution structure solved by NMR (*234*); (b) crystal structure solved by x-ray crystallography (*162*).

Quadruplexes containing bases other than guanine have also been reported. Mixed quartets exist which contain G and at least one other base, such as[$d(G_3CT_4G_3C)_2$, which forms a four-quartet structure where the outer two tetrads each contain two cytosine residues (*116*). Entire non-guanine quartets have also been identified; the *S. cerevisae* telomere motif $d(TG_2TG_2C)_4$ contains a T-tetrad (*166*); the SV40 repeat sequence $d(TG_3CG_2T)_4$ contains a C-quartet (*165*); and an A-quartet has been identified in the human telomeric sequence $[d(AG_3T)]_4$ (*73;167*).

4.1.6 Roles of Counter Ions

Negative charge contributed by the phosphate backbone of DNA mean that it is stabilised by positively charged ions. In the case of quadruplex DNA counter ions are explicitly required for formation. The most abundant physiological cations are Na^+ and K^+ and hence these are most commonly used in the study of quadruplexes *in vitro*. As discussed, monovalent cations bind between stacked G-tetrads with 8-fold coordination to the lone pair of the each *O6* atom. The nature of the coordinating ion can affect the stability of certain types of quadruplex structure. Some sequences will adopt different conformations dependent on ion type and concentration, which suggests that conformational switching *in vivo* may be brought about by localised changes

in intracellular ion concentration. Some general rules for ion requirements in quadruplex formation have been proposed by Patel and coworkers (*164*) based on NMR experiments in the presence of varying K^+ and Na^+ concentration. They have stated that four-stranded parallel quadruplexes have no specific requirement for K^+ or Na^+ and will form provided that at least one of these ions is present at sufficient concentration. However, electrophoretic and UV studies have shown that K^+ is generally more stabilising than Na^+ (*139*). Due to the polymorphic nature of the quadruplex sequence and conformation, there are some cases where these rules do not hold, particularly for quadruplexes involving adenine or cytosine (*117;244*). In four-stranded parallel quadruplexes, K^+ has been shown to coordinate between every pair of adjacent G-tetrads. Binding can take place at other sites, such as at the ends of stacks, or between G- and non-G-tetrads. Ion binding at these sites is largely dependent on the surrounding sequence (*73;139;164;167*).

4.1.7 A-tetrads

As previously stated, there is a limited potential for non-guanine tetrads to form in DNA quadruplexes. These structures are, however, liable to be much less stable than G-tetrads, and are more likely to be found flanked on either or both sides by their more stable relatives than in guanine-free quadruplex sequences. The alternative tetrad structure with the greatest potential to form is the A-tetrad, as these have the greatest potential to form stabilising hydrogen bonds in the core of the quadruplex (Fig.4.7). Unlike their guanine containing cousins however, A-tetrads have little potential to bind potassium ions in their central channel, as they lack the electron rich *O6* atom, or any substitute in their purine ring structure.

Octahedrally-coordinated cations between adjacent G-tetrads must first be displaced in order for a drug molecule to intercalate. Since A-tetrads lack the requisite electronegative functional groups, intercalated ions may be more easily dislplaced, or absent. As such, A-tetrads are of particular interest in the

context of drug stabilisation of quadruplexes. As will be discussed further in Section **4.1.10**, cellular DNA must be converted from its paranemic form to unstructured single strands in order to be enzymatically activated and so drugs that stabilise the paranemic form can inactivate gene.expression

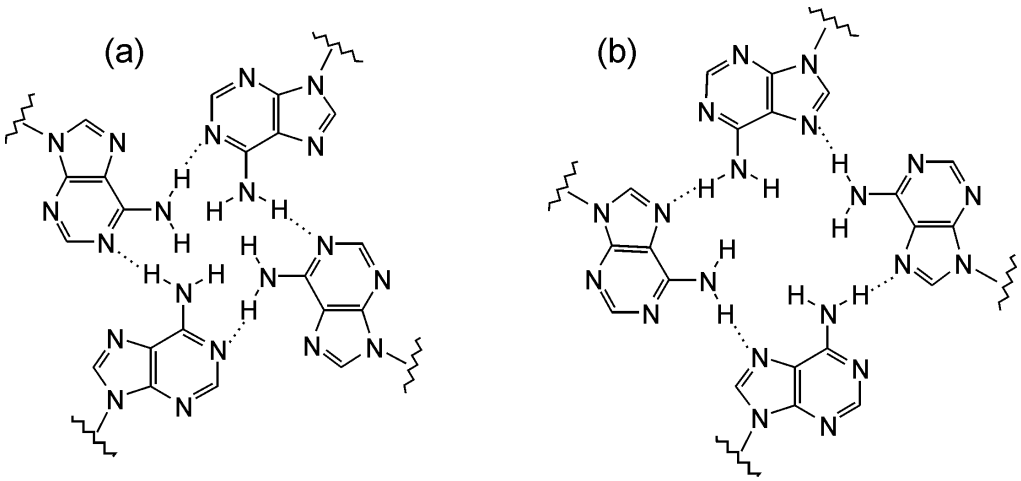


Fig.4.7: The two possible forms of A-tetrad alignment; with N6–H···N1 (a) and N6–H···N7 (b) hydrogen bonding patterns.

A previous study of drug recognition and stabilisation of the parallel-stranded quadruplex $d(TTAGGGT)_4$, determined the solution structure of a 2:1 complex of an intercalating drug with the oligomer (*73*). They observed that the drug molecules bound on the ends of the G-quadruplex, at the ApG and GpT steps, rationalising that the stabilising effects of (potassium) ions within the central channel precluded intercalation at a GpG step because the energetic cost of displacing a potassium ion is not sufficiently compensated by stacking interactions with the drug.

Patel & co-workers (*167*) have described NMR structural studies on two truncated human telomeric DNA sequences, $d(AG_3T)_4$ and $d(TAG_3T)_4$ in solutions containing K^+ ions. The G_3 stretches in both the oligonucleotides were seen to form parallel-stranded quadruplexes. Their report included the formation and characterization of a novel K^+-stabilized $d(AG_3)_4$ quadruplex in the sequence $d(AG_3T)_4$. This is a parallel-stranded quadruplex in which an A-

tetrad is formed at the 5' end. Such a feature was not seen in the $d(TAG_3T)_4$ sequence where the AG_3 stretch was flanked by T nucleotides at the 5' end. They observed that the G_3 parts had overall similar characteristics in both the molecules, but in $d(AG_3T)_4$, the adenines had *syn* glycosidic conformations and the A-tetrad stacked well over the adjacent G-tetrad. The data also indicated dynamism in the alignments of adenine in the A-tetrad over 100 ps of rMD simulation.

At the end of the above calculations for $d(AG_3T)_4$ quadruplex, the final structures were seen to be distributed between two distinct patterns of A alignment, with hydrogen bonds formed between the *N6* amino nitrogen of one adenine and either the *N1* or *N7* nitrogen of an adjacent adenine, consistent with Fig.4.7. A rapid equilibrium between these N61 (Fig.4.7(a)) and N67 (Fig.4.7(b)) models has been postulated, as only one resonance was observed for each proton. The exchange rate, thus, must be fast compared to the chemical shift differences between the two models, but slower than the NOE time scales, and was estimated in the study to be on the nanosecond timescale.

Gavathiotis & Searle (*73*) have also resolved a $d(AG_3T)_4$-containing quadruplex by NMR methods as part a study of the human telomeric repeat. In this case, the A-tetrad is flanked on its 5' face by two T-tetrads, which do not form any recognisable hydrogen bonding alignment during the 1 ns of restrained MD simulation they describe. The A-tetrads in this model had *anti* glycosidic conformations, and, again, stacked well over the adjacent G-tetrad. The NOE assignment, in this case, was consistent with the formation of an N61-aligned A-tetrad (Fig.4.7(a)). However, the RMSD of hydrogen bond distances was greater for the A-tetrads than G-tetrads, attributed to the proximity of the As to the amorphous 5'-end of the molecule.

These previous studies have failed to unambiguously resolve the conformation, N61 or N67, preferred by A-tetrads incorporated into DNA quadruplexes. Additionally, little effort has been made to address the exact number of potassium ions bound to these adenine-containing structures, though in the latter study, K^+ ions bind between G-tetrads consistent with existing stochiometric studies (*233*).

4.1.8 Telomeres

Telomeres are the specialised nucleoprotein structures found at the end of eukaryotic chromosomes, which are essential for chromosome stability. Telomeric sequences are characterised by tandem repeats of short, guanine-rich DNA. The exact sequence of the repeat is species-dependent. Examples of such repeat sequences are TTGGGG, TTTTGGGG, TTTAGGG, TTTTAGGG and TTAGGG, which are found in the telomeric regions of *Tetrahymena*, *Oxytrichia*, *Arabidopsis*, *Chlamydomona* and human DNA, respectively (*254*). Early genetic studies found that chromosomes with broken ends fused to produce unstable structures such as circular chromosomes (*19*). Subsequent molecular biological studies identified another crucial role: the prevention of chromosome shortening on the 5' strand with successive rounds of semi-conservative replication. Further research has also identified telomeres as involved in transcriptional silencing, recombination events, and nuclear organisation of chromosomes (*180*).

In terms of sequence and structure, telomeres are highly conserved throughout even the most distantly related eukaryotes. Telomeres consist of short, G-rich, tandemly repeated motifs, paired with C-rich repeats on the complementary strand (Fig.4.8). A section of G-rich strand overhangs the C-rich strand, running 5' to 3' with respect to the chromosome end. The length of telomeres varies widely throughout eukaryotes, from a few tens of base pairs in *Oxytrichia* to 300-450 bp in yeast and between 1.5 and 150 kb in vertebrates (*180*). Furthermore, the length by which the G-strand overhangs also varies.

In ciliated protozoa the G-strand overhangs by 12-16 nucleotides, compared with up to 300 nucleotides in yeast (*237*).

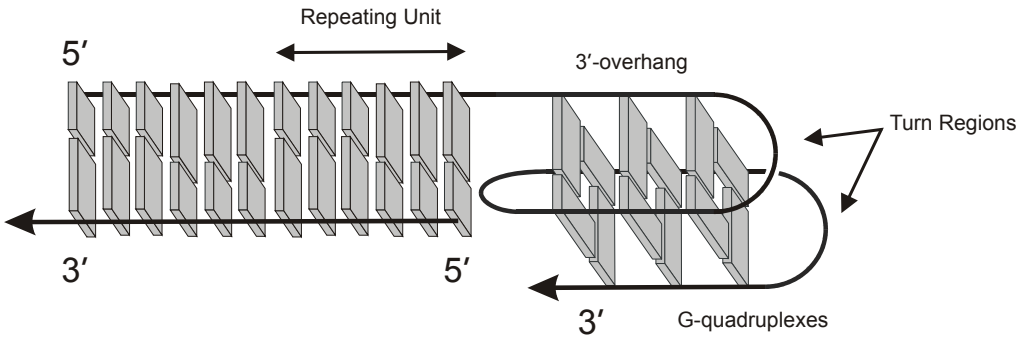


Fig.4.8: Schematic representation of the telomere.

In vertebrates, the majority of telomeric DNA is organised into nucleosome-like structures analogous to packaging of the majority of chromosomal DNA by histones (*138;223*), but with the significant difference that some histone proteins are absent from packaged telomeric DNA. Rat hepatocyte telomeric DNA or chromatin consists of a structure that repeats every 157bp, as opposed to 200bp in bulk chromatin. This is due to the absence of the linker histone H1 (*138*). However, chicken erythrocytes do contain a linker histone; usually H5 as this has a higher binding affinity for telomeric DNA than H1 (*153*). Yeast telomere organisation depends primarily on the protein *Rap1p*, which is essential for regulation of telomere length, as well as stimulation and repression of transcription (*199*). In lower eukaryotes no telomeric DNA organisation or packaging appears to be present (*138*).

4.1.9 Quadruplexes *in vivo*

A quadruplex containing a telomeric DNA sequence was first reported in 1989, when the *Tetrahymena* telomeric DNA motif $d(T_2G_4)_2$ was found to form a two-stranded fold-back quadruplex (*220*). Over the next decade many different DNA sequences, from a wide range of eukaryotic organisms, were found to form stable quadruplex structures *in vitro* under physiological ion concentrations. These studies have presented convincing evidence for

quadruplex formation *in vivo*. Proteins have been identified in lower eukaryotes (*61;79*) and vertebrates (*58*), which bind to DNA quadruplexes and in some cases stimulate quadruplex formation. Recently, binding of quadruplex-specific antibodies to telomeric DNA in *Stylonychia lemnae* macronuclei has been detected using immunofluorescence (*187*), providing the first direct evidence of quadruplex formation *in vivo.*

So far the most compelling evidence of quadruplex formation in human cells is the identification of two different quadruplex structures in the Nuclease Hypersensitivity Element III_1, which forms part of the promoter of the *c-myc* oncogene (*200*). Identification of quadruplexes at human cell telomeres is hindered by low concentrations of telomeric DNA. In contrast, macronuclei of ciliates contain gene-sized chromosomes each with telomeric ends, providing a high concentration of telomeric DNA relative to total nucleic acid concentration, which is well-suited to the study of telomeres (*176*).

4.1.10 Biological Roles

With the evidence for *in vivo* existence of telomeric quadruplexes becoming more compelling, their possible function or functions must be addressed. G-quadruplex formation may be important in stabilising the 3' overhang of telomeres (*220*) during cell quiescence, with rapid protein-mediated unfolding occurring during the replicative stage of the cell cycle. Immunofluorescence studies of *Stylonychia lemnae* macronuclei support the theory that quadruplexes at telomeres are disassembled during replication (*187*). It has been suggested that quadruplexes are present at telomeres as part of higher order structures, in a form that can be assembled and disassembled rapidly (*162*).

In somatic cells, telomeres shorten with each round of semi-conservative replication; typically by 100bp per cell cycle. Below a certain telomere length the cell becomes senescent, thus restricting the cell to a finite number of

replications (*91*). In cancer cells, telomere length is shorter than normal, but is maintained by the enzyme telomerase, which is inactive in normal somatic cells but activated in 85-90% of all human cancers (*120*). Inhibition of telomere maintenance in cancer cells leads to telomere shortening, lack of cell proliferation and apoptosis (*255*). Quadruplex DNA cannot be recognised by telomerase; disasembly of the telomere to unwound single-stranded DNA is required for telomere extension (*253*). Inhibition of telomerase via G-quadruplex-binding ligands presents an important target for development of anti-cancer drugs.

G-quadruplex forming sequences are also found in several transcriptional regulatory regions of important oncogenes, including c-*MYC*, c-*MYB*, c-*FOS* and c-*ABL* (*205*). Because of the polypurine–polypyrimidine nature of these duplex sequences, which contain four or more runs of clusters of three or more guanines on the purine-rich strand, they often show a single-stranded character and hence are hypersensitive to nucleases. Expression of the c-*MYC* oncogene is linked to potentiation of cellular proliferation and to inhibition of differentiation, leading to its association with a number of human and animal malignancies, including carcinomas of the breast, colon, and cervix, as well as small-cell lung cancer, osteosarcomas, glioblastomas, and myeloid leukemias (*60;140;168;215*).

The nuclease hypersensitivity element III_1 upstream of the P1 promoter of c-*MYC* controls 85-90% of the transcriptional activation of this gene (*175;186;224*). In the c-*MYC* promoter, the purine- and pyrimidine-rich strands bind transcription factors (cellular nucleic-acid-binding protein, CNBP, and heterogeneous nuclear ribonucleoprotein, hnRNP) that are required for transcriptional activation. As these elements can also potentially form G-quadruplex and i-motif structures (*204;206*), it is possible that the secondary DNA structures inactivate transcription, and their conversion to the duplex form is required for transcriptional activation. Hurley *et al*. (*201*) have

demonstrated that the proposed chair G-quadruplex, and possibly an i-motif structure formed in the complementary strand, that can form in the c-MYC promoter region need to be converted to unstructured purine and pyrimidine single-stranded forms before c-*MYC* can be transcriptionally activated (Fig.4.9). Stabilised G-quadruplex structures that would prevent this transition would therefore inactivate c-*MYC* expression. As downregulation of c-*MYC* expression in tumour cells by only 30% leads to a marked reduction in their transformation by RAS or RAF (*14*), the selective stabilisation of G-quadruplex structures in the promoter regions of c-*MYC* should lead to selective effects on cancer cells.

To date, no structure of the paranemic form of the NHE III_1 region of the c-*MYC* oncogene has been determined by NMR or X-ray crystallographic methods. Although the proposed structure of the quadruplex is an intramolecular complex stabilised by G-tetrads, this is based solely on DMS footprinting studies (*201*). The proposed structure also contains one heterogenous tetrad. Furthermore, this region of the c-*MYC* ongogene contains several d(GGAGG) repeats along its 27 base-pair length, which, as we have discussed in Section **4.1.7**, can potentially form A-tetrads stabilised by flanking G-tetrads.

A number of molecules have been identified that bind to and stabilise G-quadruplexes formed from telomeric DNA sequences *in vitro* (*62;178;219;238*) and *in vivo* (*81*). These are aromatic compounds, which are frequently heterocyclic in nature, and carry an overall positive charge. G-quadruplex stabilising molecules have been shown to bind DNA through intercalation at both ends of, but not within, a stack of tetrads. We have stated that current ligands cannot penetrate a quadruplex and bind between two G-tetrads is most likely due to their inability to displace counterions normally present between the tetrads (Section **4.1.1**). In this context, incorporation of non-guanine bases into tetrad structures could greatly expand the biological

function and diversity of quadruplex forming structures, particularly from the point of view of stabilising and destabilising these structures by pharmacological intervention. The potential of drug molecules to stabilise non-guanine tetrads, which would not normally form, warrants a structural study into these structures, particularly A-tetrads, which have the greatest potential to form stable quadruplexes.

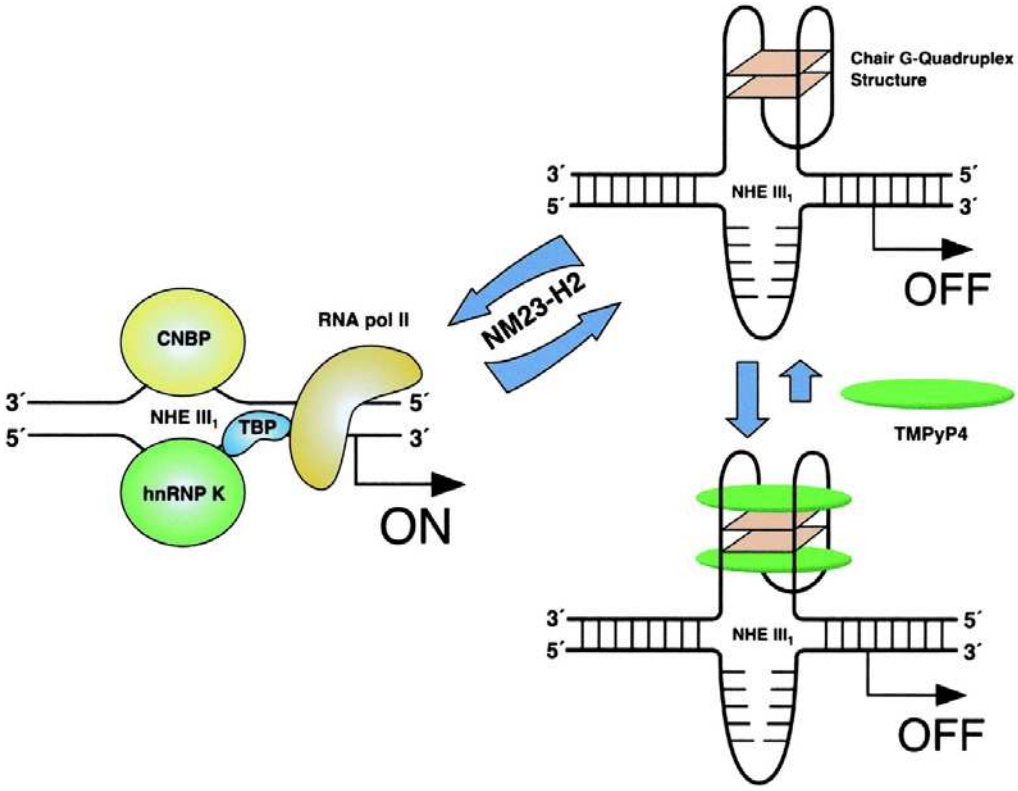


Fig.4.9: Model for the activation and repression of gene transcription involving the accessory role of NM23-H2 in interconversion of the unstructured purine and pyrimidine single-stranded DNA forms to the paranemic secondary DNA structures. Interaction of the G-quadruplex structure with the drug molecule TMPyP4 stabilises the gene-off form by inhibition of conversion to the single-stranded gene-on forms. Reproduced from (*201*).

4.2 Experimental Design

We have carried out a careful NMR/MD approach to determine the details of a family of quadruplex structures, each containing a $d(GGAGG)_4$ core.

Intermolecular G-quadruplexes may have less biological significance than intramolecular G-quadruplexes, which can be formed at the 3' overhang of telomeric DNA, but are likely to play a role *in vivo* in phenomena such as recombination and end to end pairing. In addition to the strong interest in the structure and stability of G-quadruplexes, there is also the need to explore the role of guanine and other neighbouring residues in the G-quadruplex structure *vis-à-vis* stability and how adjacent residues are perceived in molecular recognition by proteins and small ligands.

Previous studies of A-tetrads (Section **4.1.7**) have been described in terms of the human telomeric repeat, $d(TTAGGG)_n$. The 3' overhang of telomeric DNA is believed to form an intramolecular quadruplex in which the adenines do not form a tetrad, but rather form part of the connecting loops (*162;234*). However, the c-*MYC* promoter region contains d(GGAGG) repeats, which could potentially assemble to form A-tetrads. For this reason, it was decided to study sequences containing this repeat within the purine-rich sequence. We have, however, endeavoured to keep this study relevant to those that have gone before, while attempting to maximise the potential for formation of a stable A-tetrad.

The initial sequence $d(TAGGAGGT)_4$ was chosen as it contained two A-tetrads, one sandwiched between two sets of G-tetrads, and another at the end of a run of guanines, analogous to those studied previously. It was hoped that this could provide an insight into how the flanking sequence affected the conformation of an A-tetrad. This structure, however, was a mixture of one minor & one major species, and so a second sequence, $d(TGGAGGC)_4$ was designed in order to concentrate on the more stable sandwiched A-tetrad. A cytosine residue was incorporated into this second sequence to provide an additional reference for assignment purposes, and from which to derive distance restraints using the relaxation matrix approach. Also, to remove the

symmetry of the purine-rich sequence, an extra guanine residue was added on the 5' side of the A-tetrad, giving rise to the sequence d(TGGGAGGC)$_4$.

4.3 Materials & Methods

4.3.1 Synthesis & Purification

The oligonucleotides were synthesised and purified as described in Section **2.3.1**, and were shown to be > 95% pure and in the quadruplex form by ^{1}H NMR spectroscopy. The oligonucleotide sample was quantified spectrophotometrically. The NMR samples were prepared in the single strand concentration range of 5.0-8.0 mM (1.3-2.0 mM in quadruplex) in 600 µl of mixed solvent, having a final salt concentration of 100 mM KCl and 10 mM K_2HPO_4, for WATERGATE experiments. The quadruplexes were lyophilised and redissolved in D_2O solution (600 µl) for the observation of non-exchangeable proton resonances.

4.3.2 NMR Analysis

NMR experiments were performed on a Bruker Avance™600 spectrometer. All one-dimensional NMR experiments were recorded over 32768 data points in the f2 dimension, with spectral width of 16.022 ppm and a delay time of 2 s. Phase-sensitive Double Quantum Filtered COrrelated SpectroscopY (DQF-COSY) (*171;177*), TOtal Correlated SpectroscopY (TOCSY) (*11*), and NOE SpectroscopY (NOESY) (*12*) experiments were performed, collecting either 1024 or 2048 points in f2, and between 400 and 512 points in f1. NOESY data at mixing times ranging from 75 ms to 300 ms were collected with a spectral width of 16.022 ppm in each dimension for D_2O experiments, and 20.028 ppm for H_2O experiments, each with a delay time of 2 s. Quadrature detection in f1 was achieved using time-proportional phase incrementation (TPPI). The solvent signal was suppressed by pre-saturation or using the WATERGATE pulse sequence (*173*). WATERGATE NOESY spectra at 300 ms, 150 ms, 100 ms and 75 ms mixing times were acquired using samples dissolved in 90% H_2O and 10% D_2O. All TOCSY experiments employed a

spin-locking field of 7 kHz. Two-dimensional data were zero-filled to 2048 x 1024 points prior to Fourier transformation, optimised with a shifted sine squared function in both dimensions, and treated with automatic baseline correction.

4.3.3 Distance Restraints

Interproton distances were derived from integration of NOE cross peak volumes from NOESY WATERGATE experiments with mixing times of 150 ms, 100 ms, and 75 ms (90% H_2O, 10% D_2O). Distances were determined by extrapolation to zero mixing time as before (Section 2.3.4). For $d(TAGGAGGT)_4$, the thymine H6-Me fixed distance (3.0 Å) was used a reference distance for all NOEs involving base protons, and the sugar H2'-H2" fixed distance (1.85 Å) for NOEs involving only sugar protons. For $d(TGGAGGC)_4$ and $d(TGGGAGGC)_4$, the cytosine H5-H6 distance (2.45 Å) was applied to all distances involving base protons, except those involving methyls, where the H6-Me distance was again employed as a reference.

For well resolved non-exchangeable cross peaks the distances were given upper and lower error bounds of 10% of the calculated distance. Partially overlapped non-exchangeable cross peaks distances were given upper and lower error bounds of 20%. Due to the relative broadness of their exchangeable crosspeaks, only non-exchangables were used for distance restraint derivation for $d(TAGGAGGT)_4$ and $d(TGGGAGGC)_4$. For $d(TGGAGGC)_4$, where exchangeable crosspeaks were considered, the resulting distances were given upper and lower error bounds of 25%. As well as their broadness relative to their non-exchangable relatives, the solvent system used for WATERGATE experiments means that the maximum theoretical intensity of exchangeable crosspeaks is 90% of that observed for non-exchangables. H-bond restraints were included for atoms involved in the ideal hydrogen-bonding geometry of the G-tetrad. The heavy atom-heavy

atom distance restraints for O6–N6 and N7–N7 distances were set to 2.85 ± 0.10 Å.

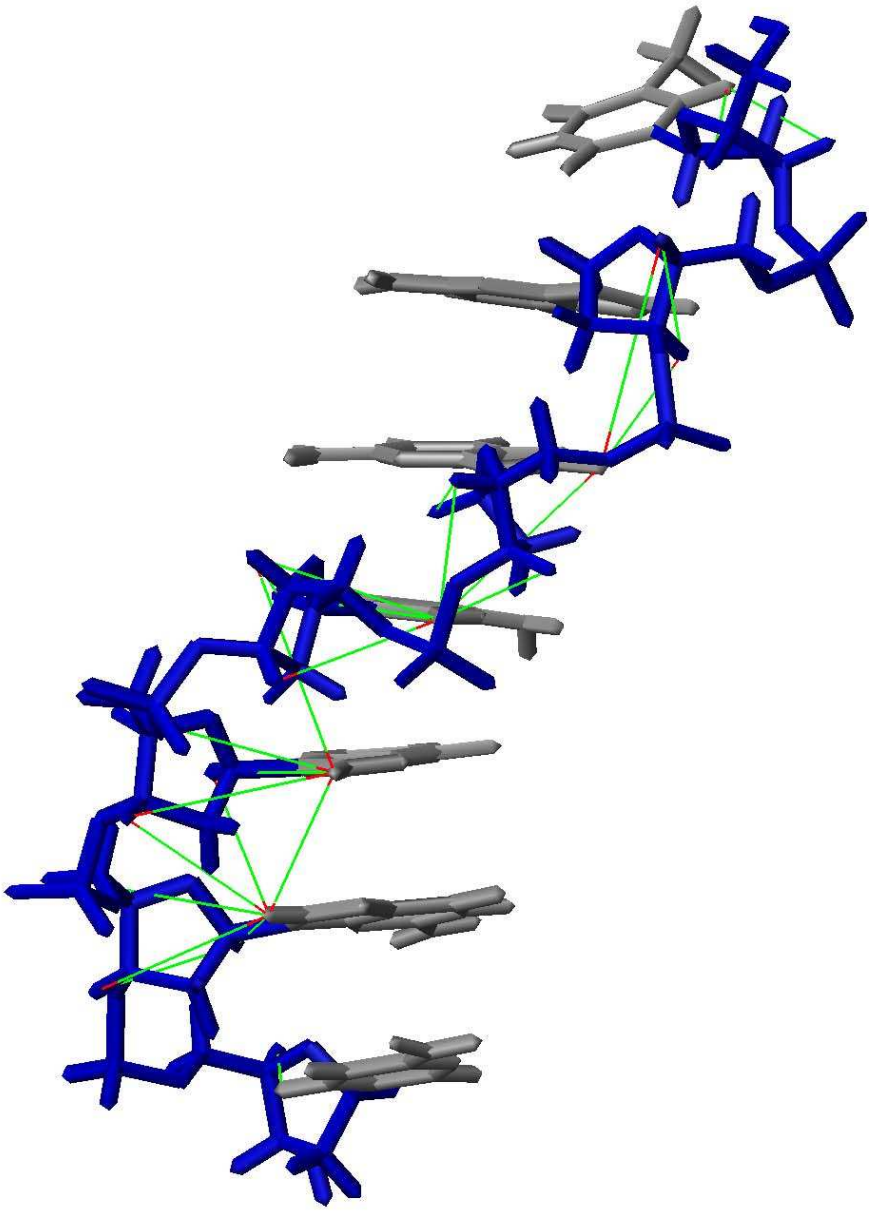


Fig.4.10: Selected distance restraints (green) and associated violations (red) for one strand of the energy minimised structure of d(TGGACCG)$_4$. Displayed using MolMol.

Distance restraints were checked for large geometrical inconsistencies by comparing visually with the distances of the energy-minimised quadruplex structure derived from unrestrained molecular dynamics simulation using MolMol (Fig.4.10). In order to distinguish between interstrand and intrastrand distance restraints resulting from the symmetry of the quadruplex structure, the energy-minimised structure was examined using MolMol. Interstrand restraints were manually determined by matching significantly violated intrastrand distances with their equivalent symmetry translated interstrand distances from the initial energy-minimised structure.

4.3.4 Starting Structures

The initial model for the quadruplex structures $d(TGGAGGC)_4$ was generated using the LEaP module of AMBER 6, and representations of $d(TGGGAGGC)_4$ and $d(TAGGAGGT)_4$ were constructed by editing the PDB file of the parent structure.

G- and A-tetrads were made by aligning four nucleotide bases, each in the *anti* glycosidic conformation, symmetrically in a plane to form the appropriate H-bonds. The A-tetrads were aligned in the N61 pattern (*167*) with the amino hydrogens in H-bonding distance range from the N1 receptor nitrogens. Since no stable H-bonding pattern has been documented for either C- or T-tetrads, these were constructed by arbitrarily aligning the bases symmetrically in a plane, having overlaid the heavy atoms of the sugars on those of a G-tetrad to allow easier construction of the quadruplex. The quadruplex itself was made by stacking these constructed tetrads on top of each other in the desired order, with a separation of 3.4 Å between consecutive planes and a 36º twist, consistent with B-form DNA (*52*). Dephosphorylation of the 5' ends of each strand and capping of 3' ends were achieved automatically when the quadruplex coordinates were read into LEaP.

An initial model of $d(TGGGAGGC)_4$ was generated by reading two copies of the coordinates of $d(TGGAGGC)_4$ into MOLMOL and overlaying them with a one-residue stagger (Fig.4.11). This was achieved by superimposing the deoxyribose heavy atoms, C1'-C5' and O4', of the overlapping nucleotides. The RMSD to the mean averaged over both molecules for the selected atoms was 1.02 Å. Residues 1T and 2G were removed from each strand of one quadruplex, and residues 4A to 7C removed from each strand of the other. The two PDB files were then combined and edited, and the residues renumbered to give the required starting structure.

RMSD fitting of individual A- and T-tetrads to $d(TGGGAGGC)_4$ and subsequent editing of the PDB file was also used to generate the structure of $d(TAGGAGGT)_4$.

Subsequently, each system was treated as follows: potassium ions were inserted between the planes of each tetrad (separated by 3.4 Å). Further potassium counter-ions were then used to negate the negative charge on the backbone, and the model was solvated to within a distance of 5 Å using boxes of 216 TIP3P H_2O molecules. The periodic box of the $d(TAGGAGGT)_4$ system measured 47x49x56 Å and contained 2537 water molecules. The 2196 H_2O molecules surrounding $d(TGGGAGGC)_4$ were enclosed in a periodic box measuring 50x47x50 Å. The periodic box of $d(TGGAGGC)_4$ was significantly smaller than the others, at 48x44x49 Å, reflecting the shorter sequence length, and housed 1961 TIP3P waters. In each case, the solvated systems were energy minimised at constant pressure using the same procedure employed for the duplex and hairpin models.

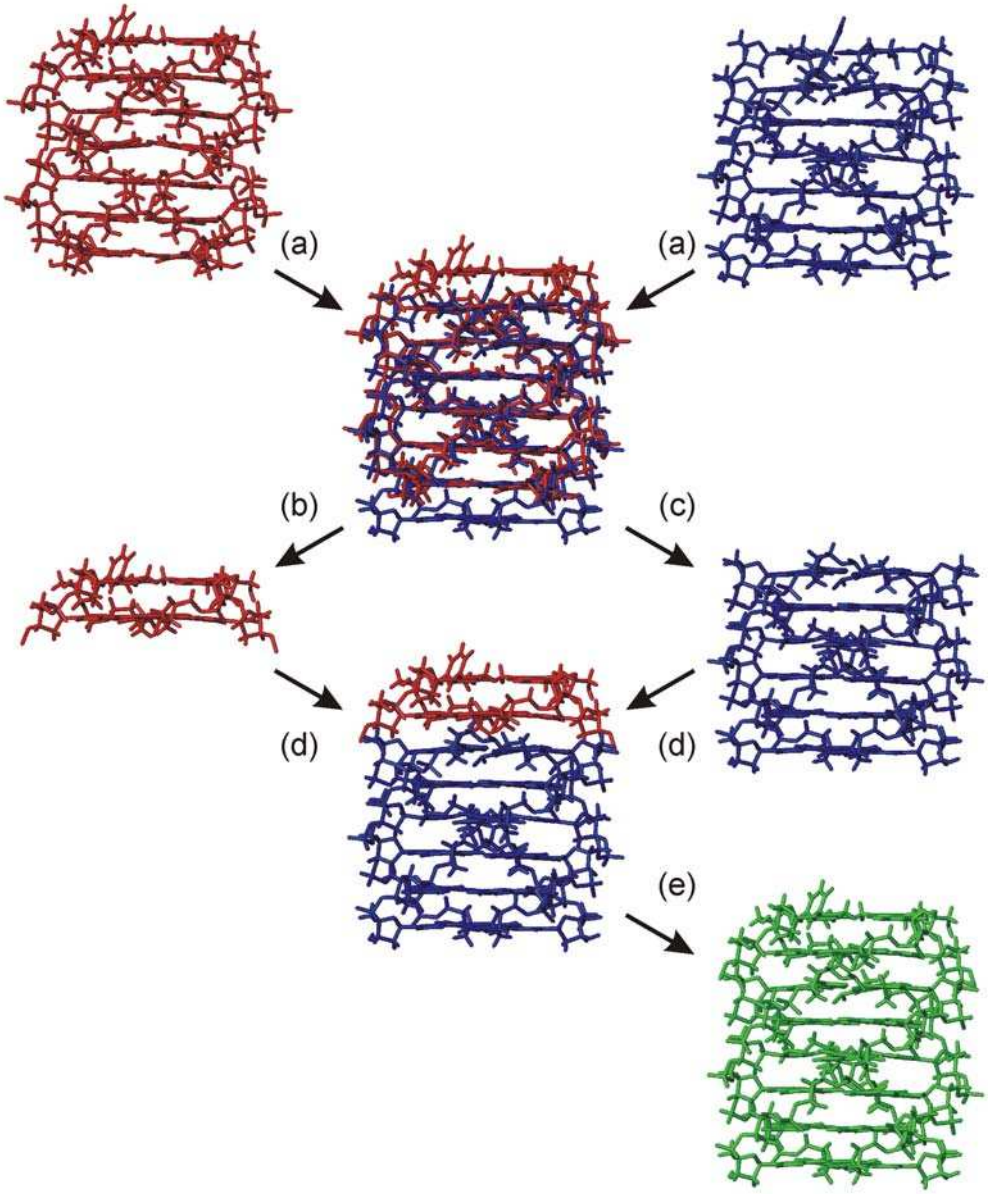


Fig.4.11: Steps involved in the generation of an initial model for d(TGGGAGGC)$_4$. Two copies of d(TGGAGGC)$_4$ were overlaid (a); the top two stacks (d(TG)$_4$) were selected from one copy (b), and the bottom four stacks (d(GGAGGC)$_4$) selected from the other (c). The selected residues were then combined into one structure (e).

4.3.5 Structure Calculations

Minimisation was followed by ten iterations of molecular dynamic simulation. Each of the first nine rounds of MD comprised 10 ps of simulation, split into 5000 steps of 0.002 ps duration. During the opening three iterations, the restraining field of 100 kcal Å^{-2} mol was applied only to the DNA and

counter-ions, while the temperature of the system was increased from 100K to 300K. At this temperature, the potassium ions were released from the restraining field and another round of simulation was run. Over the next five iterations the restraining field was reduced to 50, 20, 10, 5 and 2 kcal $Å^{-2}$mol. The final round of MD was made up of 50000 steps, totalling 100 ps, during which the entire system was unrestrained.

Following minimisation and unrestrained MD simulations, NMR restraints were introduced to the equilibrated structure and the system allowed to undergo 2 ns of restrained MD to satisfy the NMR restraints and search adequate conformational space under the influence of water molecules and cations. Typically, the NMR restraints are satisfied well within the 2 ns simulation with a tendency towards only small violations associated with the terminal residues.

4.3.6 Quadruplex Structures

For $d(TAGGAGGT)_4$, a total set of 332 restraints was used for energy minimizations and restrained molecular dynamics simulations. No overlapped peaks, or peaks not occurring at all three mixing times were considered, and thus error bounds of 10% were applied to all distances. Restrained molecular dynamics continued for 1 ns with all NOE restraints active. This was followed by energy minimisation. Snapshots from the final 100 ps of rMD were extracted every picosecond. In the final averaged energy minimised structure, none of the applied restraints showed restraint violations >0.5 Å. The mean pairwise RMSD calculated over the final 100 snapshots was 0.82 (±0.1) Å over all heavy atoms. $d(TGGAGGC)_4$ underwent 2 ns of molecular dynamics simulation, restrained by a set of 656 NOE-derived restraints and 16 H-bond restraints. Of the experimentally-derived restraints, 620 were developed from well-resolved non-exchanging crosspeaks, and 36 involving exchangeable base protons. The latter restraints were thus subject to the highest error bounds (25%). In this case, only 6 of the 155 per strand NOEs showed

restraint violations >0.5 Å but <0.8 Å, all of which corresponded to very weak NOEs. The mean RMSD for the final 100 snapshots of d(TGGAGGC)$_4$ was 0.58 (±0.1) Å. d(TGGGAGGC)$_4$ was modelled as above for 2 ns, subject to 112 high quality (10% error bounds) NOE derived restraints per strand (448 total for quadruplex) and a total of 20 H-bonding restraints. Again, the mean pairwise RMSD calculated over the final 100 snapshots was low, at 0.55 (±0.1) Å.

		d(TAGGAGGT)	d(TGGAGGC)	d(TAGGAGGC)
NOE-derived restraints				
	non-exchangable	332	620	448
	exchangable	N/A	36	N/A
Implied restraints				
	H-bonds	N/A	16	20
Restraint Violations				
	>0.5Å	0	24	0
	>0.8Å	0	0	0
RMSD				
	heavy atoms	0.82 (±0.1) Å	0.58 (±0.1) Å	0.55 (±0.1) Å

Table 3.1: Summary of modelling statistics for the final 100 ps of 1 ns rMD trajectories of d(TAGGAGGT) and d(TAGGAGGC), and the final 100 ps of a 2 ns trajectory for d(TGGAGGC).

4.4 Results & Discussion

4.4.1 Non-Exchangable Proton Assignments

Quadruplex DNA structures have been shown to exhibit helical parameters of twist and rise close to those of B-DNA duplexes, in both crystalline (*170*) and solution (*2*) states. As such, many of the assignment pathways for B-form DNA described in Section **2.4.2** are also relevant for tetraplex structures. Resonance assignments of exchangeable and non-exchangable protons were based on information from NOESY spectra at various mixing times (75-300 ms) and through-bond scalar coupling interactions observed from TOCSY and DQF-COSY spectra. Full WATERGATE NOESY spectra (t_m=300 ms) are shown for d(TGGAGGC)$_4$ (Fig.4.12), d(TAGGAGGT)$_4$ (Fig.4.13), and d(TGGGAGGC)$_4$ (Fig.4.14).

Expanded portions of the NOESY spectrum of d(TGGAGGC)$_4$, recorded at 278 K with a mixing time of 300 ms, and showing the sequential pathway followed for the assignment of the sugar H1' protons and the base H6/H8 protons of each nucleotide, can be plotted (Fig.4.15(b)). The sequential connectivities of each H6/H8 base proton involving its own and its 5'-flanking H1' sugar protons can be traced in the same manner as for B-like duplex DNA structures. Fig.4.15(a) shows NOE connectivities between the base H6/H8 protons and their own and their 5'-flanking H2' and H2" sugar protons. These pathways have also be traced out on the NOESY contour plots of d(TGGAGGC)$_4$ and d(TAGGAGGT)$_4$ (data not shown).

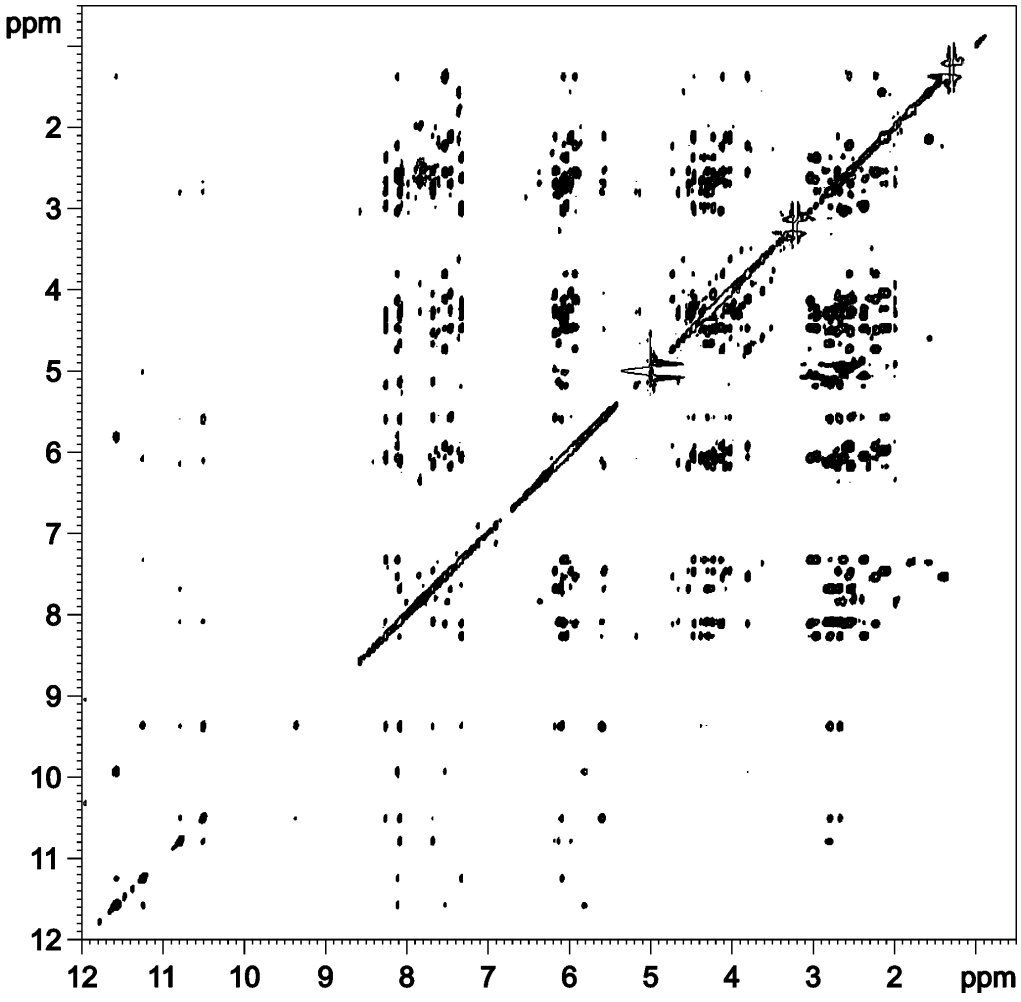


Fig.4.12: NOESY spectrum of d(TGGAGGC)$_4$ in H_2O solution, recorded at 278 K with a mixing time of 300 ms.

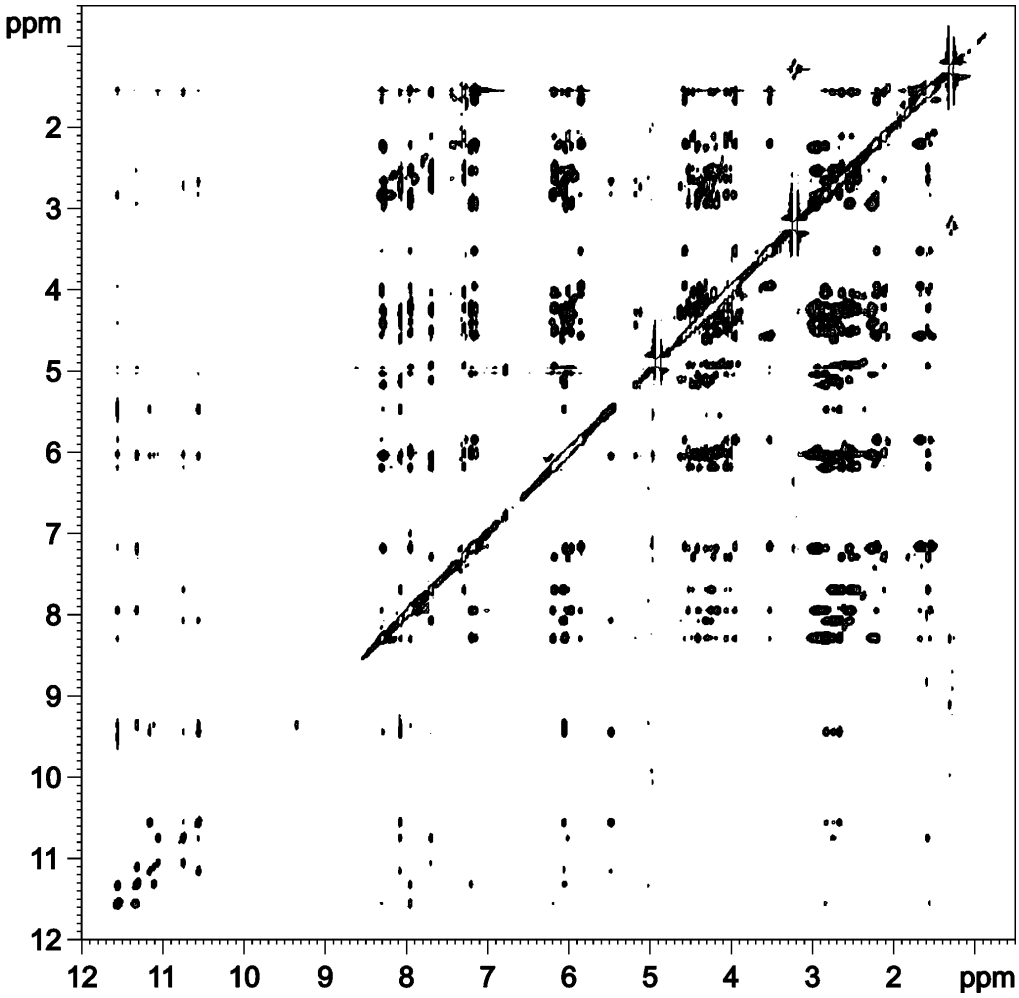


Fig.4.13: NOESY spectrum of d(TAGGAGGT)$_4$ in H_2O solution, recorded at 273 K with a mixing time of 300 ms.

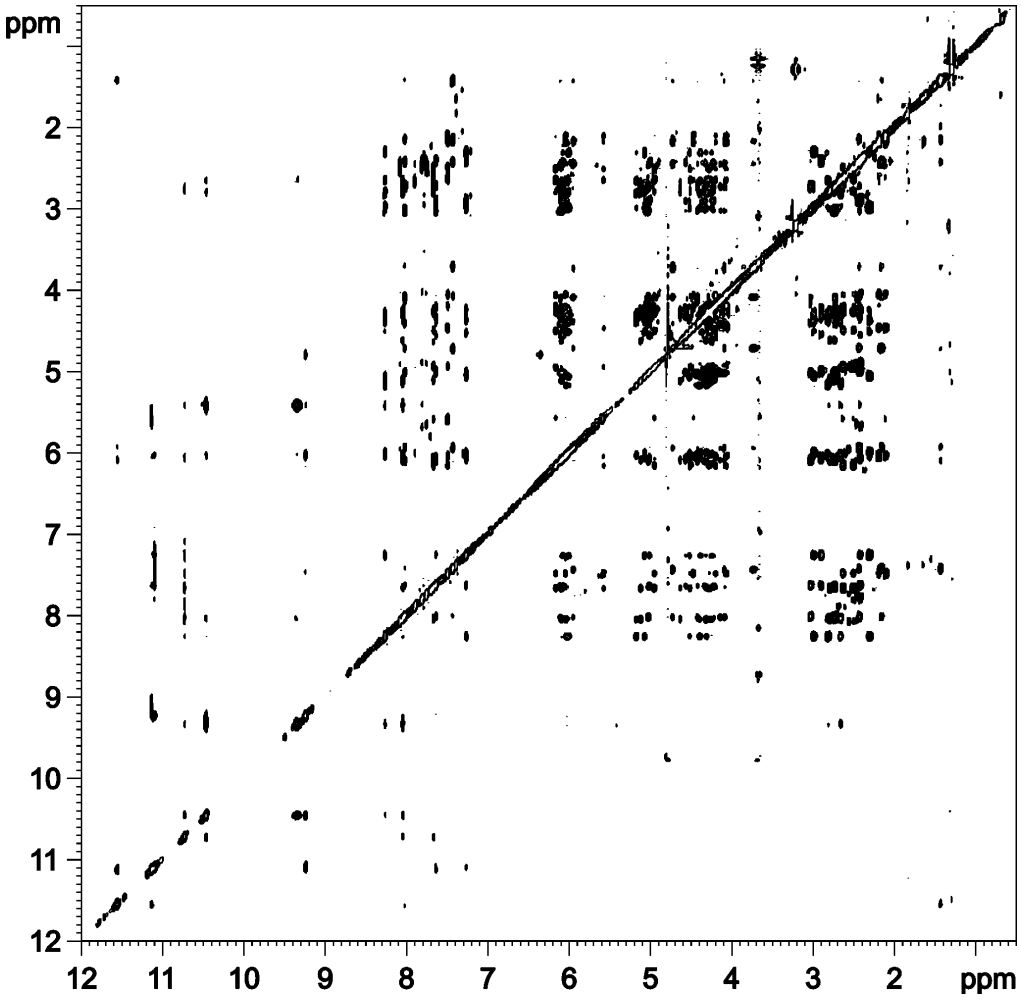


Fig.4.14: NOESY spectrum of d(TGGGAGGC)$_4$ in H_2O solution, recorded at 298 K with a mixing time of 300 ms.

Assignments of all the deoxyribose spin systems, encompassing H1', H2', H2", H3', H4', H5' and H5" protons, were obtained from a combination of NOESY and TOCSY spectra. As was the case for the duplex and hairpin loop structures, the H2' & H2" and, where possible, H5' & H5" resonances were assigned stereospecifically on the basis of their NOE cross peak intensities in the 125 ms NOESY spectra. According to Shugar & Rhemin's rule (*179*), in any sugar conformation the interproton distance between H1'-H2' is always shorter than the H1'-H2" distance. The opposite holds true for H4'- H5'/H5" interproton distances. Thus the resonances of the H2' for all the nucleotides in d(TGGAGGC)$_4$ and its sister quadruplexes appear upfield of the resonance of

the H2" protons, while H5' protons come into resonance downfield of their neighbours.

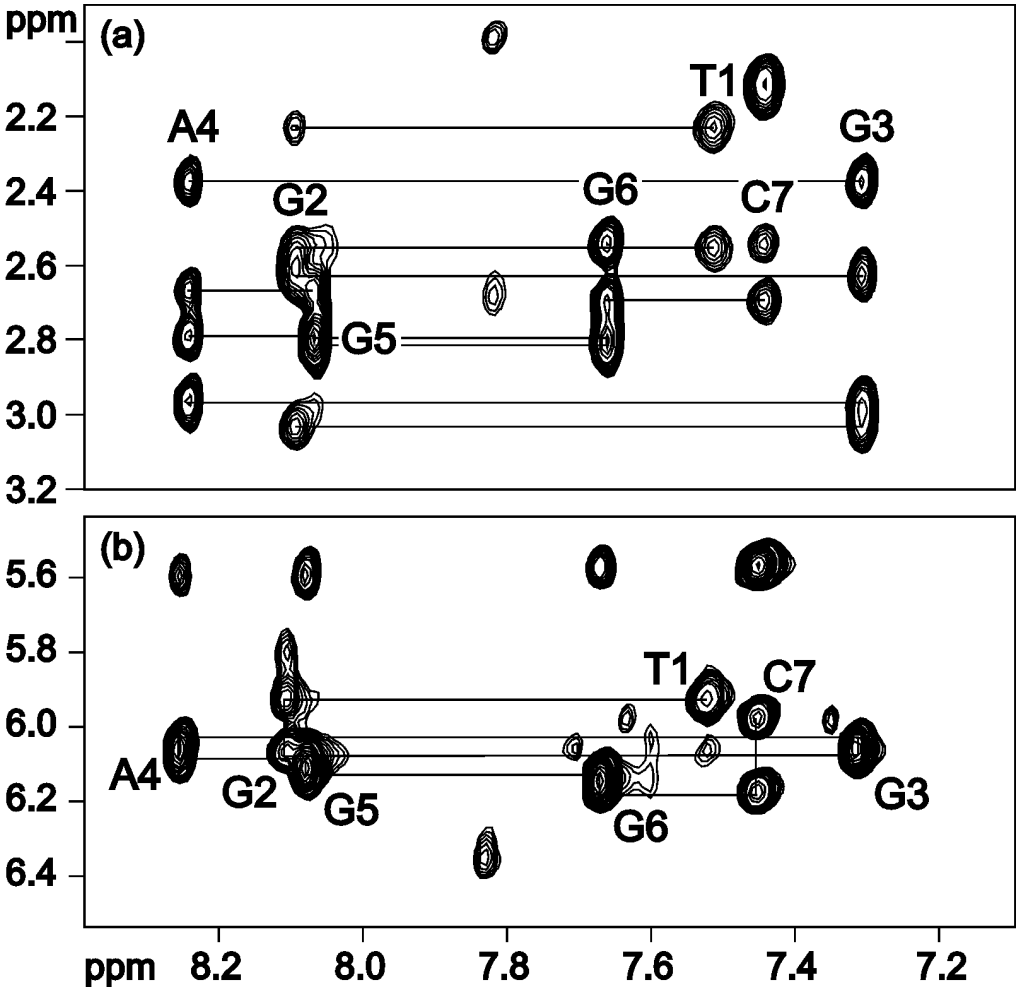


Fig.4.15: Expanded regions of the NOESY spectrum of d(TGGAGGC)$_4$, showing NOE connectivity pathways between (a) H6/H8 & H2'/H2" protons, and (b) H6/H8 & H1' protons.

In each model system, an unusually downfield shifted non-exchangeable resonance at approximately 9.3 ppm can be observed (Fig.4.16). This gives strong NOEs to the imino protons of the guanines flanking the adenine base in the conserved central region (G3 & G5 in d(TGGAGGC)$_4$, and G4 & G6 in d(TAGGAGGT)$_4$ & d(TGGGAGGC)$_4$). These observations imply that the peak corresponds to the sandwiched adenine's H2 proton. If this is the case, the proton would have to be subjected to strong ring current effects, or near to

another strongly desheilding environment in order to perturb the proton to its observed chemical shift, some 1.1 ppm downfield of that typically seen for AH2 protons. This phenomenon is discussed more fully in the context of A-tetrad arrangement in Section **4.4.4**.

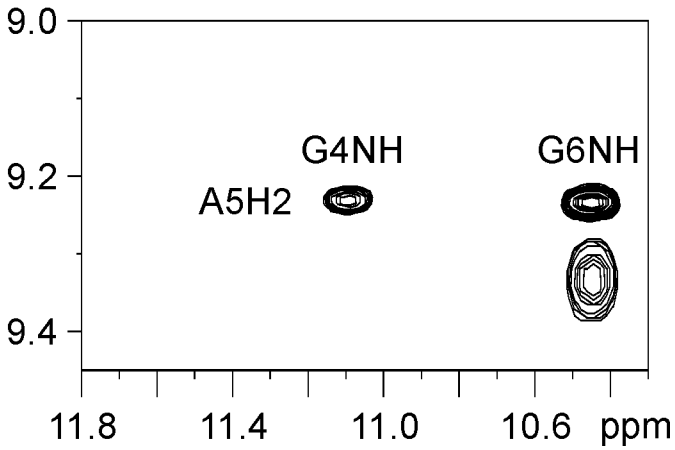


Fig.4.16: A section of the NOESY contour plot of d(TGGGAGGC)$_4$, expanded to show the NOEs between the H2 proton of the adenine base and the imino protons of the flanking guanines, G4 & G6.

The intensities of the NOE crosspeaks between base H6/H8 protons and the sugar protons H1', H2' and H2" on their 5'-flanking residues are similar for all residues in these models, not just for their highly stabilised d(GGAGG)$_4$ core sequences (Fig.4.15). This suggests that the 5'- and 3'-terminal pyrimidine residues are not randomly located or flipped-out of the helix, as suggested by some previous studies (*162;170*), but are involved in stacking with their neighbouring nucleotides. Although stable pyrimidine tetrads have been observed previously (*166*), these have always been flanked on both sides by stable G-quartets, and have shown no evidence for H-bond formation. It is therefore unlikely that these terminal residues hydrogen bond with their interstrand neighbours, but that stacking interactions are their main stabilising force.

The sequential NOE connectivity pathways along the DNA backbone of these systems give patterns and intensities of NOEs consistent with their adenine

bases stacking within the quadruplex with all purine glycosidic torsion angles in the *anti* range.

Glycosidic *anti* and *syn* conformations can be differentiated by distances of 3.7 Å and 2.5 Å, respectively, between the H6/H8 and H1' protons (*131*). These distances were calculated to be in the range 3.3–3.8 Å for all bases in these three models, using the relaxation matrix approach, as described in Section **2.3.4**. This indicates that all residues in $d(TGGAGGC)_4$, and the related quadruplex structures, adopt *anti* glycosidic torsion angle conformations. This agrees with what has been found with all the intermolecular G-quadruplex structures, with the exception of $d(AGGGT)_4$ (*232*), in which the terminal adenines have been shown to adopt a *syn* conformation. The data presented here are consistent with all bases adopting the *anti* conformation.

4.4.2 Exchangable Proton Assignments

1D and 2D NMR spectra of this series of quadruplex structures (pH 7.0, 100 mM KCl, 10 mM phosphate) show that these structures display four-fold symmetry with well-resolved guanine NH resonances consistent with the presence of stable G-tetrads. Both $d(TAGGAGGT)_4$ and $d(TGGAGGC)_4$ exhibit four exchangeable imino proton resonances, while $d(TGGGAGGC)_4$ gives rise to five such signals on the 1D ^{1}H NMR spectrum.

The presence of these four imino signals, one for each G-tetrad, on the NOESY contour plot of $d(TGGAGGC)_4$ (Fig.4.17) suggests a four-fold symmetry, where imino-imino contacts within the tetrad are between magnetically equivalent protons. This gives rise to only one signal per tetrad, and implies a parallel stranded structure. The equivalent anti-parallel structure would be expected to give rise to two sets of intra-tetrad imino signals (*2*), one for each pair of parallel strands.

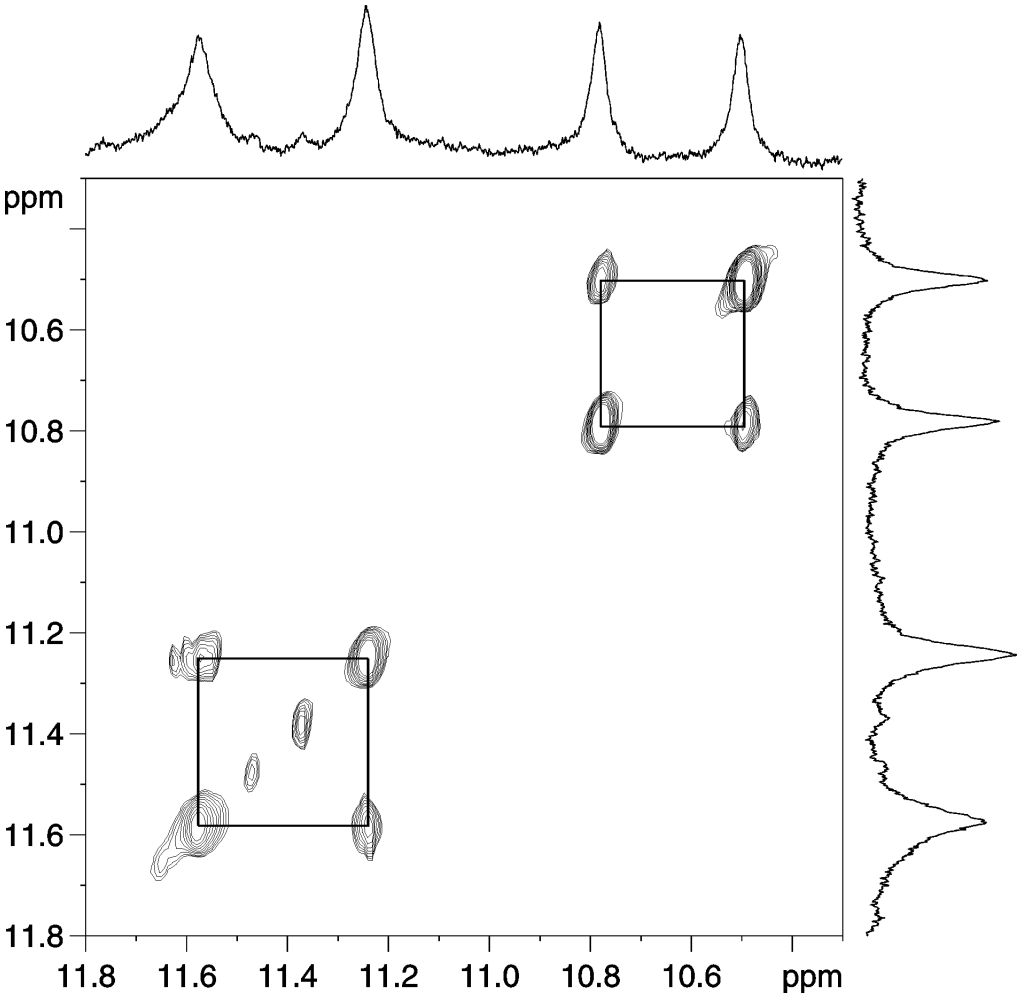


Fig.4.17: ^{1}H NMR and NOESY spectra (600 MHz) of d(TGGAGGC)$_4$ at 279K (90% H_2O, 10% D_2O), expanded to highlight the imino region (10.3-11.8 ppm). Imino resonaces are evident for each of the four G-tetrads. NOE crosspeaks are clearly evident between adjacent tetrads G2 (11.6 ppm) & G3 (11.2 ppm), and G4 (10.7 ppm) & G5 (10.5 ppm).

The equivalent expanded region of the NOESY spectrum of d(TAGGAGGT)$_4$ (Fig.4.18) clearly shows two conformations, one major and one minor, involved in chemical exchange. The significance of this exchange process will be discussed in detail in Section **4.4.3**. For the purpose of our current discussion, it is sufficient to note that there are four imino signals evident for each conformer, which again implies a parallel, four-stranded structure containing four G-tetrads.

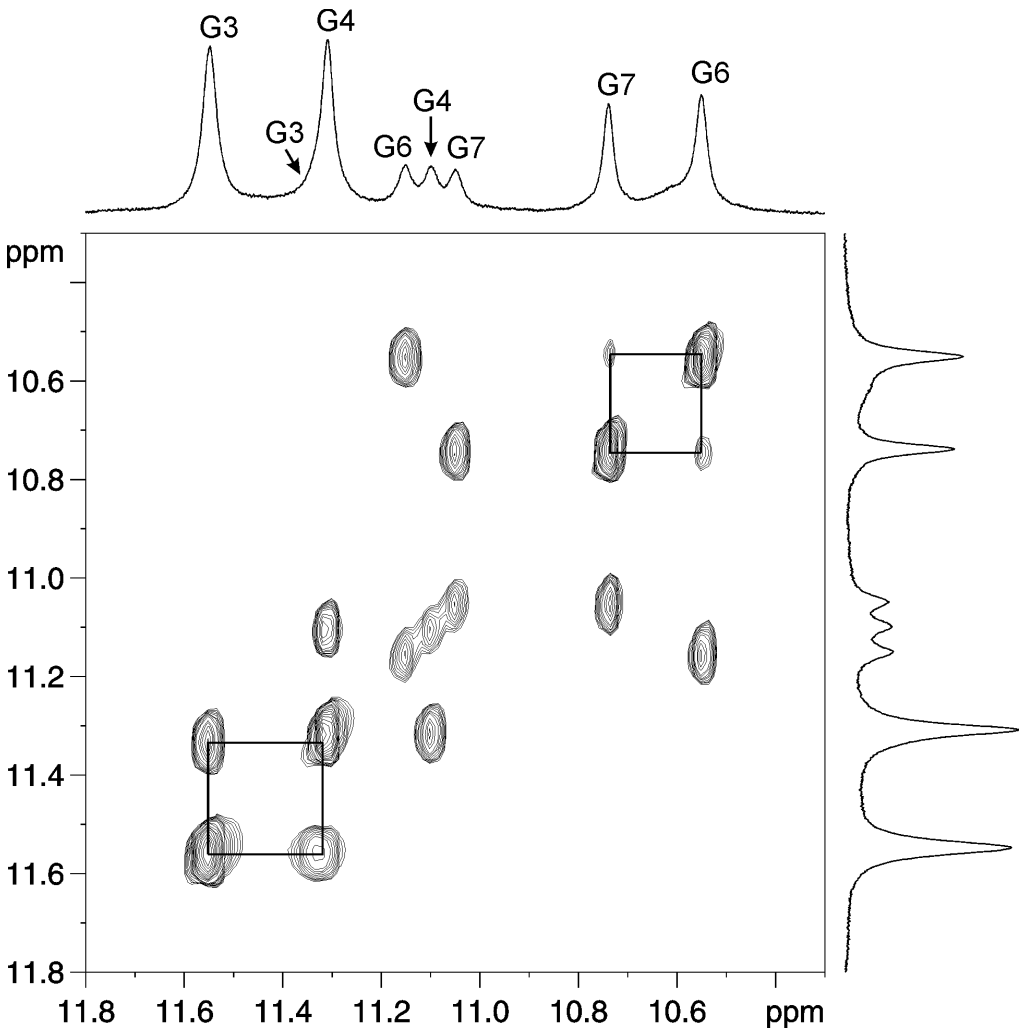


Fig.4.18: ^{1}H NMR and NOESY spectra of d(TAGGAGGT)$_4$ at 279K (90% H_2O, 10% D_2O), expanded to highlight the imino region (10.3-11.8 ppm). In addition to the sequential crosspeaks, as described in Fig.1, exchange crosspeaks between minor (11.0-11.3 ppm) and major conformers are evident.

Similarly, the imino region of the NOESY spectrum of d(TGGGAGGC)$_4$ (Fig.4.19) clearly exhibits five distinct resonances; one for each of the G-tetrads in the conserved d(GGAGG)$_4$ core, and an additional signal corresponding to the stable 5'-flanking G-tetrad.

Additionally, guanine imino protons give NOEs with their own base protons and to the base protons of their 5'-flanking neighbours in the conserved d(GGAGG)$_4$ portion of each system. Thus, for the shorter 7-mer sequence, d(TGGAGGC)$_4$, the 11.6 ppm imino proton of G2 gives NOEs to the base

protons G2H8 and T1H6, the 11.2 ppm imino proton of G3 interacts with base protons G3H8 and G2H8, while downstream of the A-tetrad, NOE interactions can be seen between the 10.5 ppm G5NH and the base protons G5H8, A4H8 and A4H2, and between the G6NH at 10.7 ppm and both G6H8 and G5H8 (Fig.4.20(a)). Similar interactions can be traced out for the longer 8-mer sequences $d(TAGGAGGT)_4$ and $d(TGGGAGGC)_4$. The latter sequence contains an extra step, between the additional G2 imino at 11.5 ppm and the G2H8 and T1H6 base protons (Fig.4.20(b)), which is absent in its adenine containing sister. This is, of course, due to the lack of an imino proton on adenine bases.

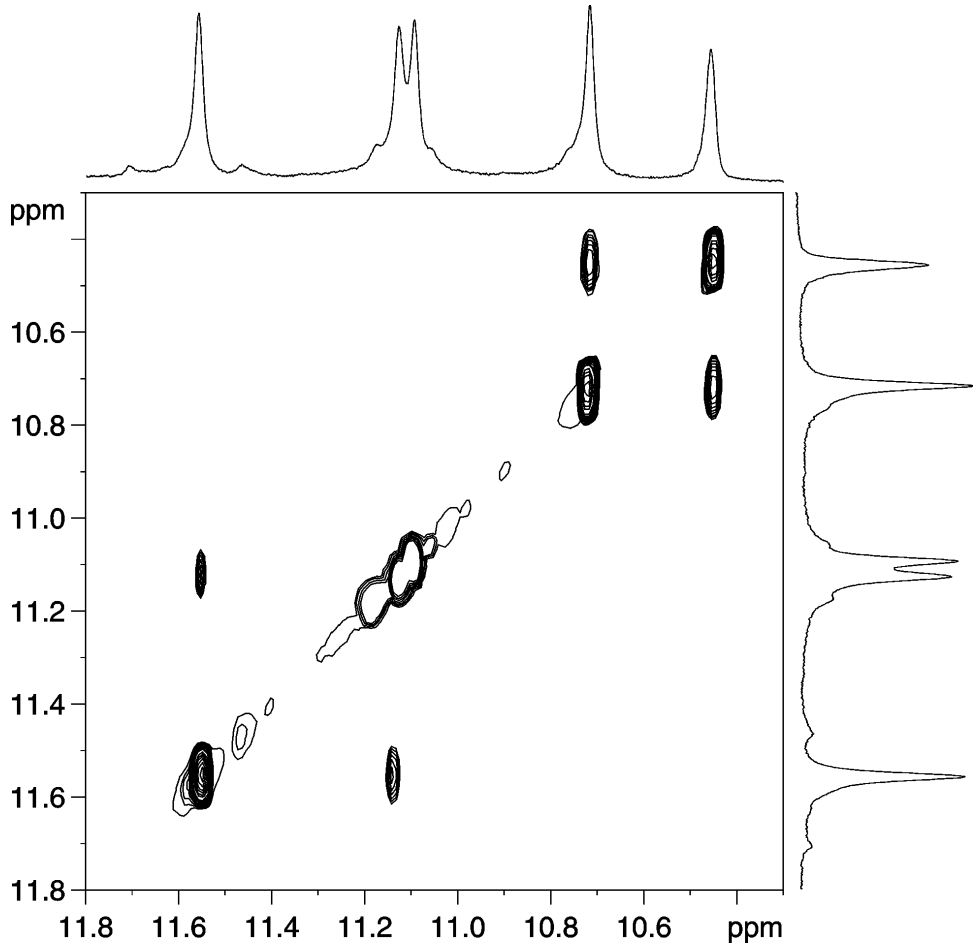


Fig.4.19: ^{1}H NMR and NOESY spectra of $d(TGGGAGGC)_4$ at 298K (90% H_2O, 10% D_2O), expanded to highlight the imino region (10.3-11.8 ppm). Note the crosspeaks between G3NH (11.12 ppm) and G4NH (11.09 ppm) are partially obscured due to their closeness to the diagonal.

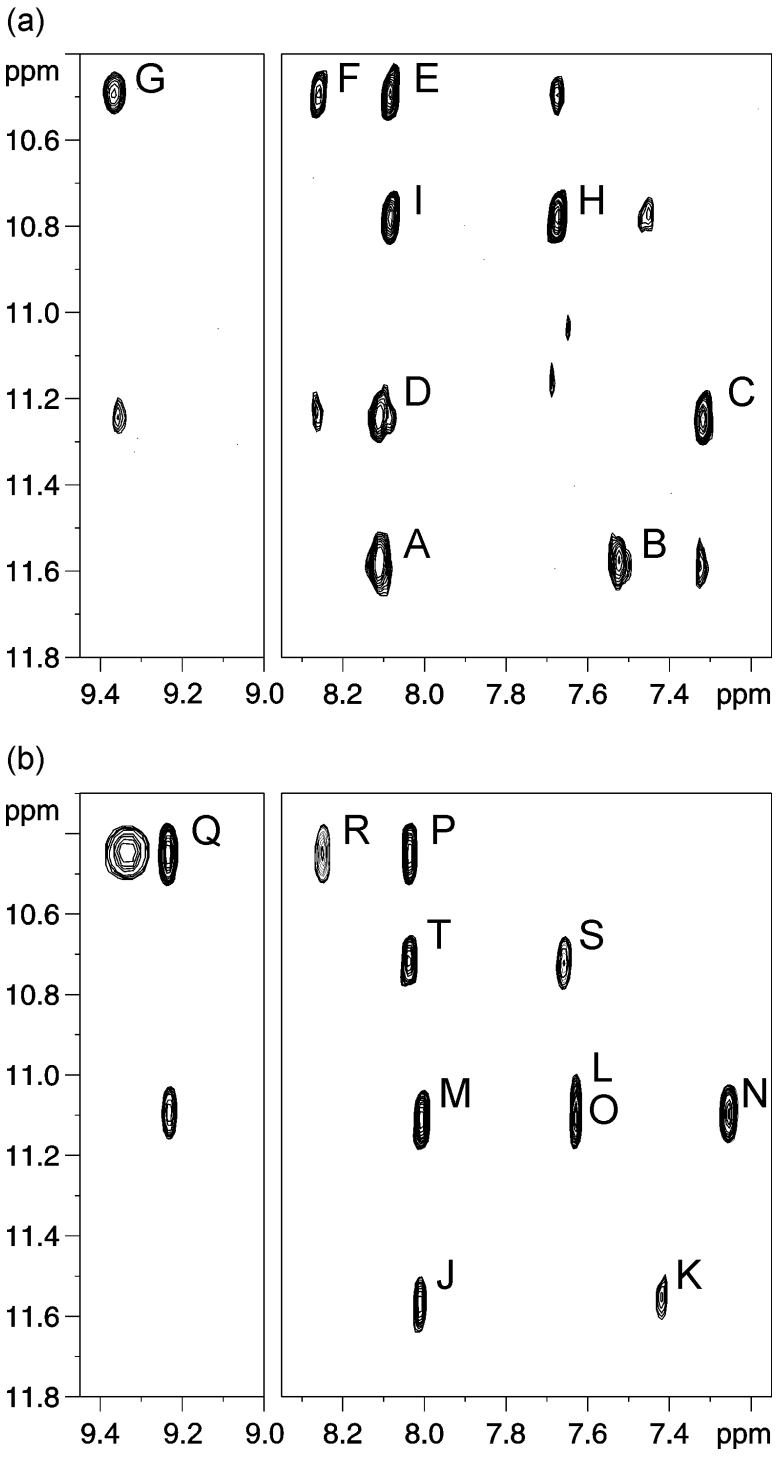


Fig.4.20: Expanded region of the NOESY spectra of (a) d(TGGAGGC)$_4$, and (b) d(TGGGAGGC)$_4$, showing interactions between base protons. NOEs, labelled (A)-(T), are assigned as follows: (A,J) G2NH-G2H8, (B,K) G2NH-T1H6, (C,L) G3NH-G3H8, (D,M) G3NH-G2H8, (E) G5NH-G5H8, (F) G5NH-A4H8, (G) G5NH-A4H2, (H) G6NH-G6H8, (I) G6NH-G5H8, (N) G4NH-G4H8, (O) G4NH-G3H8, (P) G6NH-G6H8, (Q) G6NH-A5H2, (R) G6NH-A5H8, (S) G7NH-G7H8 and (T) G7NH-G6H8.

The expanded region of the NOESY spectrum of $d(TGGAGGC)_4$ in Fig.4.21 includes distinct NOEs between each guanine imino proton and its own amino protons. Fig.4.21 also demonstrates NOEs between amino protons on the same bases, which resonate at two distinct chemical shifts. The observation of two distinct amino protons is consistent with slow rotation of the C–N bond at low temperature. One hydrogen of the amino group resonates near 6.00 ppm and is exposed to the solvent, while the other hydrogen near 9.00 ppm exhibits a NOE with the its own base H8 proton. Since the amino proton is over 7.3 Å from its own H8 proton, the NOE can be attributed to the closeness of this amino group to a symmetry related guanine base within the G-tetrad. The G2 imino proton gives NOEs with its 5.8 ppm and 9.9 ppm amino protons, the G3 imino with its 6.1 ppm and 9.4 ppm aminos, and the G5 imino with its 5.6 ppm and 9.4 ppm aminos. The G6 amino proton of this sequences, as with the G7 amino in its longer cousins, could not be unambiguously assigned, due to the high density of crosspeaks in this region of the spectrum. Characteristic of the amino NOEs is broadness implying a degree of exchange with the solvent.

Particularly strong NOEs are also observed between A4H8 and $G5NH_2$, suggesting strong stacking interactions across the 5'-AG step (Fig.4.22). These interactions are reflected in the relative stabilities of the G-tetrads as evident from NH$\rightarrow$ND exchange experiments (Section 4.4.5), where the disappearance of signals in the 1D spectrum after redissolving the lyophilised sample in D_2O can be observed.

These NOEs provide evidence of the interstrand interactions between the guanine residues that are involved in G-tetrad formation, as they cannot be accounted for by NOEs within the same strand. In addition, such NOEs impart evidence of good stacking interactions between adjacent G-tetrads. More importantly, NOEs between adjacent A- and G-tetrads demonstrate favourable interactions between these tetrads in this conserved family of quadruplex structures.

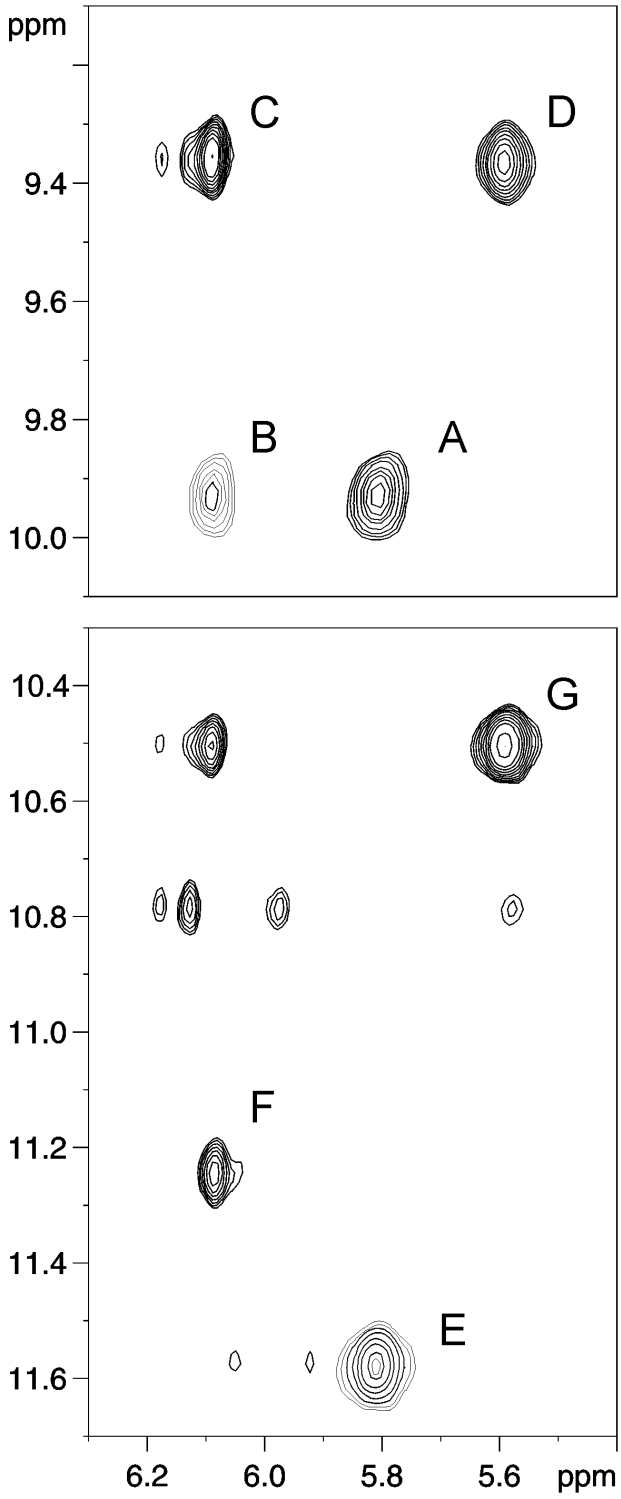


Fig.4.21: Expanded region of the NOESY spectrum of d(TGGAGGC)$_4$ showing NH2-NH2* interactions, and NH-NH2* interactions. NOEs, labeled (A)-(G), are assigned as follows: (A) G2NH2-G2NH2*, (B) G2NH2-G3NH2*, (C) G3NH2-G3NH2*, (D) G5NH2-G5NH2*, (E) G2NH-G2NH2*, (F) G3NH-G3NH2* and (G) G5NH2-G5NH2*. Solvent exposed amino hydrogens are denoted NH2*.

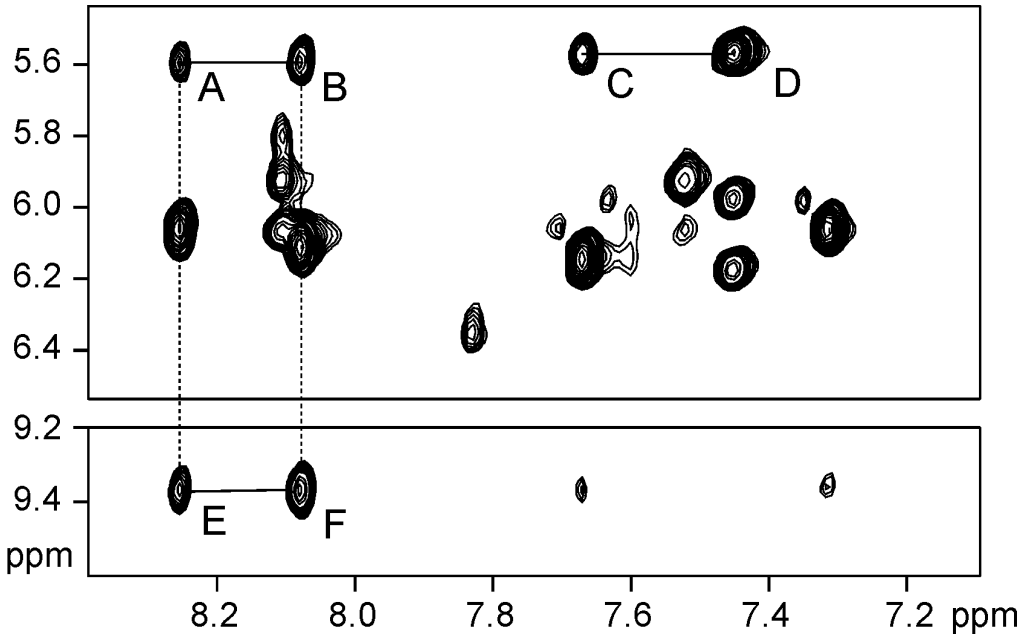


Fig.4.22: Portion of the NOESY spectrum of d(TGGAGGC)$_4$ showing amino-base proton interactions in the conserved central portion. NOEs, labeled (A)-(F), are assigned as follows: (A) G5NH2-A4H8, (B) G5NH2-G5H8, (C) C7H6-G6H8, (D) C7H5-H6, (E) G5NH2*-A4H8 and (F) G5NH2*-G5H8.

The amino protons of the sandwiched adenine were not observed in the 1D NMR spectra of d(TGGAGGC)$_4$, d(TGGGAGGC)$_4$ or d(TAGGAGGT)$_4$. Prior NMR studies have rationalised the absence of ANH_2 resonances in terms of an A-tetrad that is not significantly stabilised by H-bonding (*167*). In studies where these amino resonances have been observed, they have been assigned on the basis of NOE connectivities with 3'-flanking guanine imino protons. In such cases, the A-tetrad was located either on the 5' terminal end of the structure (*167*) or flanked on its 5' face by non-tetrad forming pyrimidine bases (*73*), so that purine-purine stacking interactions across the *N*pA step did not factor, and the formation of stable H-bonds was intrinsic to tetrad formation.

4.4.3 Patterns of Potassium Ion Binding

Previous theoretical studies looking at electrostatic surface potentials (ESP) of A-tetrads using density functional theory (DFT) have shown the negatively charged binding pocket present in guanine tetrads is lacking, leading to the

conclusion that adenine is unlikely to coordinate a monovalant cation (*83*). This is in agreement with the molecular mechanical picture that shows the relatively weak electrostatic interactions with the nitrogen lone pair on the adenine 6-NH_2 group do not appear to compensate for the repulsion between potassium ions. These models show that one potassium cation bound between the GpA or ApG site is tolerated but gives no indication as to the relative stability of models in which the GpA or ApG sites are devoid of potassium ions.

The DFT calculation carried out by Gu and Leszczynski (*84*) suggests that out-of-plane positive electrostatic density on the A-tetrad itself could interact with the ESP located around the guanine carbonyls in a stacked GpA, or ApG, quadruplex. Contour plots of the electrostatic potentials of planar adenine tetrads in both N61 and N67 conformations revealed positive areas in the central part approximately 1.5 Å above the tetrad plane. Since the negative ESP appears around the central area 1.6 Å above the guanine tetrad plane (*129*), this positive centre of the adenine tetrads is expected to reinforce the stacking between the A-tetrads and the G-tetrad in stacked tetraplexes. This strongly suggests a mechanism that placates guanine-adenine tetrad interactions without the need for an intervening cation.

The role of K^+ ion binding in A-tetrad stabilisation is at present unclear and was further investigated by molecular dynamics simulations using the AMBER 5.0 package employing an explicit solvent model.

NMR structural models of $d(TGGAGGC)_4$ and its related quadruplexes were generated on the basis of a large number of NOE-derived distance restraints as described in Section **4.3.6**. Initially, K^+ ions were inserted only between G-tetrads (G•G-A-G•G). These structures give a stable MD trajectory over 2 ns of simulation. Similar structures were subsequently modeled with K^+ ions inserted between each pair of tetrads (G•G•A•G•G). After careful

equilibration of the solvent, the DNA co-ordinates were slowly released according to the standard protocol (see Section **4.3.5**) to undergo free dynamics at 300K. Very rapidly (<2 ps) it was observed that one K^+ ion was ejected from the structure at either the 5'-AG or 5'-GA step without apparent discrimination. Repeat simulations with the homologous sequence $d(TGGGAGGC)_4$ demonstrated this to occur irrespective of the bases flanking the conserved $d(GGAGG)_4$ core. The resulting structures with one vacant site (G•G-A•G•G and G•G•A-G•G) are stable for at least 1 ns of unrestrained dynamics and are of comparable stability to the initial structure with both 5'-GA and 5'-AG sites vacant.

The ANAL module of AMBER 5.0, a function of which is energy decomposition of different groups of atoms, was used to find the interaction energies between different parts of our systems. Previous studies have shown that the negative areas of ESP in the plane of an A-tetrad are limited to the space around the N1 and N6 nitrogens, and that the *O6* oxygens are the only area of significant negative ESP in G-tetrads (*83;84*). Electrostatic components of interactions between these selected groups and intercalated potassium cations, and between the entire tetrad and the ions were found to be in good agreement. Dissection of the electrostatic energy components from the molecular mechanics data, as represented in Fig.4.23, shows that repulsions between K^+ ions bound between adjacent tetrads is large, with a mean value of approximately 30 kJ mol^{-1} averaged over all such interactions in both $d(TGGAGGC)_4$ and $d(TGGGAGGC)_4$. However, when the tetrads consist of guanines in the form (G•G•*N*) or (*N*•G•G), favourable interactions between K^+ ions and the G bases, which exceed 100 kJ mol^{-1} for each tetrad-ion pair, more than offset the repulsion between charges. This is not the case when K^+ ions are bound on either face of an A-tetrad; (G-A•G) or (G•A-G). In such cases, there is a favourable interaction of 30 kJ mol^{-1} between the ion and the A-tetrad, which even when added to the complementary relationship between that ion and the adjacent G-tetrad (100 kJ mol^{-1}), does not offer

nearly as much compensation for the repulsive force between ions. It was not possible to analyse the electrostatic energy components for the (G•A•G) because, as we have discussed, this system was stable for less than 2 ps of the MD trajectory.

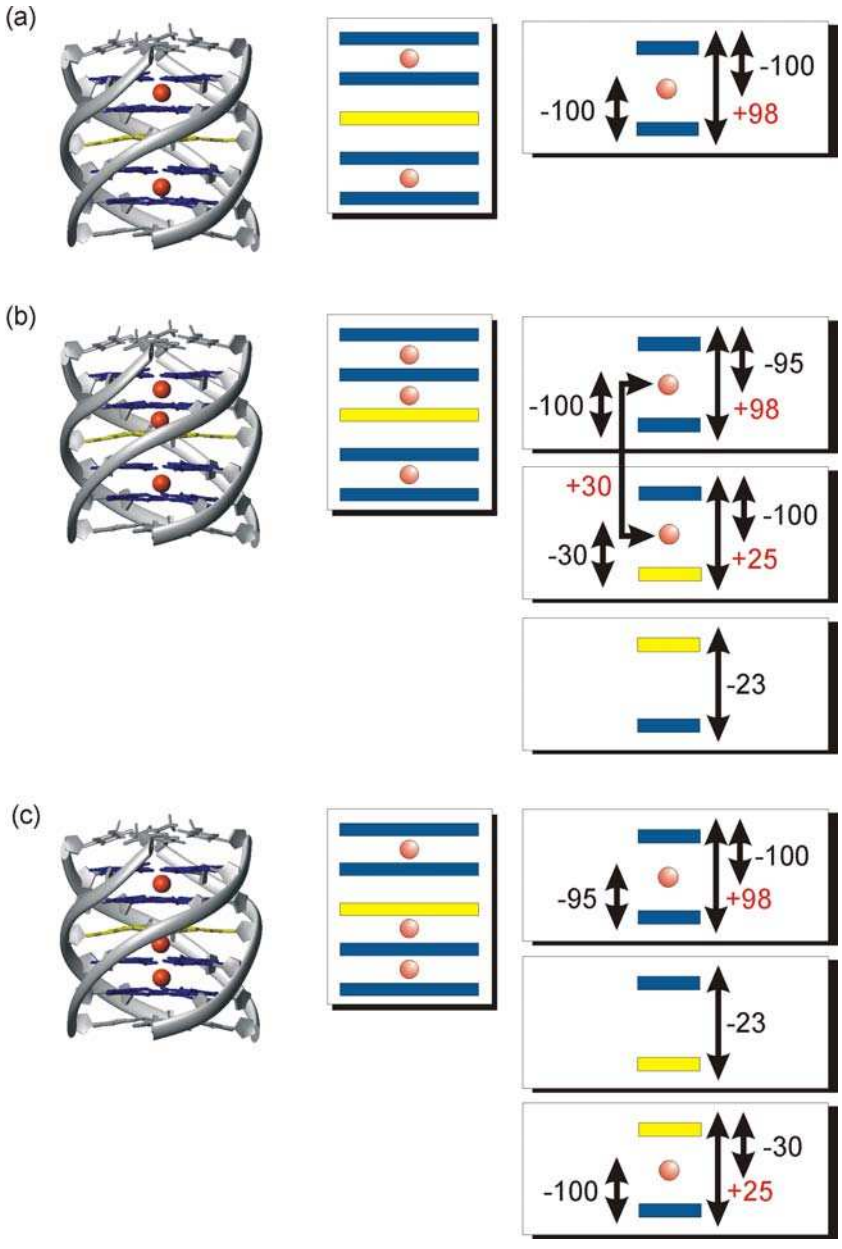


Fig.4.23: Electrostatic energy components of the conserved $d(GGAGG)_4$ core of $d(TGGAGGC)_4$ and $d(TGGGAGGC)_4$, averaged from 1 ns of MD simulation of each quadruplex. Analysis was carried out for models with no ions on either face of the A-tetrad (a), with a vacant ApG step (b) and with a vacant GpA step (c). Attractive (-) and repulsive (+) interactions are in units of kJ mol^{-1}.

The data seem to favour versions of both $d(TGGAGGC)_4$, $d(TGGGAGGC)_4$, and of the major conformer of $d(TAGGAGGT)_4$, without potassium ions intercalated at either GpA or ApG steps. There are insufficient stabilising interactions, as calculated by *ab initio* DTF calculations, to favour a potassium bound site over a vacant one, which is stabilised by complementary electrostatic surfaces on the stacked tetrads. Indeed, the electrostatic energy components from the molecular mechanics data show that the constructive electrostatic interaction (-23 kJ mol^{-1}) between stacked A- and G-tetrads is reversed to an adverse interaction of 25 kJ mol^{-1} upon the introduction of a K^+ ion (Fig.4.23(b);(c)).

Surprisingly, a change in the sequence of the unstructured 3'- and 5'-termini to $d(TAGGAGGT)_4$ results in the partial stabilisation of a minor conformer (~20%) with guanine imino protons resonating over a much narrower chemical shift range (11.0-11.4 ppm) (Fig.4.18). 2D NOESY spectra at 298K (pH 7.0, 100 mM KCl) reveal strong chemical exchange cross-peaks between the two species, which we estimate to be interconverting on a time scale of a few seconds. The major conformer shows the same pattern of protection towards NH/ND exchange seen for $d(TGGAGGC)_4$. The observed time scale of interconversion between conformers demonstrated in Fig.4.18 is fast enough to preclude a conformational equilibrium that involves disruption of A- or G-tetrads, since strand dissociation and re-annealing has been shown to occur over a period of hours at 298K (Section **4.4.5**).

The presence of a minor conformer of $d(TAGGAGGT)_4$ can be rationalised in terms of a K^+ ion binding at either the 5'-GA or 5'-AG site. Though either model has been shown to possess comparable stability, the G4H8 resonance corresponding to the guanine at the 5'-GA step is perturbed more significantly than any other base proton in the NOESY spectrum at 279K (Fig.4.24). Although the minor conformer is only weakly populated, the G4H8 is clearly split into two resolved signals consistent with a K^+ ion binding with low

affinity at the 5'-GA step (G•G•A-G•G), rather than at the 5'-AG step where the purine-purine stacking interactions appear to be stronger (Section **4.4.4**). Additionally, the A5H8 resonance is significantly broadened, though two signals could not be resolved. Based on the ratio of major and minor conformers observed by NMR, we estimate a binding energy of only ~3 kJ mol^{-1}, although the kinetic barrier to exchange is clearly significant since the two conformers are in slow exchange. It is uncertain why a terminal 5'-TA versus 5'-T should produce such a subtle difference in K^+ binding affinity at the 5'-GA step resulting in the observation of two conformers for $d(TAGGAGGT)_4$ but only one for $d(TGGAGGC)_4$ and $d(TGGGAGGC)_4$. It has, however, been shown that the presence of extra bases at the ends of quadruplexes can have a destabilising effect (*89;147;167*).

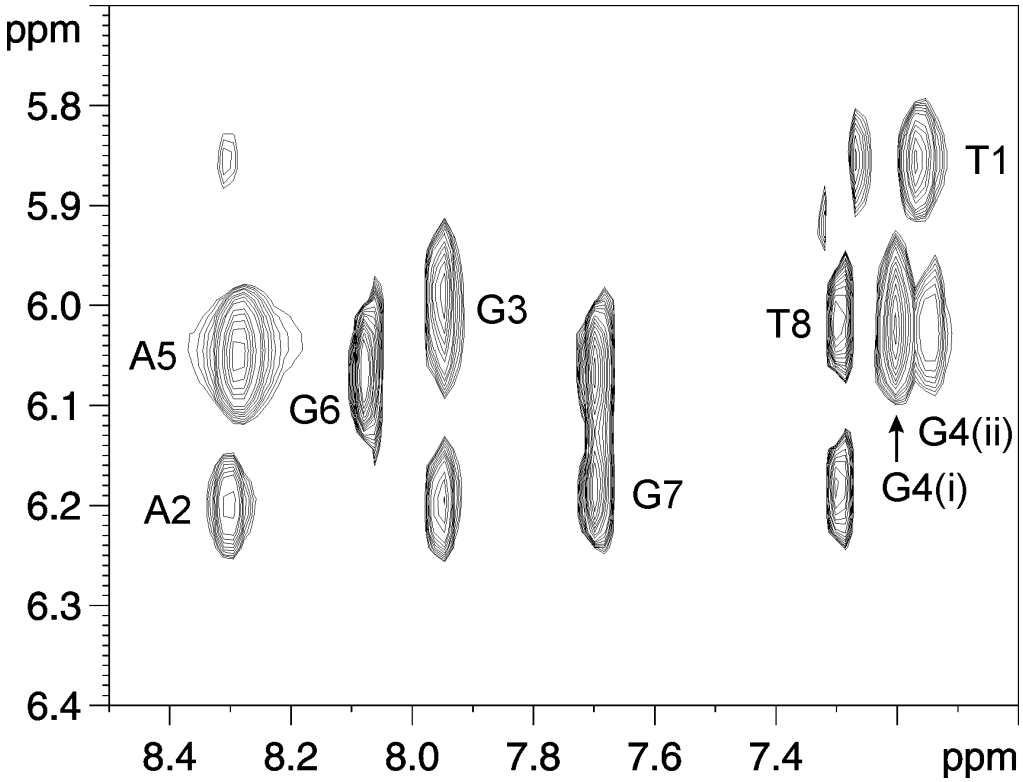


Fig.4.24: Expanded region of the NOESY spectrum of $d(TAGGAGGT)_4$, showing NOE connectivity pathways between H6/H8 & H1' protons with a particular emphasis on G4H8 signal splitting.

4.4.4 A-Tetrad Orientation

As we have discussed in Section **4.4.1**, both $d(TGGAGGC)_4$ and $d(TGGGAGGC)_4$, exhibited strong NOEs between the AH2 base protons and the imino protons of the flanking guanines, providing evidence for good stacking across the GpA and ApG steps.

Observed NOEs can be used to rationalise which A-tetrad orientation is preferred by these models: Distances between base protons across the GpA step are broadly similar in both N61 and N67 models. The distance between GH1 and AH2 in an idealised quadruplex is 3.8 Å when the A-tetrad is in the N61 orientation and 4.1 Å when the N67 conformer is preferred. The GNH_2-AH2 distance across the GpA is maintained at approximately 4.0 Å in both models. This is of little use in determining which conformation is adopted, as both models are consistent with the NMR data, in which NOESY crosspeaks can be observed for these interactions (Fig.4.16).

Across the ApG step, however, there are significant differences in interproton distances between the two proposed A-tetrad forms. The AH2-GH1 interatomic distance for the N61 model is 4.9 Å, compared with 4.2 Å for the N67 conformer. Similarly, the AH2-GNH_2 distance is much larger for the N61 model, at 5.6 Å, while this distance is 3.8 Å for an idealised N67 A-tetrad sandwich. In the NOESY spectra of both $d(TGGAGGC)_4$ and $d(TGGGAGGC)_4$, relatively intense crosspeaks can be observed between AH2 and GH1 protons, which are inconsistent with the larger distance measured in the N61 model. The NOE crosspeaks observed for interactions involving GNH_2 protons were generally broad, and of little use in distinguishing between the two models.

To assist further with the task of deciding which conformation is preferred by the A-tetrads in our models, the same NOE derived restraints were applied to two structural models with different A-tetrad alignments (Fig.4.7), following

unrestrained molecular dynamics. Over the course of the 1 ns rMD simulation, both models should ideally converge to a similar alignment, and thus give an insight into which of these two is the preferred orientation of an A-tetrad.

For these simulations, the potential ion binding sites on each face of the A-tetrads were vacant, consistent with Section **4.4.3**. In each case, the N67 model reverted to N61 orientation before the first 100ps of rMD was completed, with the resultant structure giving a stable MD trajectory over at least 1 ns. The transition from N67 to N61 occurred over one 2 ps timestep, with no intermediate orientation evident on the MD sampling timescale. Both final averaged energy minimised structures of $d(TGGAGGC)_4$ (calculated over the final 100 ps of each simulation) were superimposable and gave a pairwise RMSD fit of 0.78 (±0.1) Å, indicating that the final structures had converged from very different starting structures.

In order to facilitate a comparison between backbone torsion angles in both N61 and N67 versions of $d(TGGAGGC)_4$ over a meaningful timescale, a full nanosecond of molecular dynamics simulation was carried out on an enforced N67 model. This system was subject only to ideal H-bonding restraints for each G-tetrad (Section **4.1.1**), and heavy atom-heavy atom distance of 3.15 (±0.1) Å for N6-N7 distances in the A-tetrad. This allowed for the study of a N67 oriented A-tetrad over a timescale for which it would not normally be stable. Dials plots of backbone torsion angles were plotted for A-tetrad nucleotides in each system for the final 500 ps of their 1 ns rMD simulation (Fig.4.25).

The dials plots clearly illustrate the difference in backbone conformation at the A-tetrad following conversion from N67 to N61 species. Although the overall conformation is relatively unchanged, certain angles in the N67 species do show a much larger range of backbone dynamics than in the N61 type. The

increased amplitude of fluctuations of ζ, α, and χ angles reflects the inherent instability of the former model. The unfavourable torsion angle values seem to be imposed by the relative inflexibility of the enforced N67 tetrad. Backbone torsion angles for all G-tetrads are consistent for both models, as are pyrimidine tetrads (data not shown). Terminal residues display large fluctuations in torsion angles, consistent with the large volume of conformational space they sample over the duration of the rMD run.

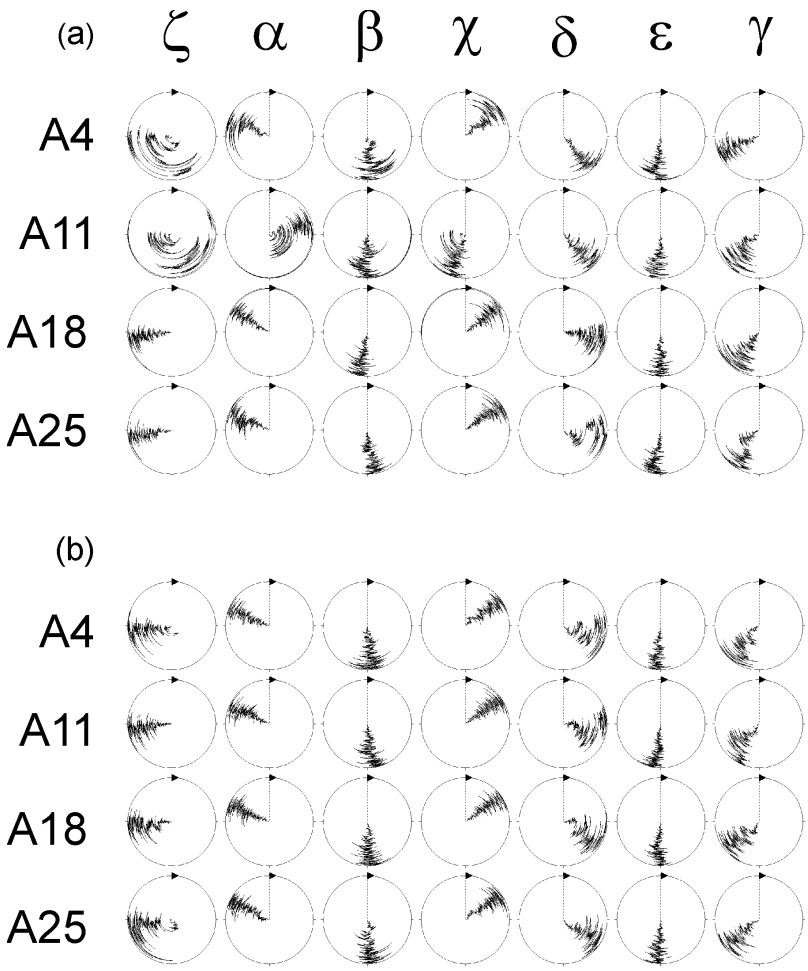


Fig.4.25: Dials plots of backbone torsion angles of adenosine nucleotides in (a) N67 and (b) N61 conformers of d(TGGAGGC)$_4$ representing the final 500 ps of a 1 ns rMD simulation.

Although a large number of NMR studies of nucleic acids have been carried out over the years, and general chemical shift trends are known (*250*), attempts to understand this behaviour have lagged behind similar work for proteins. X-

ray derived structures of proteins are often good models for protein behaviour in solution, but this is less true for nucleic acids, since differences between water and electrolyte conditions in solid- and solution-states can be associated with significant structural changes (*61;62;162;234*). The quality of published solution structures of nucleic acids has been adversely affected by a general lack of long-range constraints that help define global conformations for proteins. In spite of these difficulties, there now exists a large enough body of high quality structural information about nucleic acids to allow for the empirical examination of chemical shifts (*49*).

Empirical analysis of shift dispersion involves using an average value of the chemical shift for each type of proton from a database of nucleic acids, or a random coil value from single stranded DNA. These shifts are then adjusted to allow for environmental factors in the context of the structure of interest that give rise to shielding or deshielding at the probe nuclei. There are many potential contributions to chemical shift dispersion in nucleic acids, including ring currents, other magnetic anisotropies, electrostatics and solvent effects. When considering only sugar H1' and base protons, it has been shown to be sufficient to consider only ring current contributions from aromatic bases and electrostatic interactions from partial charges in the sugars and bases (*239*). A substantial part of the observed chemical shift dispersion can be explained by ring current theories, with smaller estimated contributions coming from electrostatic effects. As the name suggests, empirical analysis involves simple addition of the database values and the calculated proton shift contributions to give predicted values of chemical shift amenable to comparison with experimental results. This process can prove very useful when trying to determine which of a series of starting structures best matches with experimentally derived data.

To compare expected chemical shifts in the two possible A-tetrad conformations, SHIFTS 4.1.1 (*30*), a program to estimate chemical shifts in

nucleic acids, was employed. SHIFTS uses a database of 20 structures, including 3 quadruplexes, and 1631 proton shifts. Only protons bonded to carbon are considered, since protons bonded to electronegative atoms are subject to H-bonding and solvent effects not measured by the program (*207*). SHIFTS approximates the origin of the ring current as a single point-dipole located in the center of each aromatic ring, recognizing a single ring for pyrimidines and two distinct rings for each purine base. The chemical shift perturbation is then calculated based on the distance and the angle between the proton and the center of the aromatic ring. Electrostatics contributions are calculated by *ab initio* methods using the Cornell empirical force field (*47*). This is also the force field employed by AMBER to calculate electrostatics.

SHIFTS was used to compute H1', H2 and H8 proton chemical shifts for various models containing N61 and N67 oriented A-tetrads, with a view to comparing the calculated shifts with those from the NOESY assignments of $d(TGGAGGC)_4$, $d(TGGGAGGC)_4$ and $d(TAGGAGGT)_4$, and so ascertaining which orientation the A-tetrads prefer in these systems.

Although SHIFTS analysis of both N61 and N67 conformers of $d(TGGAGGC)_4$ did not accurately estimate the observed chemical shift of the A4H2, analysis of isolated component elements of each did offer an insight into the unusually downfield shifted resonance of the H2 proton. Predicted shifts for the A4H2 protons in the N61 and N67 models were 7.75 ppm and 7.80 ppm respectively. However, in these models SHIFTS calculated the downfield shifting effect of the neighbouring N7 would be cancelled out by the shielding effect from the adjacent G-tetrads. Empirical analysis of an isolated A-tetrad, where ring current effects from flanking G-tetrads are no longer a consideration, did predict a chemical shift of 9.27 ppm for A4H2 in the N61 conformation, which is consistent with the NMR data. Combined with the lack of evidence for stable hydrogen bonding of the adenine NH_2 protons, this suggests that the orientation of A-tetrads is not determined by the

formation of stable N1···H···N1 bonds. Molecular modelling shows that the structure does adopt a conformation amenable to the formation of these hydrogen bonds, though it may be that some other stabilising force, such as electrostatic and stacking interactions, are involved. This strongly suggests that transient, unstable H-bonds do play some part in the overall stabilisation of these A-tetrads.

Hunter (*102*) analysed the sequence dependent structure of duplex DNA on the basis of base stacking (π-π) interactions. This study showed that sugar-phosphate backbones are relatively flexible and, while they constrain the bases to a small region of conformational space, they play no role in determining the sequence-dependent structure of DNA. It is the stacking interactions that provide the crucial link between sequence, structure and properties. Stacking interactions refer to not only the interactions between π-electrons on consecutive bases, but also include Van Der Waal's interactions and other electronic terms. There are four principal components of π-π interactions. VdW interactions are the major contributing factors, but electrostatic interactions between partial atomic charges, between charge distributions associated with π-electron density, and between charge distributions and partial atomic charges must also be considered. The latter term in particular plays an important role in determining sequence-dependent effects in DNA. Hunter analysed the stacking interactions for the ten possible base-pair steps in Watson-Crick DNA.

Although this and other previous studies (*103*) concentrated solely on duplex DNA, they are still largely amenable to the study of quadruplex DNA. Though the six co-ordinates that describe the degrees of freedom within a single base pair (buckle, propeller-twist, etc.) are relevant only to duplex DNA, the co-ordinates which define an individual base step (rise, twist, etc.) are also relevant to quadruplex DNA, as each individual strand of the tetraplex is *B*-like.

These studies have shown that, in *B*-form duplex DNA, a 5'-AG base step is more stable than a 5'-GA base step, due largely to less repulsive electrostatic interactions between π-electron charge distributions and partial atomic charges on consecutive bases. This also holds true to the analogous base steps in quadruplex DNA, as demonstrated in Section **4.4.5**.

Attractive VdW interactions are maximised when overlap between stacked bases is maximised. In the N61 orientation of the A-tetrad, the six-membered ring of the adenine base lies over the five-membered ring of the 5'-flanking guanine. This is a similar stacking arrangement to that found between adjacent G-tetrads in parallel quadruplex structures (*170*), and in poly(dG)-poly(dA) *B*-form duplex DNA. In projection, this orientation of the rings resembles a decagon, where the vertices of one five-membered ring lie between those of the ring below (Fig.4.26(a)). As well as demonstrating 5'-GA ring-stacking interactions similar to duplex DNA, the five-membered ring of the adenine eclipses the six-membered ring of its 3'-flanking guanine in an identical manner to that observed in its duplex analogue (Fig.4.26(b)). This has been shown to be a favourable interaction, allowing the five-membered ring of the adenine base to maximally overlap the six-membered ring of the 3'-flanking guanine (*146*). It is interesting to note that a similar stacking interaction has been observed accross 5'-GG steps in quadruplex structures (*111;234*). In poly(dG) quadruplex structures, this type of stacking geometry brings the carbonyl oxygens of the upper and lower tetrads into proximity. In this case, potassium binding adds an additional driving force for the bases to adopt this orientation that is absent in adenine-containing steps.

However, these interactions are largely absent in the N67 conformation. In both 5'-GA and 5'-AG steps, there is minimal overlap, and minimal VdW contact, between consecutive bases. On the 5'-GA step, the ring systems of the adenine and guanine bases are offset entirely from each other (Fig.4.27(a)),

while on the 5'-AG step the six-membered ring of adenine staggers both ring systems of guanine so that the π-electron densities are offset from each other (Fig.4.27(b)).

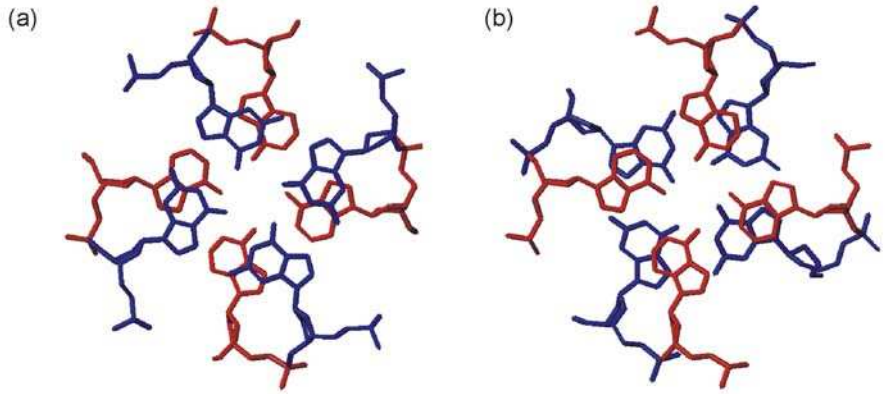


Fig.4.26: Stacking patterns of the 5'-GA (a), and 5'-AG (b) steps of the average minimised structure of d(TGGAGGC)$_4$ in the N61 conformation. For clarity, adenines are displayed in red and guanines in blue.

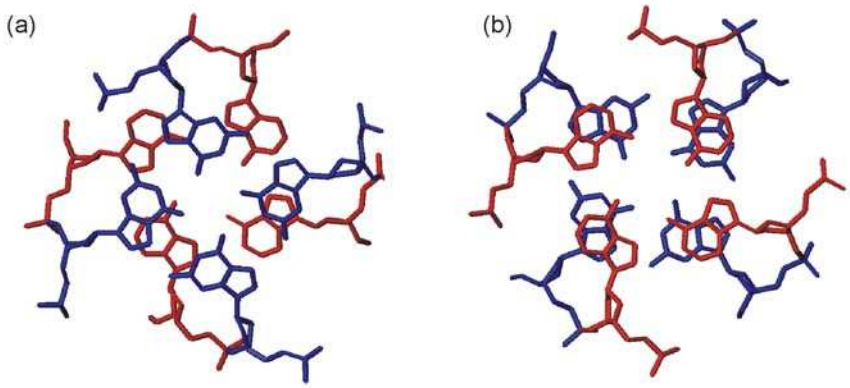


Fig.4.27: Stacking patterns of the 5'-GA (a), and 5'-AG (b) steps of the average minimised structure of d(TGGAGGC)$_4$ in the N67 conformation.

A quantum chemical study of the stability and structures of the two isolated forms of the adenine tetrad (*84*), which made the assumption that the stability of tetrads in the isolated form is the main contribution to the formation of tetrads in quadruplexes, showed that the H-bonds in the A-tetrad do not significantly alter the geometric parameters of the adenines. This was not unexpected because the H-bonding between the H atom of the amino group and the N1 or N7 atom was shown not to be very strong. The hydrogen bonds holding the adenines together in N61 were shown to be 1.3 kcal mol^{-1} stronger

than those in N67, though H-bonding energies were shown to be insignificant compared with potential stacking energy between the A-tetrad and a flanking G-tetrad (Section **4.4.3**).

For these model systems, a downfield ring-current shift for the AH2 resonance (Section **4.4.1**) appears to be diagnostic of a stabilised A-tetrad with a putative 6-NH_2 to N1 hydrogen bonding geometry. The absence of clearly resolved resonances for the 6-NH_2 group, analogous to those of the 2-NH_2 group of guanine, suggests that base-specific electrostatic interactions from π-stacking of A and G bases, rather than hydrogen bonding between adenines, may be largely responsible for A-tetrad stabilisation in this context.

4.4.5 Relative Stabilities of G-tetrads

The relative stabilities of the G-tetrads in d(TGGAGGC)$_4$, d(TAGGAGGT)$_4$ and d(TGGGAGGC)$_4$ were assessed by NH/ND exchange experiments, where the disappearance of signals in the 1D spectrum after redissolving the lyophilised sample in D_2O were observed. A prior analogous experiment reported the high kinetic stability of G-tetrads in quadruplex structures (*73*).

Our studies show that exchange is extremely slow at 298 K, at which temperature the G5NH proton of d(TGGAGGC)$_4$ was still detectable 50 days after the sample was redissolved in D_2O. The exchange process can, however, be accelerated at elevated temperatures (308 K). Even under these conditions marked sequence-specific effects can be seen showing that the G5NH (G6NH in d(TAGGAGGT)$_4$) resonance persists considerably longer than other guanine NH signals. The reverse experiment, in which the quadruplex is first melted at high temperature and reannealed in D_2O to trap ND groups, and then redissolved in H_2O solution, shows that the G6NH signal of d(TAGGAGGT)$_4$ is very slow to reappear, indicating specific protection and stabilisation against ND/NH exchange (Fig.4.28). These observations are again consistent with particularly good base stacking interactions at the 5'-AG step, as indicated in Section **4.4.4**.

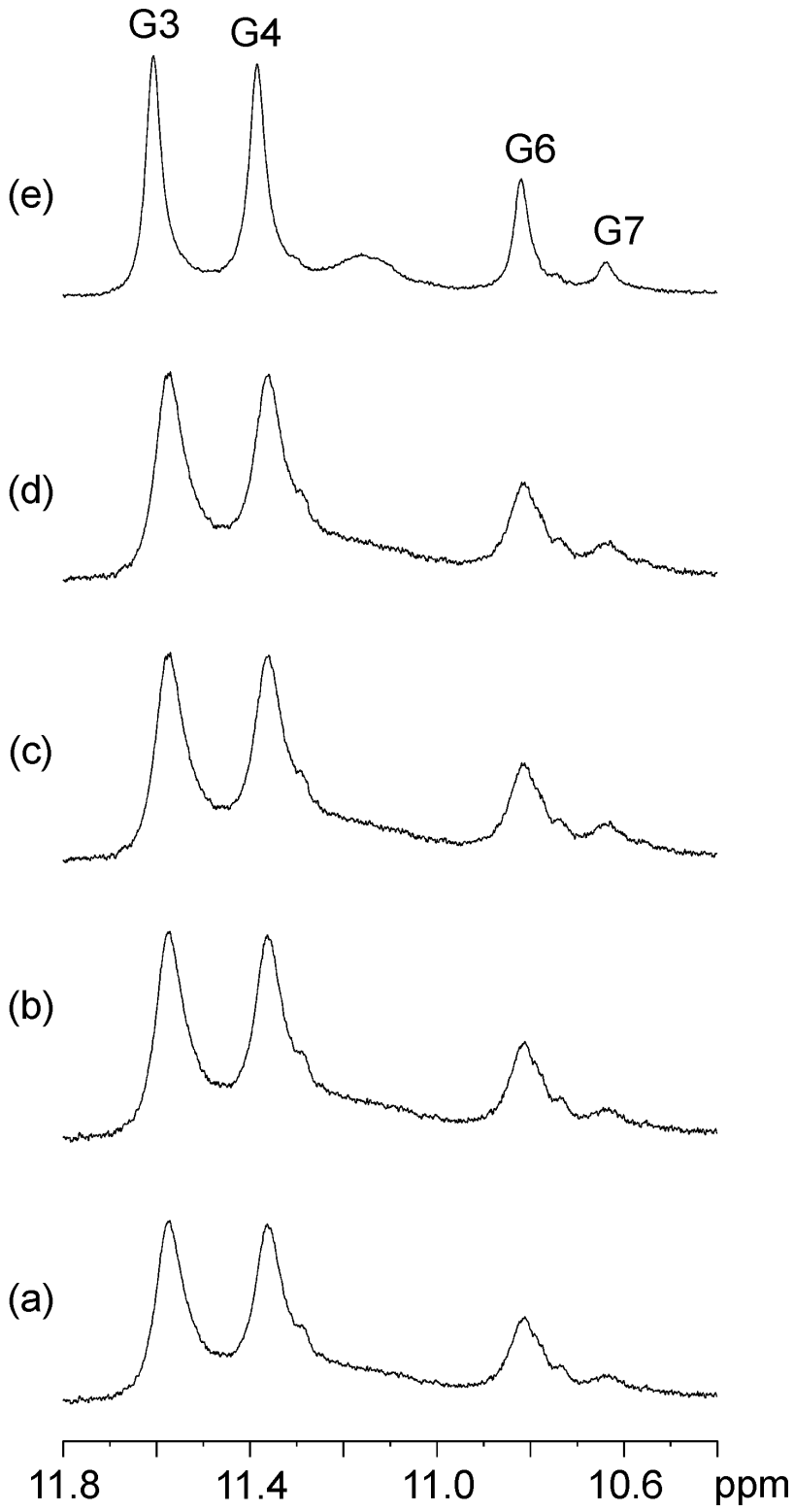


Fig.4.28: Stacked plots of 1D WATERGATE spectra of d(TAGGAGGT)$_4$ (308 K) collected (a) 15 mins, (b) 1 hour, (c) 6 hours, (d) 18 hours, and (e) 7 days after reconstituting the dry sample in H_2O, demonstrating protection against ND/NH exchange.

In d(TGGGAGGC)$_4$, the analogous G6NH also shows enhanced stability following resolvation in D_2O. However, the longer run of guanines upstream of the A-tetrad significantly alters the relative stabilities of the other tetrads. The iminos of the outermost guanines, G2 and G7, still exchange rapidly after redissolving in D_2O, and have disappeared completely from the 1D WATERGATE spectrum after the first sampling interval (15 mins) (Fig.4.29). The G3 and G4 imino protons are still detectable after 24 hours at 298K, and show no apparent loss of intensity. A temperature study of d(TGGGAGGC)$_4$ (Fig.4.30) shows that the iminos of G3, G4 and G6 are all still detectable to temperatures in excess of 338 K, though there is a noticeable decrease in the intensity of the latter two peaks above 328 K. G3NH is stable at temperatures in excess of 348K, and its resonance shows no intensity drop after 6 hours at this temperature.

The increased stability of the G3NH proton in the longer sequence can be attributed to the fact that the G3 tetrad is flanked on both faces by stable G-tetrads, which decrease solvent accessibility to the core of the quadruplex in this region. The exchange data from d(TGGAGGC)$_4$ suggest that the 5'-AG step has stronger stacking interactions than the 5'-GA step, and this would not be expected to change in d(TGGGAGGC)$_4$, as the core sequence in these models is conserved. The stability of the G4 tetrad is greater in the longer sequence due to the presence of a greater number of neighbouring G-tetrads adjacent to its 5' face. Combined with the unusually stable G7 tetrad on the 3' side of the quadruplex, these act like clamps, preventing solvent from reaching the G4NH protons from the long axis of the structure.

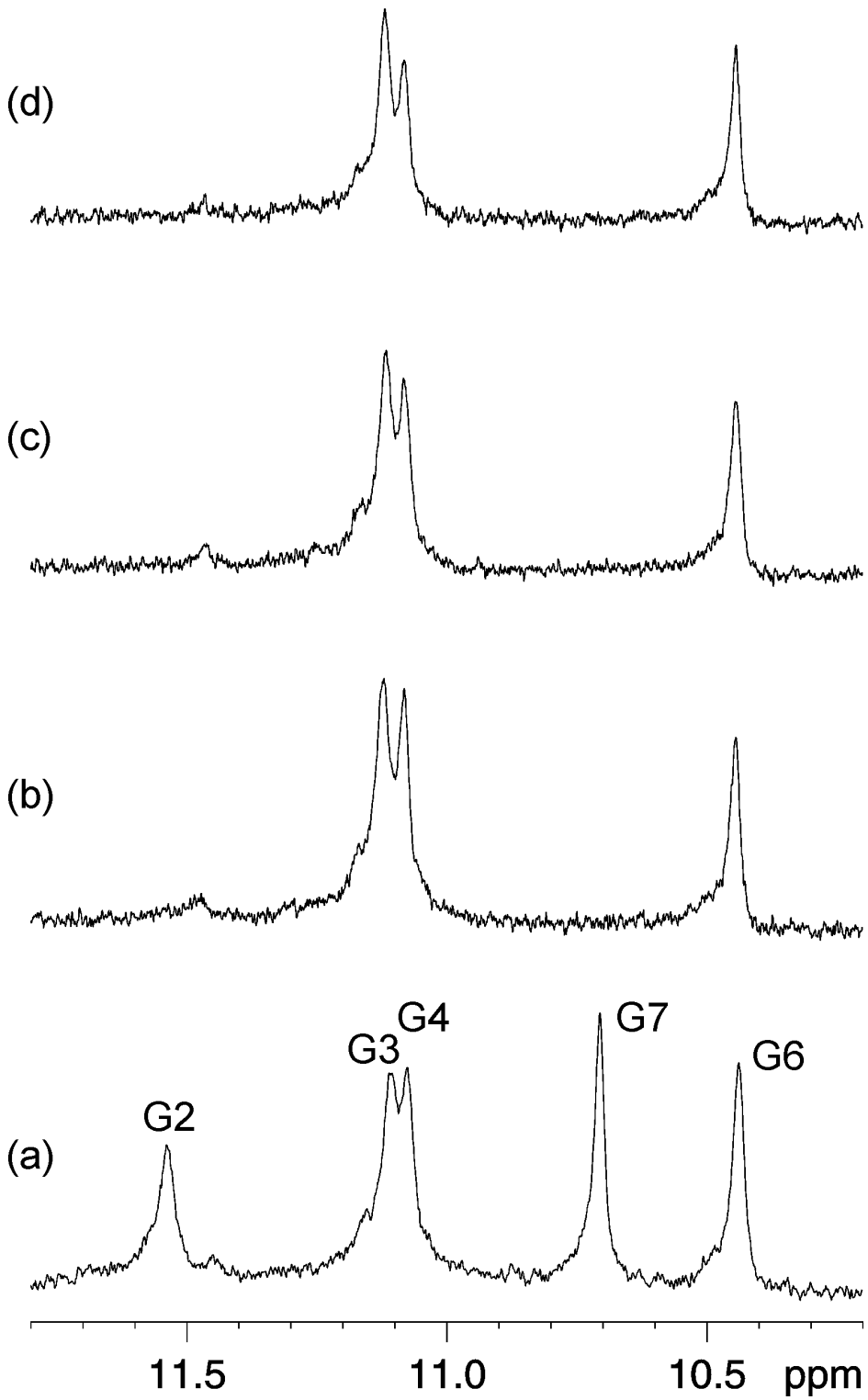


Fig.4.29: Stacked plots of 1D WATERGATE spectra of d(TGGGAGGC)$_4$ (298 K) collected (a) before and (b) 15 mins, (c) 6 hours, and (d) 18 hours after reconstituting the sample in D_2O.

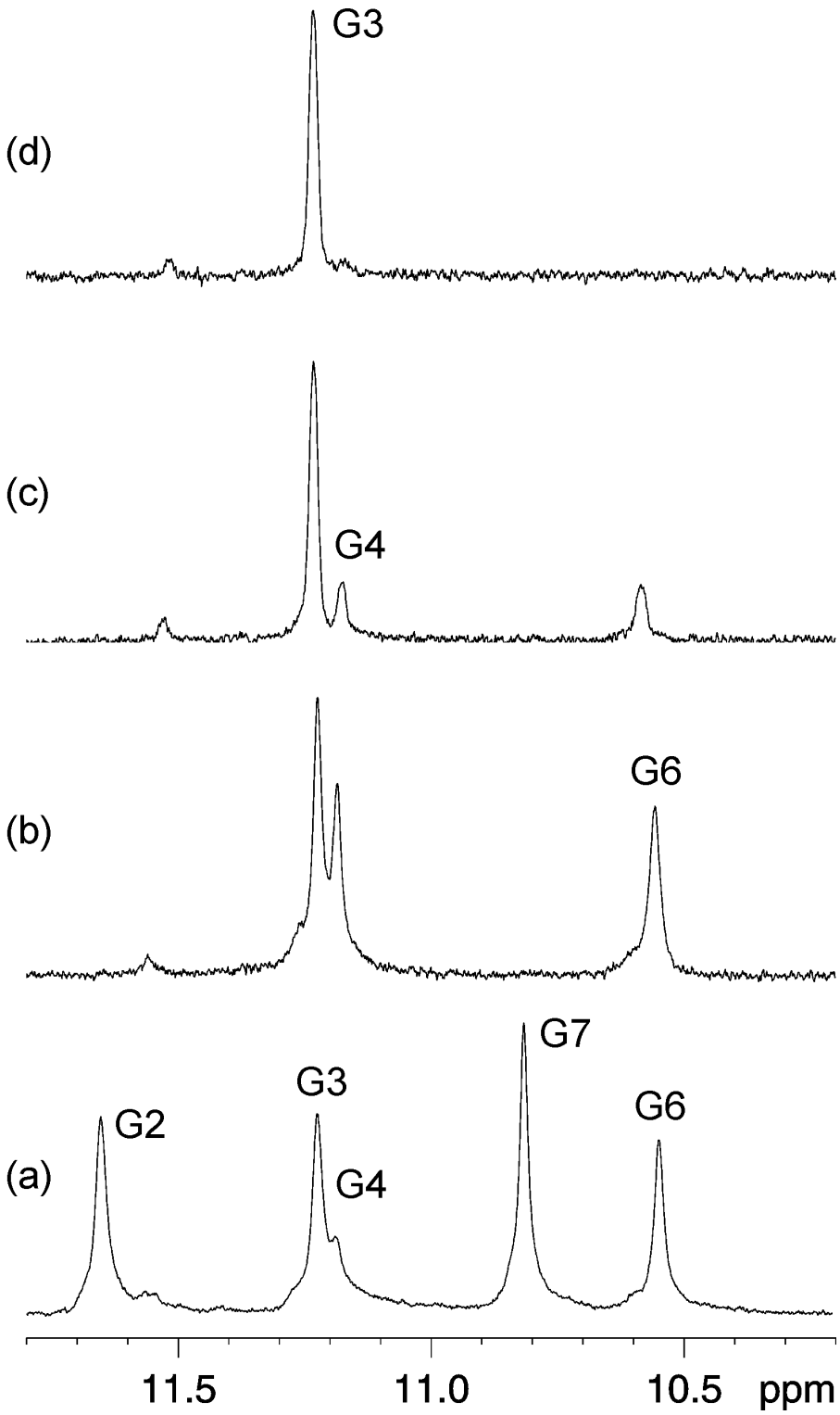


Fig.4.30: Stacked plots of 1D WATERGATE spectra of d(TGGGAGGC)$_4$ collected at (a) 298 K, (b) 308 K, (c) 338 K, and (d) 348 K.

4.5 Conclusions

This study has determined the high-resolution structure of the quadruplex structures $d(TAGGAGGT)_4$, $d(TGGAGGC)_4$, and $d(TGGGAGGC)_4$ by an NMR-restrained molecular dynamics approach that includes explicit treatment of solvent and stabilising ions. These structures are each suitable for deposition to the *Nucleic Acids Database*. These sequences form similar stable structures in the conserved $d(GGAGG)_4$ purine core region, which are enhanced by strong stacking interactions.

As well as verifying the presence of stable G-tetrads, the structure determination, supported by NMR data, demonstrated that the adenine nucleotides on separate strands could also form H-bonded tetrads. These A-tetrads have been shown to adopt *anti* glycosidic torsion angles, and have a single preferred conformation with H-bonds between N6 and N1 atoms on adjacent tetrads, as demonstrated by chemical shifts analysis and rMD simulation. In this way, these structures differ from the previously reported structures of $d(AGGGT)_4$ and $d(TAGGGT)_4$, which were shown to be in equilibrium between N61 and N67 A-tetrad orientations. In the current context, the A-tetrad conformation is strongly influenced by favourable base stacking geometries. Terminal pyrimidine residues do not seem to have the potential to form stable tetrads, as stabilising purine tetrads do not surround them. Thymine and cytosine nucleotides do not form H-bonds of any kind, but do appear to stack with their adjacent nucleotides in a B-like conformation.

MD simulations and dissection of electrostatic energy components have shown that stabilising potassium ions cannot reside on both faces of an A-tetrad simultaneously. Though temperature studies suggest that the 5'-AG step has stronger stacking interactions than the 5'-GA step, structures with either site occupied have proved to be equally stable. However, the exclusively observed species in $d(TGGAGGC)_4$ and $d(TGGGAGGC)_4$ have both sites

vacant, while only in the minor conformer of $d(TAGGAGGT)_4$ is the 5'-GA step occupied by a cation.

These high-resolution structures add valuable information to the pre-existing family of parallel quadruplex structures studied so far.

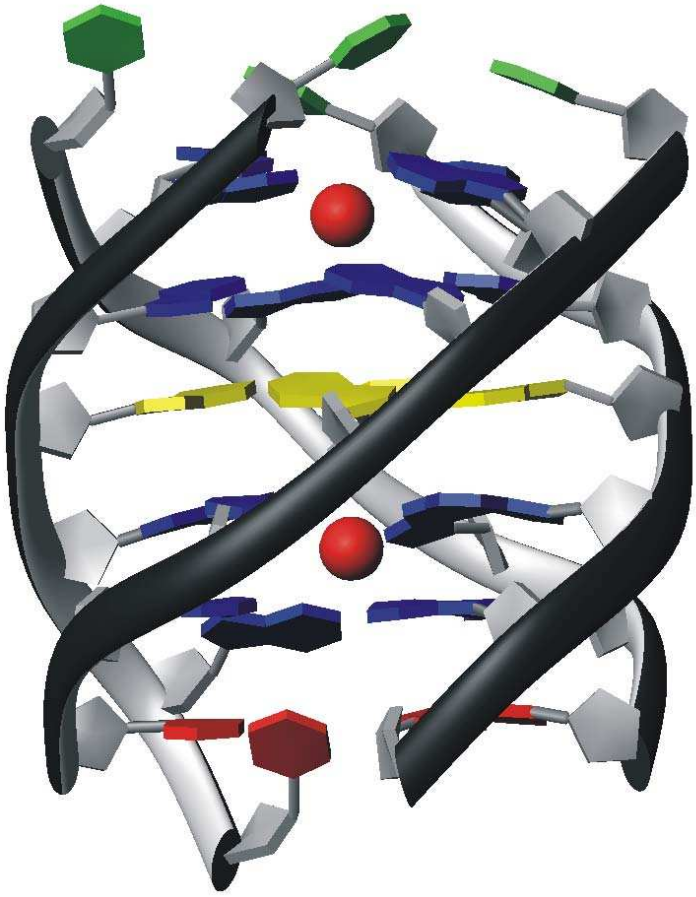


Fig.4.31: Schematic representation of the average minimised structure of $d(TGGAGGC)_4$ from the last 100 ps of a 2 ns rMD simulation, showing interstitial potassium ions. Guanines are coloured in blue, adenines in yellow, thymines in green, and cytosines in red.

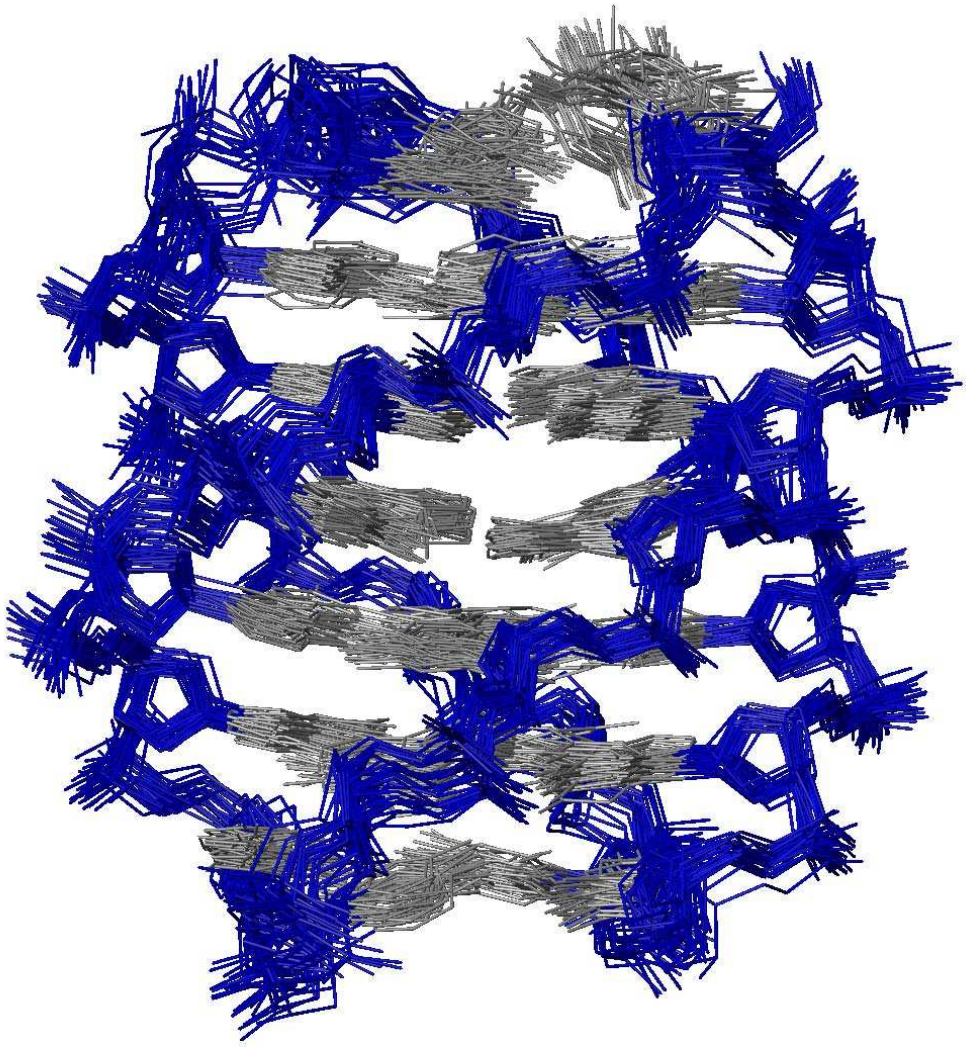


Fig.4.32: 25 randomly chosen structures from the last 500 ps of a 2ns rMD simulation of d(TGGAGGC)$_4$

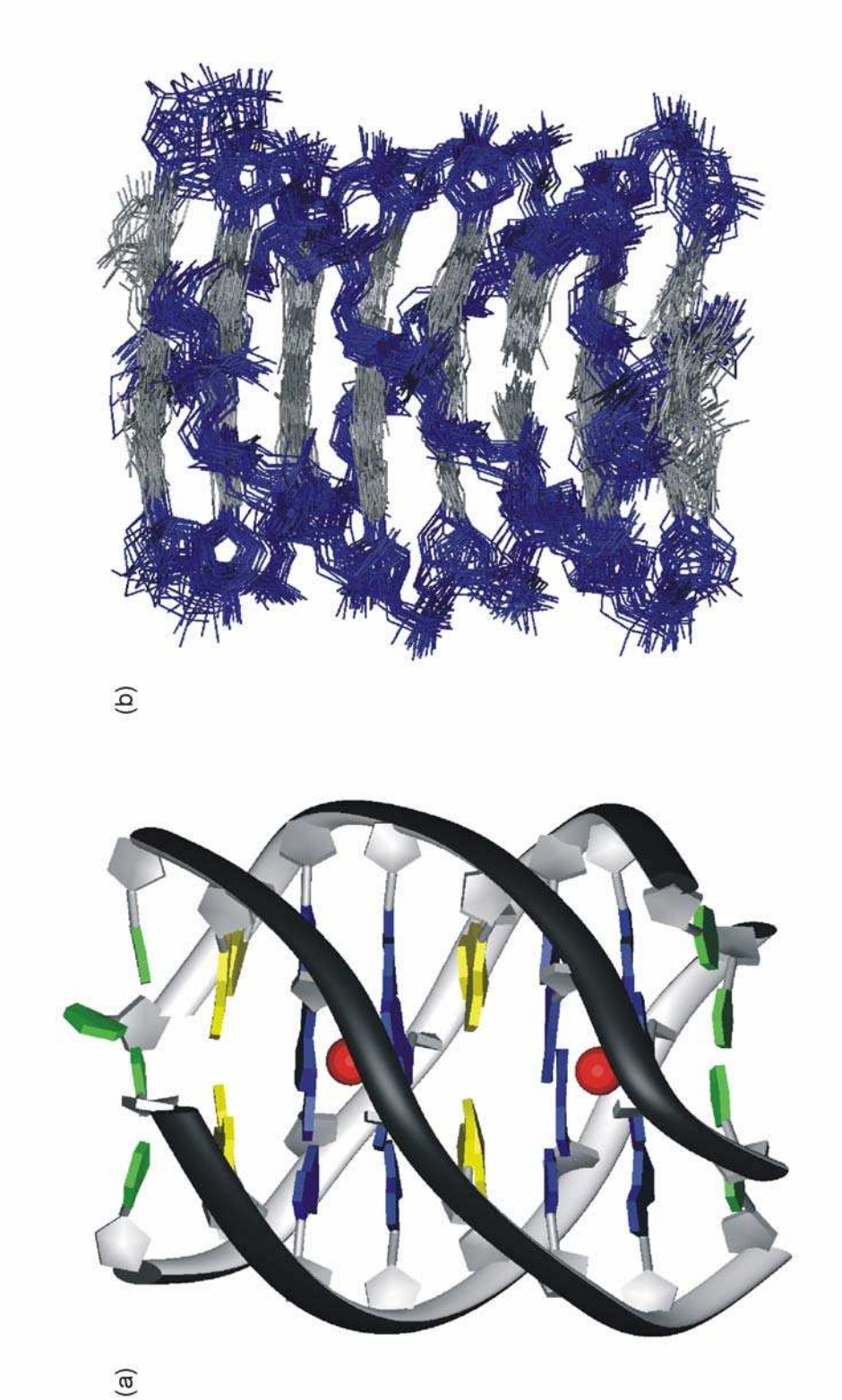


Fig.4.33: (a) Schematic representation of the average minimised structure of d(TAGGAGGT)$_4$ from the last 100 ps of a 1 ns rMD simulation, showing interstitial potassium ions. (b) 25 randomly chosen snapshots from the last 250 ps of a 1 ns rMD simulation of d(TAGGAGGT)$_4$.

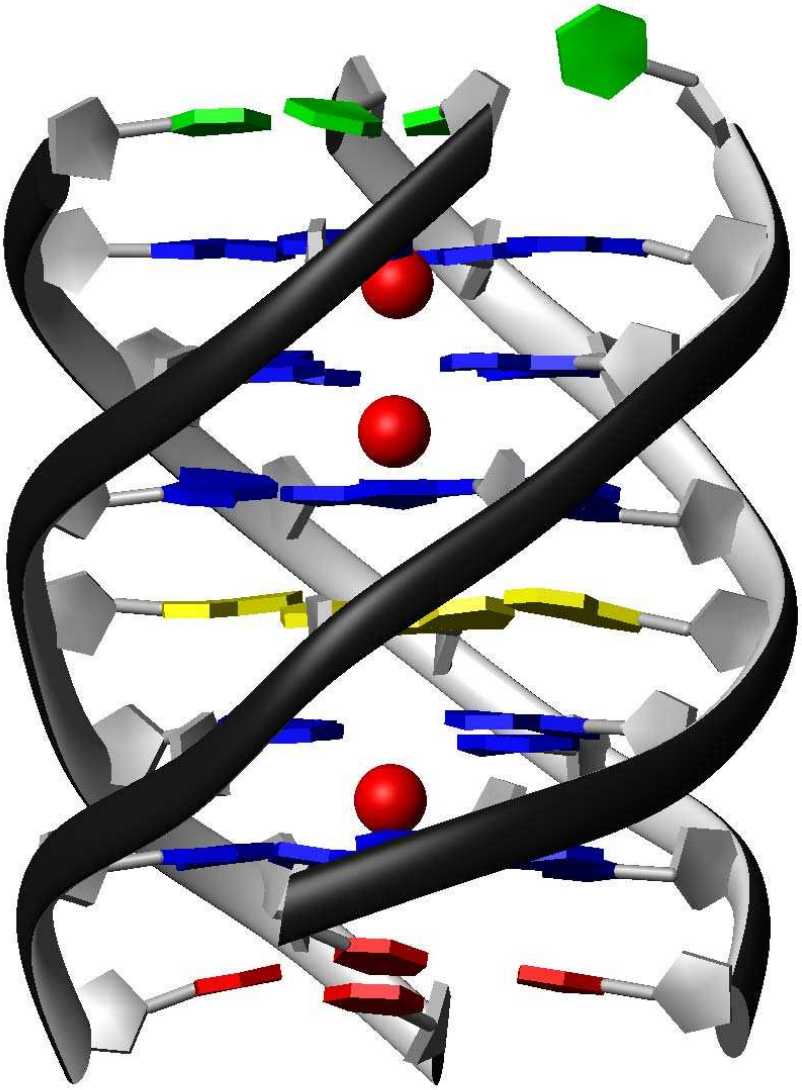


Fig.4.34: Schematic representation of the average minimised structure of d(TGGGAGGC)$_4$ from the last 100 ps of a 1 ns rMD simulation, showing interstitial potassium ions.

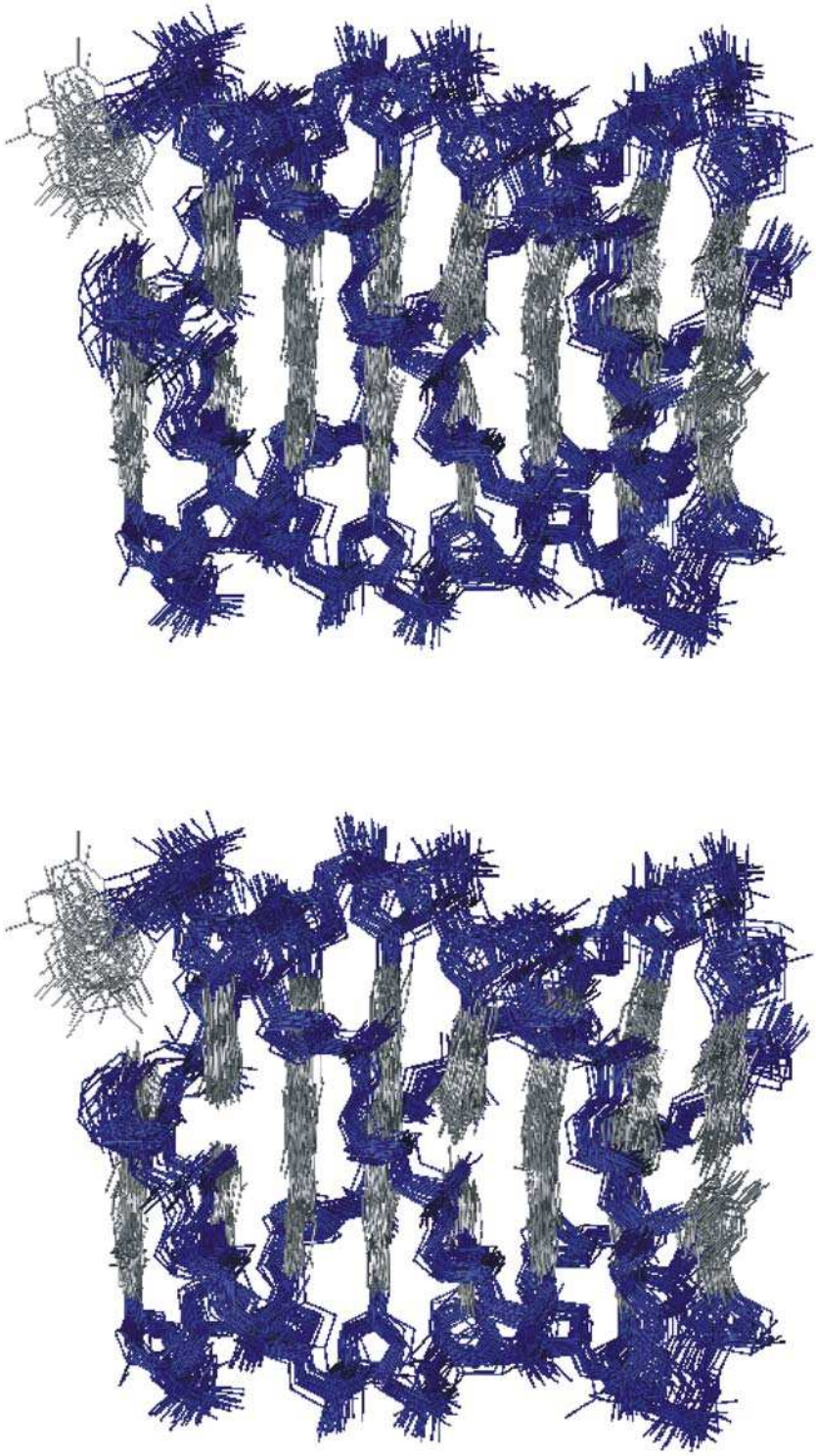


Fig.4.35: Cross-eye stereo view of 25 randomly chosen structures from the last 250 ps of a 1 ns of restrained dynamics simulation of d(TGGGAGGC)$_4$.

5. Determination of Hydration Patterns by rMD Methods

5.1 Introduction: The Role of Solvent in DNA Structure

Since water-DNA interactions are comparable in strength to the noncovalent interactions within the double helix, hydration and helical fine structure are inextricably linked. In fact, water is often regarded in NMR studies as an integral part of DNA (*104*). Solvation has been shown to play an important part in determining the overall conformation of DNA using a variety of experimental techniques, including NMR (*121;135*), X-ray crystallography (*16;56*), and thermodynamics (*33*). *B*-DNA requires the greatest hydration of all the DNA conformers, and requires about 30% water, by weight, to maintain its native conformation in the crystalline state (*144*). Partial dehydration converts it to *A*-DNA by decreasing the free energy required for A-DNA deformation and twisting, which is usefully employed by encouraging supercoiling but eventually leads to denaturation. Further dehydration will result in the least hydrated *Z*-DNA, which has a very narrow minor groove. Hydration is greater and more strongly held around the phosphate groups that run along the inner edges of the major grooves. The water molecules are not permanently situated however, due to the rather diffuse electron distribution of the phosphate groups. Hydration is more ordered and more persistent around the bases with their more directional hydrogen-bonding ability and restricted space. Consequently, water molecules are held relatively strongly around the bases. Because of the regular structure of DNA, hydrating water is held in a cooperative manner along the double helix in both the major and minor grooves. The cooperative nature of this hydration aids both the annealing and unwinding of the double helix.

The so-called *spine of hydration* in the minor groove of DNA is thought to stabilise the structure (*22;23*), particularly in AT tracts (*50;145*), and is known to favour *B*-DNA. Water molecules donate two hydrogen bonds, so bridging

between thymine 2-keto(s) and/or adenine ring N3(s) on sequential bases. This water is fully hydrogen bonded by accepting two further H-bonds from a secondary layer of hydration water, so fixing the primary hydration water more firmly in place (*50;127*). Optimisation of interactions with water molecules has been used to rationalise both the observed transition of B-DNA to A-DNA in water/alcohol mixtures (*54*) and the opposite A- to B-form conversion under fully hydrating conditions (*35*). In aqueous solution, the concentration of water is very high, at approximately 55 M. Statistically, therefore, all exposed sites on the surface of a solute particle can be expected to be close to a water molecule at any given time (*45;46*).

For the purpose of the investigation of hydration by rMD simulation, it is useful to think of water of solvation as falling into one of three categories. Bulk water is defined as that which is not associated with, nor influenced by, the solute. It is essentially the matrix in which the solute molecules are solvated. Weakly associated water molecules are typically found on the surface of the solute, where their accessibility to the bulk solvent gives short residence times at the solute-solvent interface. These weak associations may also be evident in wide grooves if no additional stabilising interactions are present (*242*). Strongly associated solvent molecules are usually isolated from the bulk solvent, in locations such as the core of proteins or deep narrow surface grooves in macromolecular complexes (*78*). In the latter cases, exchange of associated solvent with the bulk is restricted.

The term *residence time* refers to the length of time that a single water molecule is associated with a specific site on a macromolecule. *Occupancy* is a measure of the total time that the site is associated with any water molecule. Though site occupancy and residence time are two distinctly different concepts, there is some correlation between the two: if a single water molecule has a high residence time at a given site, then the occupancy of that site will increase relative to those sites with lower residence times. However, the measure of occupancy makes no distinction between individual water

molecules. Therefore, a promiscuous site, which is occupied sequentially by many different water molecules, may have high occupancy, but a low residence time for each individual water molecule. On the surface of the molecule, and in wide grooves, accessibility to the bulk solvent leads to relatively short residence times. Waters with restricted access to the bulk solvent, such as those in narrow DNA grooves, have elevated residence times, due mainly to their inability to exchange with the bulk. Due to the inability of the water molecules to escape from the groove floor, these sites would also be expected to have higher occupancy. Longer residence times are also observed for water molecules in the core of proteins, but internalised waters are not usually observed in short DNA structures.

NMR can be used to calculate residence times of those water molecules close to detectable DNA protons based on the polarity and magnitude of the observed NOE and ROE signals. The measurement of residence times extends beyond the remit of this study, and so further discussion will be limited to the concept of occupancy.

5.2 Aims & Objectives

The study of DNA structure by NMR methods can be extended beyond the examination of 3-D architecture and molecular dynamics: the involvement of water in DNA structure can also be examined using rMD methods. In this study, the principle objective is to distinguish between water molecules that form bulk solvent and those associated with the DNA, and to investigate those water molecules that are potentially interacting with DNA. We aim to achieve this by integrating the water population over the duration of molecular dynamics simulations to identify water occupancy that is significantly greater than in bulk water.

This study attempts to compare the hydration patterns of the NMR-derived rMD structures of nogalamycin-d(TGATCA) and $d(TG_4T)_4$ with those from

previously published x-ray structures (*170;212*) with a view to assessing if those waters that are strongly associated with the macromolecule, and not freely exchanging with the bulk solvent or affected by the requirements of crystal packing, are similar in both solid- and solution-states.

By way of contrast, it is also of interest to know to what degree those solvent molecules that weakly associate with the macromolecular surface differ between solid- and solution-states, and to what degree crystal-packing influences the distribution of water on the external surface of each of the DNA complexes studied.

5.3 Materials & Methods

5.3.1 rMD study of Complex Hydration

Molecular dynamics using explicit solvent models is a useful tool for probing solvent interactions in macromolecular systems. Water molecules in the rMD simulation are represented using the Jorgensen TIP3P parameterisation (*105*) within AMBER 6.0. It is a three-particle model that includes the oxygen atom and two hydrogen atoms. There is no internal flexibility, such as stretching and bending, allowed for in this model. Interactions between water molecules are described by a Lennard-Jones potential (*8*) between oxygen atoms and electrostatic contributions between all atoms. TIP3P solvent molecules have been shown to accurately reproduce experimentally derived parameters of water, including density and co-ordination (*137*).

5.3.2 Calculation of Water Occupancy

Calculations of water occupancy were carried out on snapshots taken at 1 ps intervals of a 1ns restrained molecular dynamics simulation of the complex of $d(TGATCA)_2$ with nogalamycin. Each snapshot was normalised with respect to a standardised PDB co-ordinate set of the model system taken from an arbitrary point during the molecular dynamics run. This allowed the DNA atoms from each snapshot to be superimposed within the limits of the applied

distance restraints. This procedure has been shown to give rise to atom misplacements of less than 0.5 $Å^3$ relative to the standard PDB (*242*).

For each snapshot, the space around the macromolecular complex was enclosed in a cubic grid, equal in size to that of the periodic box, and centred on the centre of mass of the complex. The periodicity of the grid was 1 Å in each dimension. The number of water molecules in each 1 $Å^3$ cube was then measured, giving a numerical measure of water occupancy around the solute (Fig.5.1(b)). A cumulative total for each cubic angstrom of space around the complex could then be calculated by summing the number of water molecules in each superimposed cube for each snapshot of the entire rMD simulation (Fig.5.1(c)).

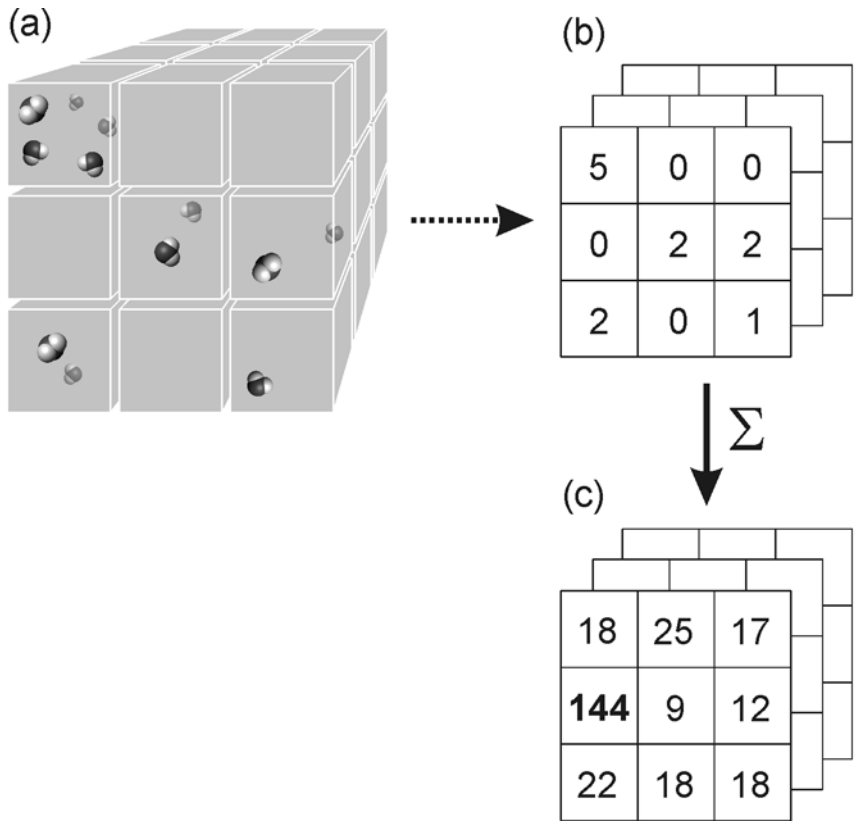


Fig.5.1: Schematic representation of the method used to determine water occupancy from rMD simulation: The periodic box is divided into 1 $Å^3$ cubes (a). The number of water molecules in each cube is summed for each snapshot of the simulation (b), and the values from each snapshot are subsequently added to give a figure for total water occupancy (c). Using this method, higher occupancy sites can readily be identified.

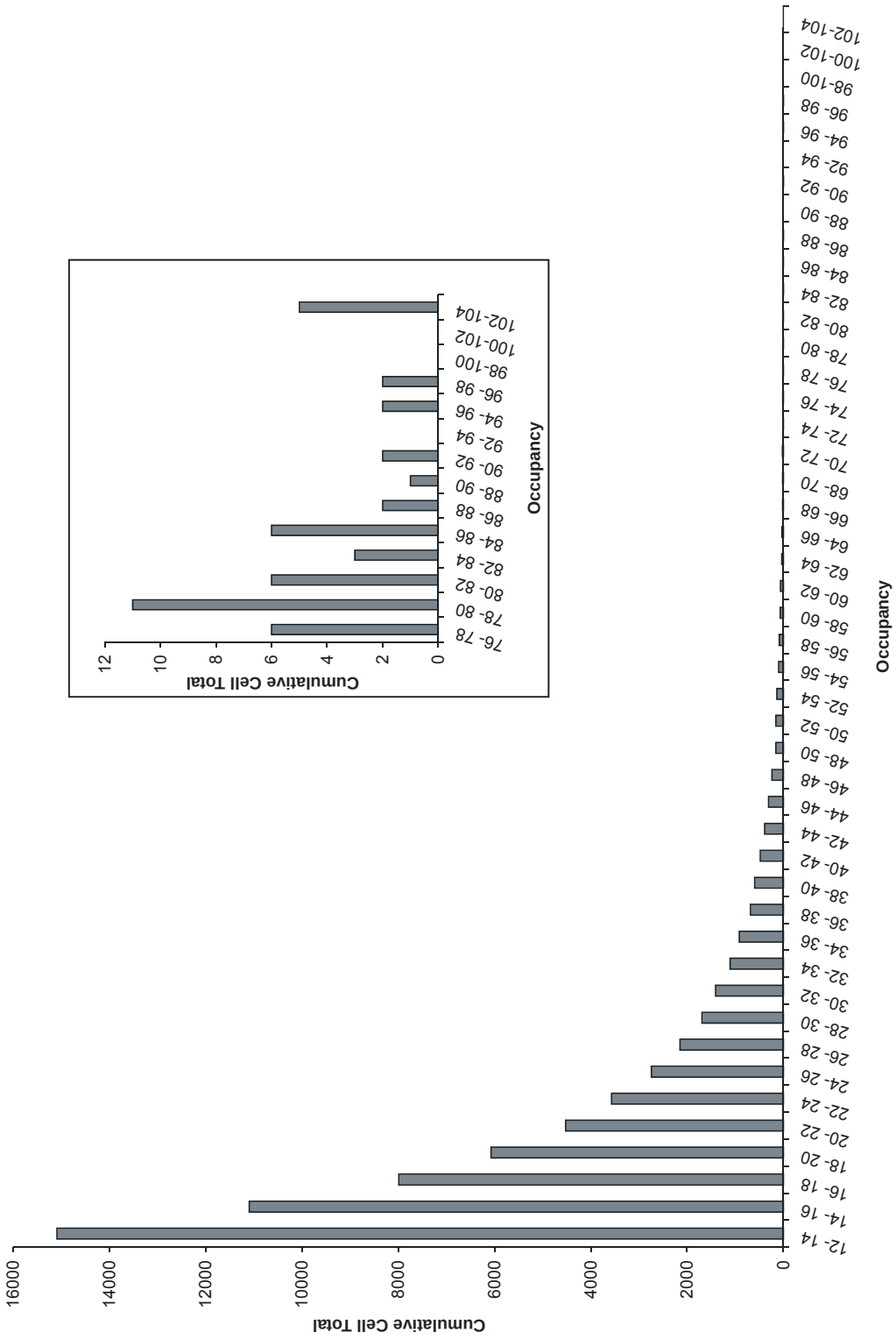


Fig.5.2: Graph of cumulative cell numbers as a function of occupancy for nogalamycin-d(TGATCA), showing the number of cells for each given occupancy value. Significant occupancy, defined by the cut-off value, shown inset.

In order to discriminate between bulk solvent and associated water molecules, it was necessary to define at what level of occupancy an individual cell exceeds background levels. For the periodic box dimension of the $d(TGATCA)_2$ complex, for example, which measures 38x40x43 Å, there are over 65000 individual cubes. The box contains approximately 4000 water molecules and 1000 snapshots were taken over the duration of the simulation. It is, therefore, reasonable to assume that every 1 $Å^3$ cube that is accessible to the solvent would contain a water molecule during some part of the dynamics run. To discriminate what constituted background occupancy, the distribution of cumulative occupancy values was tabulated and parsed (Fig.5.2), and a cut-off value was subsequently selected for each model. All cubes with cumulative totals of less than the cut-off were treated as bulk solvent. The cut-off value is specific to the system being studied, and should be chosen to give a small enough sample of cells to accurately represent water occupancy while minimising the effect of baseline noise. Although the threshold value is arbitrary, a value that selects only 2% of the total number of cells gives significant occupancy information while avoiding background untidiness.

5.3.3 Starting Structures for $d(TG_4T)_4$

In order to compare hydration patterns between solid and solution states, a starting model of the quadruplex structure of $d(TGGGGT)_4$ was generated using the atom coordinates from the crystal structure (PDB ID: 352D). The 5'- and 3'-terminal thymine bases were reoriented to maintain a separation between consecutive planes and a helical twist consistent with B-form DNA along the entire length of the quadruplex. Hydrogen atom positions were calculated and atoms added using the LEaP module of AMBER 6.0. All sodium ions were removed, and a central core of five potassium ions, coincident with the helical axes and equidistant from each tetrad, was added so that each of the three core metal ions was bipyrimidally coordinated by eight guanine carbonyl atoms. Potassium counter-ions were added with LEaP to negate the negative charge on the backbone, and the model was solvated to

within a distance of 5 Å using boxes of 216 TIP3P H_2O molecules. The structure was subsequently minimised and allowed to undergo 2 ns of unrestrained MD simulation using the protocol described in Section **4.3.5**.

5.4 Results & Discussion

5.4.1 Density Calculations for Nogalamycin-d(TGATCA)

From the 2 ns trajectory of the d(TGATCA)-nogalamycin complex, a cut-off of 78 water molecules per 1 $Å^3$ cube was selected to distinguish between bulk solvent molecules and those with significant site occupancy. Cells with a cumulative total less than 78 (>98%) were discarded. The analysis highlighted 40 cells with significant occupancy, which were subsequently overlaid on the average DNA structure for analysis using MolMol.

The crystal structure of d(TGATCA)-nogalamycin (PDB ID: 182D) was displayed using MolMol. To facilitate comparison between crystal and solution structures, water molecules within 3.8 Å of the VDW surface of the complex were selected. Those waters selected included those in direct contact with the structure while largely excluding the outer layer of water molecules, which are largely an artefact of crystal packing.

5.4.2 Comaprison of rMD and X-ray Data for Nogalamycin-d(TGATCA)

The crystal structure (*212*) contains two DNA-nogalamycin complexes, one spermine molecule, two acetates, one sodium ion, and 97 water molecules in the asymmetric unit, although the spermine and acetates do not possess any biological relevance to the drug binding. Although each DNA strand has the same sequence, differences in hydration patterns of both complexes exist, due to different solvent channels generated in each complex by crystal packing. There are also significant differences in sugar-phosphate geometry between the two complexes. This illustrates, albeit in a distinctly non-dynamic manner, the inherent flexibility of DNA, which allows conservation of water-mediated

contacts in both conformers. The existence of two conformationally-distinct structures, the absence of stabilising ions, and the presence of additional molecules, all to facilitate crystal packing, highlights the shortcomings of x-ray crystallography in the study of large molecules in a biologically-relevant environment. The solvent environment allows greater flexibility in the nature of interactions stabilising the complex through direct and solvent-mediated effects.

Each of the four nogalamycin molecules makes at least two direct hydrogen bonds to the DNA, all of which are conserved in the NMR solution structure. One H-bond between the *O15* proton of nogalamycin (Fig.5.3) and the *N7* nitrogen of guanine in the major groove (O···N, 2.8 Å) is maintained throughout the rMD simulation. In the minor groove, a second H-bond between *O7* of nogalamycin and the guanine amino proton (H21) can be observed. The hydroxyl group at position 6 on the drug chromophore has a donor-hydrogen that can be accepted by either the *O7* of the drug or the O2 of cytosine. The latter interaction can exist only in the absence of the intramolecular H-bond. Should this be the case, the *O7* oxygen is also free to form a strong H-bond (3.0 Å, θ=12 º) with the *N2* of the complementary guanine. It seems reasonable that the formation of two strong intermolecular bonds is more favourable than one intramolecular bond, and this is borne out by H-bond residency studies carried out over 1ns of rMD, where the inetrmolecular *O7*···H–*N2* bond is occupied for 63% of the simulation, compared with 31% for the equivalent intramolecular bond. No H-bonding is evident for 6% of the duration of the simulation.

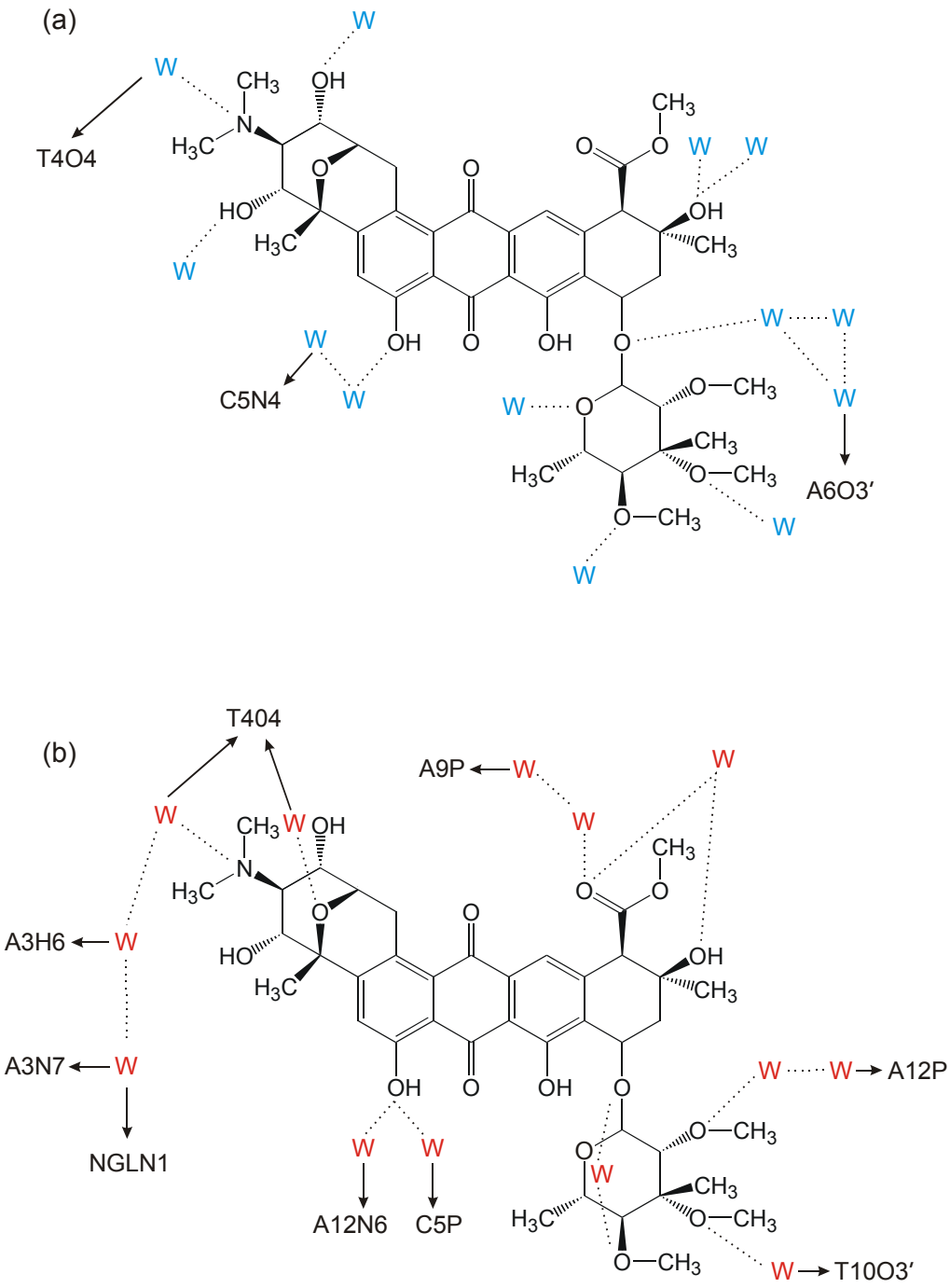


Fig.5.3: Direct water contacts and water-mediated interactions in the X-ray crystal structure (a) and NMR-derived rMD structure (b) of nogalamycin-d(TGATCA). Water molecules are depicted by the letter 'W', drug-water interactions by dotted lines, and water-DNA interactions by solid arrows.

Potential hydrogen bonding interactions mediated by water molecules were determined by a 3.8 Å distance cut-off and a 34° angular cut-off between the drug/DNA heavy atom and the centre of mass of the water molecule.

In the major groove of the crystal structure complex, there are many solvent-mediated interactions between nogalamycin and the DNA at the intercalation site. The majority of water-mediated contacts are to the bases of DNA, rather than to the sugar-phosphate backbone. There are no contacts, water-mediated or otherwise, between nogalamycin molecules in the crystal structure of nogalamycin-d(TGATCA). However, in the X-ray structure of nogalamycin-d(TGTACA) (*210*), a chain of four water molecules in the major groove links the amino nitrogen atoms of each drug molecule in the same complex (Fig.5.4). Why altering the order of the nucleotides between the two TpG binding should affect these water-mediated drug-drug interactions remains unclear (*210*). Three water molecules that fulfil the criteria for H-bond formation similarly link the amino nitrogens in the NMR-derived structure. In each case these water molecules can potentially form stabilising H-bonds to electronegative heavy base atoms on the groove floor (Fig.5.3(b)). The dimethylamino group of the bicyclic sugar is stabilised in the same orientation in the solid state by a water-mediated interaction between the *N1* atom and the N6 of the central adenine. The nogalamycin model used for the solvent structure was initially derived from the x-ray data, and its structure is maintained throughout the simulation by a solvent-mediated interaction between *N1* and the *O4* of the central thymine. Though it is also possible for the bridging water to form a H-bond with T4N6, the residency study strongly favours *O4* as the receptor.

One of the two crystal structure complexes possesses extended minor groove solvent networks. Water-mediated contacts between nogalose and adenine *N2* and *O3′* atoms stabilise the minor groove side of the complex. The solution structure is stabilised in the minor groove by bridging water molecules between the nogalose 3'-meothoxy oxygen and phosphate backbone atoms on the central thymine and terminal adenine residues, and by a similar interaction between the nogalose 1'-oxygen and the terminal adenine *O3′* oxygen (Fig.5.5). The water oxygen atom is positioned 3.95 (±0.2) Å from each of the

drug and DNA heavy atoms, and makes an angle of 69.0 (±0.5)° between them. Additionally, a water molecule acts as a bridgehead between oxygen atoms on opposite sides of the nogalose ring, providing extra stability in this region.

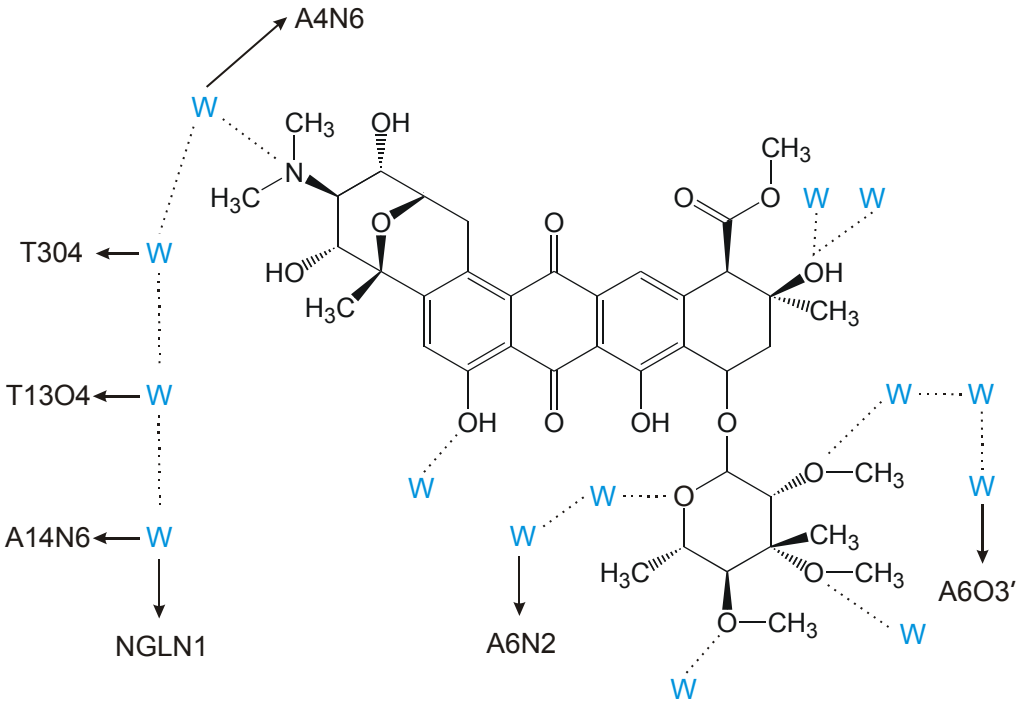


Fig.5.4: Direct water contacts and water-mediated interactions in the X-ray crystal structure of nogalamycin-d(TGTACA). Image adapted from (*210*).

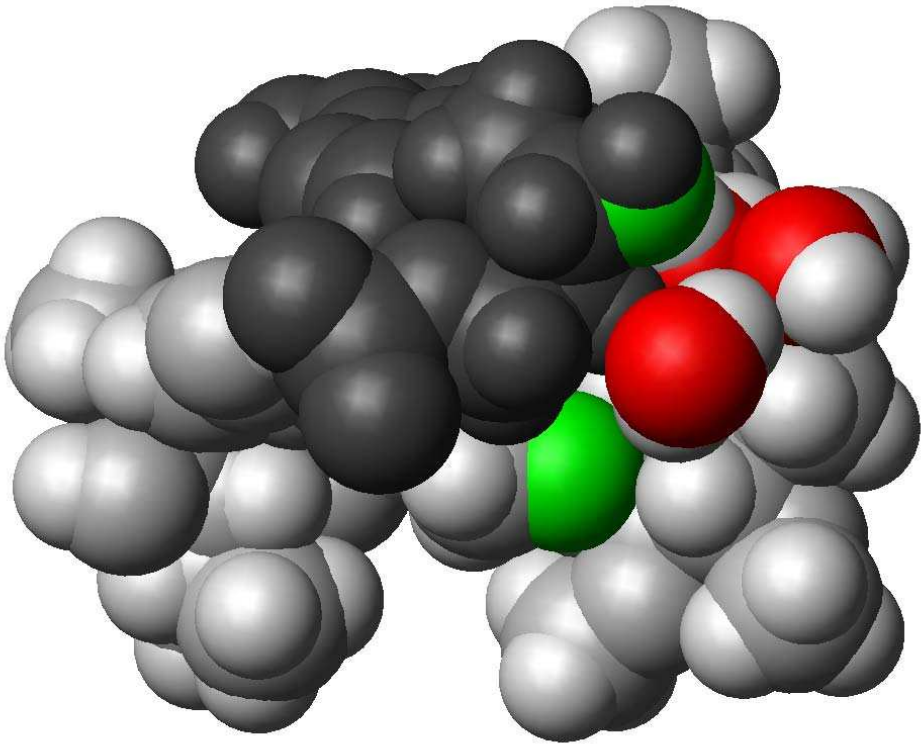


Fig.5.5: An example of a water-mediated drug-DNA interaction: between the nogalose 1'-oxygen and the terminal adenine O3' oxygen, across the minor groove. Nogalamycin is shown in grey and DNA in black. Interacting atoms are shown in green.

The spines of water molecules lining both major and minor grooves of the crystal structure (Fig.5.6(a)) are not evident in the solution structure as initially examined. However, if the cumulative water cut-off value is reduced to 68 molecules per 1 $Å^3$ cube, the number of selected cubes increases to 105, and these hydration tracts become evident (Fig.5.6(b)). To eliminate the incidence of cubes being highlighted in the bulk solvent, only the 88 waters within 5 Å of the VDW surface of the complex were selected at the new, lower cut-off. As expected, those sites with the highest occupancies were located in the DNA grooves and around the groove binding sugars of the intercalated drugs, as shown using a graduated scale of occupancy in Fig.5.6(b). This can be attributed to the inability of the water molecules in these sites to easily exchange with the bulk solvent.

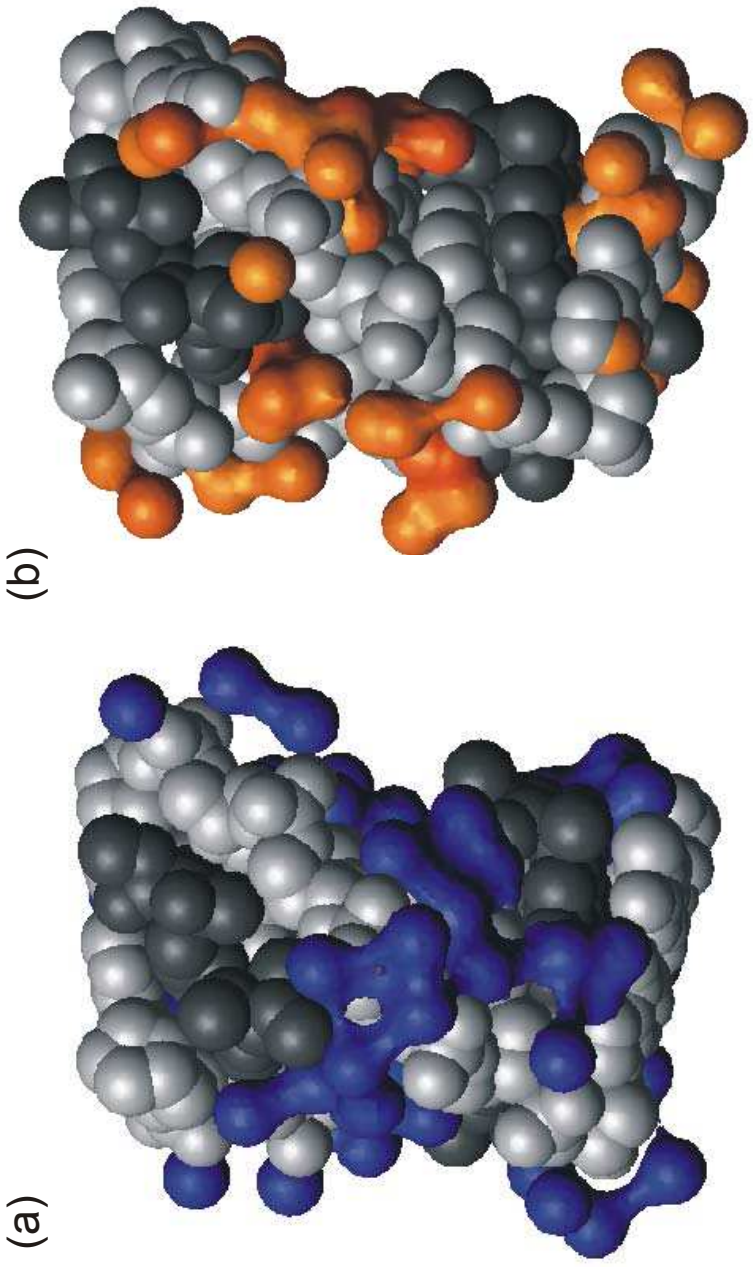


Fig.5.6: Comparison of X-ray determined (a) and rMD calculated (b) hydration patterns for the complex of nogalamycin with d(TGATCA)$_2$. DNA is shown in grey, and nogalamycin in black. Water in shown in blue for the X-ray structure. A graduated scale from orange (lowest occupancy) to yellow (highest occupancy) is used to display water on the rMD structure. View is to highlight hydration spine in DNA grooves.

5.4.3 Density Calculations for d(TG4T)$_4$

Phillips *et. al.* (*170*) have reported the crystal structure at 0.95 Å of the parallel-stranded quadruplex formed by d(TG$_4$T)$_4$ in the presence of sodium ions at 100 K. The asymmetric unit of the d(TGGGGT)$_4$ is inordinately complex (Fig.5.7). It is comprised of four independent quadruplexes, which can be grouped into two pseudo-equivalent sets. In each set, a pair of tetraplexes is coaxially stacked in a 5' to 5' orientation. This generates a ladder of eight guanine tetrads, as the terminal thymines on each pair do not stack onto the end of the quadruplexes. Calcium ions at the interface serve to stabilise contacts between adjacent tetrads.

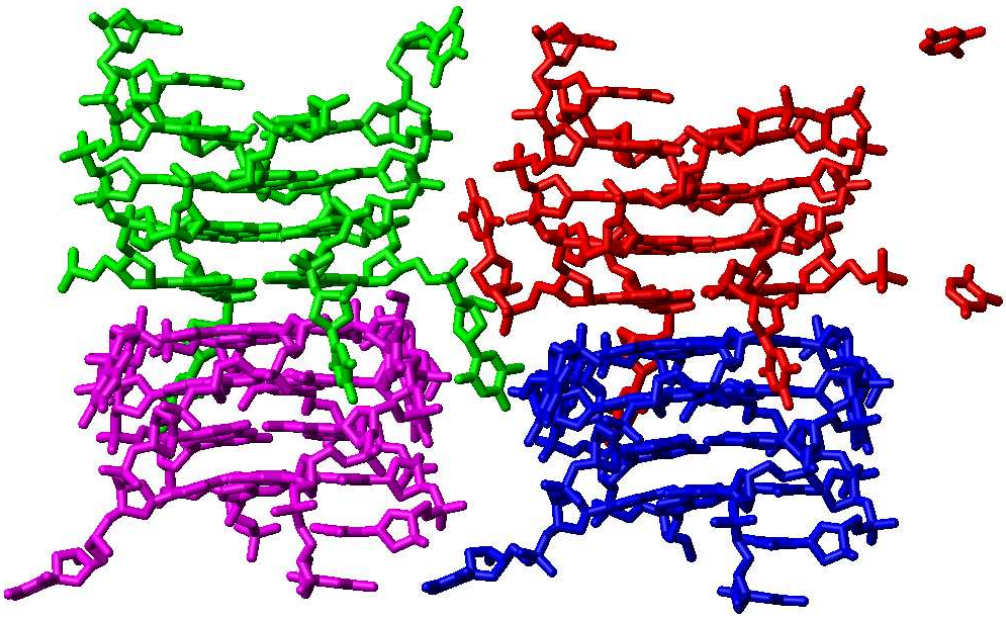


Fig.5.7: Contents of the asymmetric unit of d(TGGGGT)$_4$ (PDB ID: 352D).

5.4.2 Comaprison of MD and X-ray Data for d(TG$_4$T)$_4$

To facilitate comparison with X-ray hydration patterns, the cumulative water cut-off value of the MD structure was set to 84 per 1 Å^3 cube, which selected 96 cubes. To isolate only those cubes that interact directly with the quadruplexes only those with a centre of gravity within 4 Å of the DNA surface were considered (Fig.5.8(b)). This highlighted 42 water molecules for comparative study. A 5Å upper distance threshold was also applied to the

crystal structure leading to the selection of 68 water molecules within this distance of a single quadruplex isolated from the asymmetric unit (Fig.5.8 (a)).

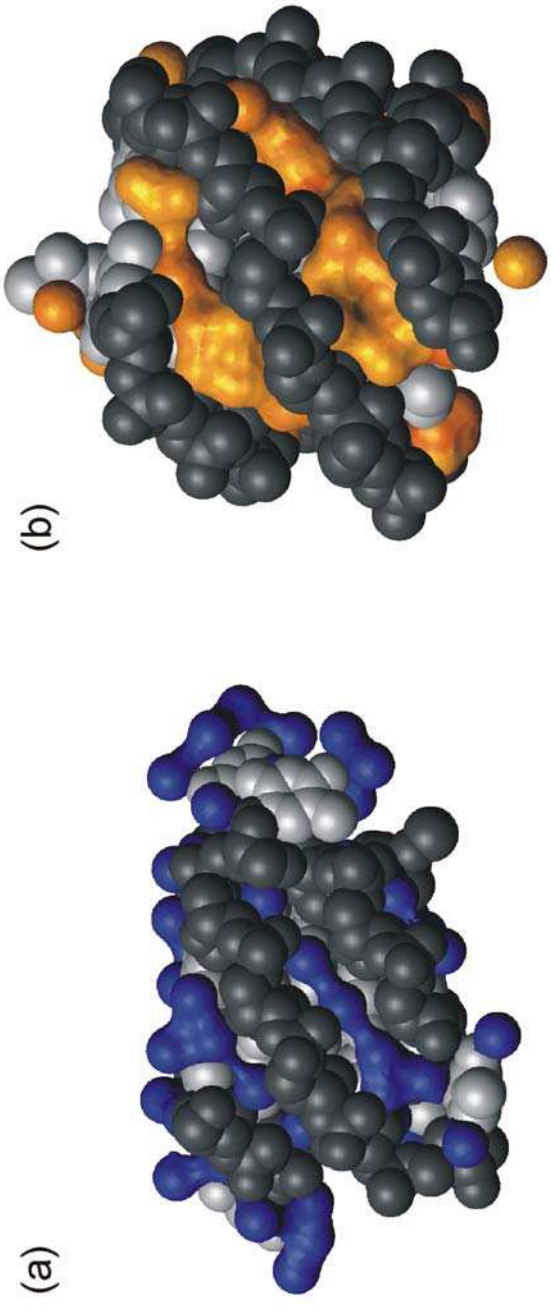


Fig.5.8: Comparison of X-ray determined (a) and MD calculated (b) hydration patterns for $d(TG_4T)_2$. DNA bases are shown in grey, and sugar & backbone atoms in black. Water is shown in blue for the X-ray structure, while a graduated scale from orange (lowest occupancy) to yellow (highest occupancy) is used to display water on the rMD structure. Only those water molecules within 5 Å of the groove floor of each molecule are highlighted.

All four grooves in the crystal structure were found to be almost equivalent with respect to hydration patterns (*170*). Water molecules were shown to cluster around individual phosphate groups, but there was no evidence of water bridging adjacent phosphate groups, as occurs in *A* and *Z* forms. The average intra-strand separation was shown to be 6.6 Å, similar to that of *B*-DNA. The four grooves are relatively narrow, measuring between 2.3 and 3.3 Å in width, as measured by cross-strand phophate separations between non-adjacent tetrads (Fig.5.9). This is in good agreement with the MD structure, where analagous measurements showed the groove width to be in the range $2.3 \leq x \leq 3.8$ Å. Hence, they are favourable binding sites for water molecules, which in both X-ray and MD structures form hydrogen bonds with the exposed guanine *N2* amino group, its heterocyclic *N3*, and with each other to create a well-supported network. The crystal structure was noted to have water molecules deep in each groove near each of the *O4'* oxygens. These molecules accept H-bonds from the *N2* atoms of the guanine bases and are linked together in a spine by a second layer of water molecules. The mean distance between the guanine *N2* and water heavy atoms in this case is 2.95 (±0.05) Å. In the solution structure, the analogous distance is 4.5 (±0.1) Å (Fig.5.10). However, the distance between the hetrocyclic *N3* atom and the water oxygen is only 3.1 (±0.1) Å. This agrees well with the previous observation that *N3* is preferred over *N2* as a H-bond acceptor in *B*-like DNA, while *N2* is preferred over *N3* under poorer hydrating conditions (*188*). In this case (Fig.5.10), the water acts as a bridging molecule between the guanine *N3* and furanose *O4'* atoms, presumably adding stability to the G-tetrad. The groove-embedded water molecules in both x-ray and NMR structures are sheltered from the bulk water by enveloping layers of water molecules. These outer layers are, in the former case, affected by the restrictions of crystal packing and, as such, the surface water molecules are not amenable to direct comparison with those in the solution structure. However the patterns of the groove lining waters are similar in both cases, forming the same H-bonds, and aligning in the same hydration spine in both cases (Fig.5.8).

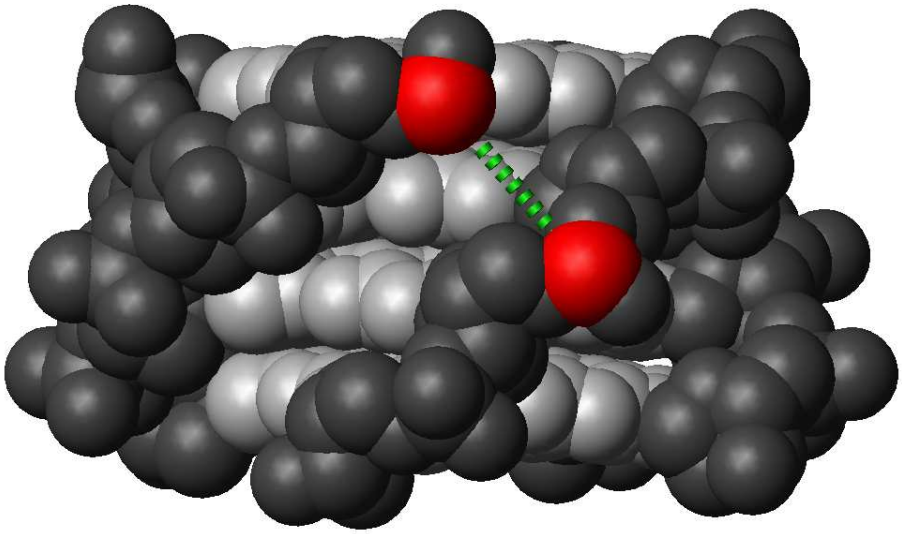


Fig.5.9: Cross-strand phosphate separations (green) between *i* and (*i*+3) tetrads were used as a measure of groove widths. Measurements were made between the centres of mass of the phosphate groups, and their VDW radii were subsequently subtracted to give a measure of groove width. Oxygens adjacent to the highlighted phosphorus atoms have been removed for clarity.

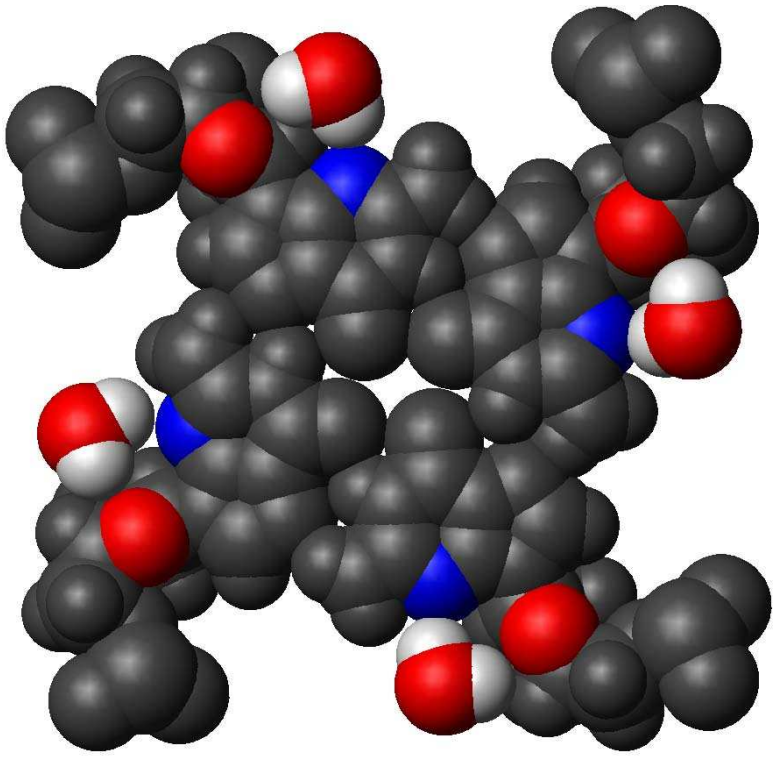


Fig.5.10: Interaction of water molecules in the groove of the solution structure of $d(TG_4T)_4$ with tetrad-forming guanines. Nucleotide is shown in black, except *N2* nitrogen (blue) and *O4'* (red).

5.5 Conclusions

Comparison of the rMD and x-ray structures reveals that both are extensively hydrated, with particularly high water occupancy in discreet regions of the complexes. The hydration patterns in the major and minor grooves of both crystal and solution structures of the complex of $d(TGATCA)_2$ with nogalamycin are broadly similar. Though there is no water mediated contact between drug molecules in the crystal structure of nogalamycin-d(TGATCA), in its solution structure and also in the crystal structure of nogalamycin-d(TGTACA), the dimethylamino groups of the aminoglucose sugars are connected by a chain of water molecules in the major groove, while the minor groove is stabilised by water interactions, both between the drug and the DNA backbone, and between adjacent methoxy groups on the nogalose.

Groove lining *spines of hydration* are also present in both x-ray crystal structure and NMR-derived solution structure of $d(TGGGGT)_4$. In each case these molecules accept H-bonds from the *N3* atoms of the guanine bases and are linked together in a spine by a second layer of water molecules, which serves to isolate them from the bulk solvent. It was not possible to compare hydration patterns at the terminal thymidine residues, as these are displaced by stabilising calcium ions at the interface between adjacent molecules in the asymmetric unit of the crystal structure, unlike the solution structure, in which they form T-tetrads (Section **4.4.1**).

Occupancy was observed to be highest in the DNA grooves for NMR–derived structures of both duplex and quadruplex structures, which is consistent with that observed in previous studies of macromolecules in aqueous media. Additionally, for both the duplex and the quadruplex studied, hydration patterns outside the DNA grooves do not correlate well between x-ray and NMR-derived structures. It is reasonable to assume that these waters are ordered to benefit crystal packing in the solid state, and so bear little or noresemblance to the solution structure.

6. References

1. IMS World Review. 2004.

Ref Type: Internet Communication

2. Aboulela, F., Murchie, A. I. H., and Lilley, D. M. J. (1992) *Nature 360*, 280-282.

3. Adhin, M. R., Alblas, J., and Vanduin, J. (1990) *Biochimica et Biophysica Acta 1050*, 110-118.

4. Agrawal, S., Temsamani, J., Galbraith, W., and Tang, J. Y. (1995) *Clinical Pharmacokinetics 28*, 7-16.

5. Anonymous (1971) *Nature New Biology 233*, 223.

6. Arai, K., Low, R., Kobori, J., Shlomai, J., and Kornberg, A. (1981) *Journal of Biological Chemistry 256*, 5273-5280.

7. Arora, S. K. (1983) *Journal of the American Chemical Society 105*, 1328-1332.

8. Atkins, P. W. (1993) in *Physical Chemistry* pp 643-678, Oxford University Press, Oxford.

9. Bailly, C. (2000) *Current Medicinal Chemistry 7*, 39-58.

10. Baleja, J. D., Moult, J., and Sykes, B. D. (1990) *Journal of Magnetic Resonance 87*, 375-384.

11. Bax, A. and Davis, D. G. (1985) *Journal of Magnetic Resonance 65*, 355-360.

12. Bax, A. and Davis, D. G. (1985) *Journal of Magnetic Resonance 63*, 207-213.

13. Bax, A. and Freeman, R. (1981) *Journal of Magnetic Resonance 44*, 542-561.

14. Bazarov, A. V., Adachi, S., Li, S. F., Mateyak, M. K., Wei, S., and Sedivy, J. M. (2001) *Cancer Research 61*, 1178-1186.

15. Berger, J. M., Gamblin, S. J., Harrison, S. C., and Wang, J. C. (1996) *Nature 379*, 225-232.

16. Berman, H. M. (1994) *Current Opinion in Structural Biology 4*, 345-350.

17. Berman, H. M., Bourne, P. E., and Westbrook, J. (2004) *Current Proteomics 1*, 49-57.

18. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., and Tasumi, M. (1977) *Journal of Molecular Biology 112*, 535-542.

19. Blackburn, E. H. (1991) *Trends in Biochemical Sciences 16*, 378-381.

20. Blackburn, E. H. and Szostak, J. W. (1984) *Annual Review of Biochemistry 53*, 163-194.

21. Boon, E. M., Livingston, A. L., Chmiel, N. H., David, S. S., and Barton, J. K. (2003) *Proceedings of the National Academy of Sciences of the United States of America 100*, 12543-12547.

22. Bostock-Smith, C. E., Harris, S. A., Laughton, C. A., and Searle, M. S. (2001) *Nucleic Acids Research 29*, 693-702.

23. Bostock-Smith, C. E., Laughton, C. A., and Searle, M. S. (1998) *Nucleic Acids Research 26*, 1660-1667.

24. Bowers, J., Tran, P. T., Liskay, R. M., and Alani, E. (2000) *Journal of Molecular Biology 302*, 327-338.

25. Bunin, B. A. and Ellman, J. A. (1992) *Journal of the American Chemical Society 114*, 10997-10998.

26. Bunin, B. A., Plunkett, M. J., and Ellman, J. A. (1994) *Proceedings of the National Academy of Sciences of the United States of America 91*, 4708-4712.

27. Caceres-Cortes, J. and Wang, A. H. J. (1996) *Biochemistry 35*, 616-625.

28. Caetanoanolles, G. and Gresshoff, P. M. (1994) *Bio-Technology 12*, 619-623.

29. Case, D. A., Caldwell, J. W., Cheatham III, T. E., Ross, W. S., Simmerling, C. L., Darden, T. A., Merz, K. M., Stanton, R. V., Cheng, A. L., Vincent, J. J., Crowley, M., Tsui, V., Radmer, R. J., Duan, Y., Pitera, J., Massova, I., Siebel, G. L., Singh, U. C., Weiner, P. K., and Kollman, P. A. AMBER 6. 1999. University of California, San Francisco.

Ref Type: Computer Program

30. Case, D. A. and Xu, X. P. SHIFTS. [4.1.1]. 2002. Department of Molecular Biology, The Scripps Research Institute.

Ref Type: Computer Program

31. Cate, J. H., Gooding, A. R., Podell, E., Zhou, K. H., Golden, B. L., Kundrot, C. E., Cech, T. R., and Doudna, J. A. (1996) *Science 273*, 1678-1685.

32. Cate, J. H., Hanna, R. L., and Doudna, J. A. (1997) *Nature Structural Biology 4*, 553-558.

33. Chalikian, T. V., Sarvazyan, A. P., Plum, G. E., and Breslauer, K. J. (1994) *Biochemistry 33*, 2394-2401.

34. Chazin, W. J., Wuthrich, K., Hyberts, S., Rance, M., Denny, W. A., and Leupin, W. (1986) *Journal of Molecular Biology 190*, 439-453.

35. Cheatham, T. E. and Kollman, P. A. (1996) *Journal of Molecular Biology 259*, 434-444.

36. Chen, H. F., Liu, X. C., and Patel, D. J. (1996) *Journal of Molecular Biology 258*, 457-479.

37. Cheong, C., Varani, G., and Tinoco, I. (1990) *Biophysical Journal 57*, A448.

38. Chou, S.-H., Chin, K.-H., and Chen, F.-M. (2002) *Proceedings of the National Academy of Sciences of the United States of America 99*, 6625-6630.

39. Chou, S. H., Tseng, Y. Y., and Wang, S. W. (1999) *Journal of Molecular Biology 287*, 301-313.

40. Colgrave, M. L. Electrospray Mass Spectrometry and Nuclear Magnetic Resonance Spectrometry of DNA Interactive Agents. 2002. University of Wollongong.
Ref Type: Thesis/Dissertation

41. Colgrave, M. L., Williams, H. E. L., and Searle, M. S. (2002) *Angewandte Chemie-International Edition 41*, 4754-4756.

42. Collier, D. A., Neidle, S., and Brown, J. R. (1984) *Biochemical Pharmacology 33*, 2877-2880.

43. Colman, P. M. (1994) *Protein Science 3*, 1687-1696.

44. Colman, P. M., Varghese, J. N., and Laver, W. G. (1983) *Nature 303*, 41-44.

45. Conte, M. R., Conn, G. L., Brown, T., and Lane, A. N. (1996) *Nucleic Acids Research 24*, 3693-3699.

46. Conte, M. R., Conn, G. L., Brown, T., and Lane, A. N. (1997) *Nucleic Acids Research 25*, 2627-2634.

47. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. (1995) *Journal of the American Chemical Society 117*, 5179-5197.

48. Cowing, D. W., Bardwell, J. C. A., Craig, E. A., Woolford, C., Hendrix, R. W., and Gross, C. A. (1985) *Proceedings of the National Academy of Sciences of the United States of America 82*, 2679-2683.

49. Dejaegere, A., Bryce, R. A., and Case, D. A. (1999) in *Modelling NMR Chemical Shifts* (Facelli, J. C. and Dedios, A. C., Eds.) pp 194-206, American Chemical Society, New York.

50. Denisov, V. P., Carlstrom, G., Venu, K., and Halle, B. (1997) *Journal of Molecular Biology 268*, 118-136.

51. Dewitt, S. H., Kiely, J. S., Stankovic, C. J., Schroeder, M. C., Cody, D. M. R., and Pavia, M. R. (1993) *Proceedings of the National Academy of Sciences of the United States of America 90*, 6909-6913.

52. Diekman, S. (1989) *Journal of Molecular Biology 205*, 787-791.

53. Doroshow, J. H. (1991) *New England Journal Of Medicine 342*, 843-845.

54. Drew, H. R. and Dickerson, R. E. (1981) *Journal of Molecular Biology 151*, 535-556.

55. Ducruix, A. and Giege, R. (1992) in *Crystallisation of Nucleic Acids and Proteins* (Ducruix, A. and Giege, R., Eds.) pp 73-98, Oxford University Press, Oxford.

56. Eisenstein, M. and Shakked, Z. (1995) *Journal of Molecular Biology 248*, 662-678.

57. Elias, P. and Lehman, I. R. (1988) *Proceedings of the National Academy of Sciences of the United States of America 85*, 2959-2963.

58. Erlitzki, R. and Fry, M. (1997) *Journal of Biological Chemistry 272*, 15881-15890.

59. Escaja, N., Gelpi, J. L., Orozco, M., Rico, M., Pedroso, E., and Gonzalez, C. (2003) *Journal of the American Chemical Society 125*, 5654-5662.

60. Facchini, L. M. and Penn, L. Z. (1998) *FASEB Journal 12*, 633-651.

61. Fang, G. W. and Cech, T. R. (1993) *Cell 74*, 875-885.

62. Fedoroff, O. Y., Salazar, M., Han, H. Y., Chemeris, V. V., Kerwin, S. M., and Hurley, L. H. (1998) *Biochemistry 37*, 12367-12374.

63. Feigon, J., Wright, J. M., Leupin, W., Denny, W. A., and Kearns, D. R. (1982) *Journal of the American Chemical Society 104*, 5540-5541.

64. Fox, K. R. and Waring, M. J. (1986) *Nucleic Acids Research 14*, 2001-2014.

65. Fox, K. R. and Waring, M. J. (1986) *Biochemistry 25*, 4349-4356.

66. Freemont, P. S., Friedman, J. M., Beese, L. S., Sanderson, M. R., and Steitz, T. A. (1988) *Proceedings of the National Academy of Sciences of the United States of America 85*, 8924-8928.

67. Fujiwara, A. and Hoshino, T. (1986) *CRC Critical Reviews in Biotechnology 3*, 133-157.

68. Furka, A., Sebestyen, F., Asgedom, M., and Dibo, G. (1991) *International Journal of Peptide and Protein Research 37*, 487-493.

69. Gallagher, C. T. and Searle, M. S. (2003) *Chemical Communications* 1814-1815.

70. Gao, X. L. and Patel, D. J. (1990) *Biochemistry 29*, 10940-10956.

71. Gavathiotis, E., Heald, R. A., Stevens, M. F. G., and Searle, M. S. (2001) *Angewandte Chemie-International Edition 40*, 4749-+.

72. Gavathiotis, E., Heald, R. A., Stevens, M. F. G., and Searle, M. S. (2003) *Journal of Molecular Biology 334*, 25-36.

73. Gavathiotis, E. and Searle, M. S. (2003) *Organic & Biomolecular Chemistry 1*, 1650-1656.

74. Geierstanger, B. H., Cole, T. D., and Wemmer, D. E. (1995) *Journal of Cellular Biochemistry* 59.

75. Geierstanger, B. H. and Wemmer, D. E. (1995) *Annual Review of Biophysics and Biomolecular Structure 24*, 463-493.

76. Geinguenaud, F., Liquier, J., Brevnov, M. G., Petrauskene, O. V., Alexeev, Y. I., Gromova, E. S., and Taillandier, E. (2000) *Biochemistry 39*, 12650-12658.

77. Gellert, M., Lipsett, M. N., and Davies, D. R. (1962) *Proceedings of the National Academy of Sciences of the United States of America 48*, 2013-2018.

78. Gerothanassis, I. P., Birdsall, B., Bauer, C. J., Frenkiel, T. A., and Feeney, J. (1992) *Journal of Molecular Biology 226*, 549-554.

79. Giraldo, R. and Rhodes, D. (1994) *EMBO Journal 13*, 2411-2420.

80. Goorin, A. M., Chauvenet, A. R., Perezatayde, A. R., Cruz, J., McKone, R., and Lipshultz, S. E. (1990) *Journal Of Pediatrics 116*, 144-147.

81. Gowan, S. M., Harrison, J. R., Patterson, L., Valenti, M., Read, M. A., Neidle, S., and Kelland, L. R. (2002) *Molecular Pharmacology 61*, 1154-1162.

82. Gowan, S. M., Heald, R., Stevens, M. F. G., and Kelland, L. R. (2001) *Molecular Pharmacology 60*, 981-988.

83. Gu, J. D. and Leszczynski, J. (2000) *Journal of Physical Chemistry A 104*, 1898-1904.

84. Gu, J. D. and Leszczynski, J. (2001) *Chemical Physics Letters 335*, 465-474.

85. Guano, F., Pourquier, P., Tinelli, S., Binaschi, M., Bigioni, M., Animati, F., Manzini, S., Zunino, F., Kohlhagen, G., Pommier, Y., and Capranico, G. (1999) *Molecular Pharmacology 56*, 77-84.

86. Guest, C. R., Hochstrasser, R. A., Dupuy, C. G., Allen, D. J., Benkovic, S. J., and Millar, D. P. (1991) *Biochemistry 30*, 8759-8770.

87. Guntert, P., Mumenthaler, C., and Wuthrich, K. (1997) *Journal of Molecular Biology 273*, 283-298.

88. Gupta, G., Garcia, A. E., Guo, Q., Lu, M., and Kallenbach, N. R. (1993) *Biochemistry 32*, 7098-7103.

89. Han, F. X. G., Wheelhouse, R. T., and Hurley, L. H. (1999) *Journal of the American Chemical Society 121*, 3561-3570.

90. Hare, D., Shapiro, L., and Patel, D. J. (1986) *Biochemistry 25*, 7456-7464.

91. Harley, C. B., Futcher, A. B., and Greider, C. W. (1990) *Nature 345*, 458-460.

92. Harris, S. A., Gavathiotis, E., Searle, M. S., Orozco, M., and Laughton, C. A. (2001) *Journal of the American Chemical Society 123*, 12658-12663.

93. Heald, R. A., Modi, C., Cookson, J. C., Hutchinson, I., Laughton, C. A., Gowan, S. M., Kelland, L. R., and Stevens, M. F. G. (2002) *Journal of Medicinal Chemistry 45*, 590-597.

94. Heus, H. A. and Pardi, A. (1991) *Science 253*, 191-194.

95. Heus, H. A., Uhlenbeck, O. C., and Pardi, A. (1990) *Nucleic Acids Research 18*, 1103-1108.

96. Hilbers, C. W., Blommers, M. J. J., Vandeven, F. J. M., Vanboom, J. H., and Vandermarel, G. A. (1991) *Nucleosides & Nucleotides 10*, 61-80.

97. Hirao, I., Ishida, M., Watanabe, K., and Miura, K. (1990) *Biochimica et Biophysica Acta 1087*, 199-204.

98. Hirao, I., Kawai, G., Yoshizawa, S., Nishimura, Y., Ishido, Y., Watanabe, K., and Miura, K. (1994) *Nucleic Acids Research 22*, 576-582.

99. Hirao, I., Naraoka, T., Kanamori, S., Nakamura, M., and Miura, K. (1988) *Biochemistry International 16*, 157-162.

100. Hirao, I., Nishimura, Y., Naraoka, T., Watanabe, K., Arata, Y., and Miura, K. (1989) *Nucleic Acids Research 17*, 2223-2231.

101. Hirao, I., Nishimura, Y., Tagawa, Y., Watanabe, K., and Miura, K. (1992) *Nucleic Acids Research 20*, 3891-3896.

102. Hunter, C. A. (1993) *Journal of Molecular Biology 230*, 1025-1054.

103. Hunter, C. A. and Lu, X. J. (1997) *Journal of Molecular Biology 265*, 603-619.

104. Johannesson, H. and Halle, B. (1998) *Journal of the American Chemical Society 120*, 6859-6870.

105. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) *Journal of Chemical Physics 79*, 926-935.

106. Joshuator, L., Rabinovich, D., Hope, H., Frolow, F., Appella, E., and Sussman, J. L. (1988) *Nature 334*, 82-84.

107. Kalnik, M. W., Norman, D. G., Li, B. F., Swann, P. F., and Patel, D. J. (1990) *The Journal of Biological Chemistry 265*, 636-647.

108. Kalnik, M. W., Norman, D. G., Swann, P. F., and Patel, D. J. (1989) *The Journal of Biological Chemistry 264*, 3702-3712.

109. Kalnik, M. W., Norman, D. G., Zagorski, M. G., Swann, P. F., and Patel, D. J. (1989) *Biochemistry 28*, 294-303.

110. Kamitori, S. and Takusagawa, F. (1994) *Journal of the American Chemical Society 116*, 4154-4165.

111. Kang, C., Zhang, X. H., Ratliff, R., Moyzis, R., and Rich, A. (1992) *Nature 356*, 126-131.

112. Ke, S.-H. and Wartell, R. M. (1995) *Biochemistry 34*, 4593-4600.

113. Kellner, U., Sehested, M., Jensen, P. B., Gieseler, F., and Rudolph, P. (2002) *Lancet Oncology 3*, 235-243.

114. Keniry, M. A. (2000) *Biopolymers 56*, 123-146.

115. Keniry, M. A., Strahan, G. D., Owen, E. A., and Shafer, R. H. (1995) *European Journal of Biochemistry 233*, 631-643.

116. Kettani, A., Bouaziz, S., Gorin, A., Zhao, H., Jones, R. A., and Patel, D. J. (1998) *Journal of Molecular Biology 282*, 619-636.

117. Kettani, A., Kumar, R. A., and Patel, D. J. (1995) *Journal of Molecular Biology 254*, 638-656.

118. Khan, I. M. and Coulson, J. M. (1993) *Nucleic Acids Research 21*, 2957-2958.

119. Khan, I. M. and Coulson, J. M. (1993) *Nucleic Acids Research 21*, 4433.

120. Kim, N. W., Piatyszek, M. A., Prowse, K. R., Harley, C. B., West, M. D., Ho, P. L. C., Coviello, G. M., Wright, W. E., Weinrich, S. L., and Shay, J. W. (1994) *Science 266*, 2011-2015.

121. Kochoyan, M. and Leroy, J. L. (1995) *Current Opinion in Structural Biology 5*, 329-333.

122. Koo, H. S., Rice, J., and Crothers, D. M. (1987) *Biophysical Journal 51*, A506.

123. Koradi, R., Billeter, M., and Wuthrich, K. (1996) *Journal of Molecular Graphics 14*, 51-55.

124. Kraulis, P. J. (1989) *Journal of Magnetic Resonance 84*, 627-633.

125. Lam, W.-C., Van der Schans, E. J. C., Sowers, L. C., and Millar, D. P. (1999) *Biochemistry 38*, 2661-2668.

126. Lamers, M. H., Perrakis, A., Enzlin, J. H., Winterwerp, H. H. K., de Wind, N., and Sixma, T. K. (2000) *Nature 407*, 711-717.

127. Lamiri, A., Albiser, G., and Premilat, S. (1998) *Biochemical and Biophysical Research Communications 253*, 809-812.

128. Lane, A. N. and Jenkins, T. C. (2001) *Current Organic Chemistry 5*, 845-869.

129. Laughlan, G., Murchie, A. I. H., Norman, D. G., Moore, M. H., Moody, P. C. E., Lilley, D. M. J., and Luisi, B. (1994) *Science 265*, 520-524.

130. Laver, W. G., Bischofberger, N., and Webster, R. G. (1999) *Scientific American 280*, 78-87.

131. Lavery, R. and Sklenar, H. (1989) *Journal of Biomolecular Structure & Dynamics 6*, 655-667.

132. Lawrence, M. C., Borg, N. A., Streltsov, V. A., Pilling, P. A., Epa, V. C., Varghese, J. N., McKimm-Breschkin, J. L., and Colman, P. M. (2004) *Journal of Molecular Biology 335*, 1343-1357.

133. Liaw, Y. C., Gao, Y. G., Robinson, H., Vandermarel, G. A., Vanboom, J. H., and Wang, A. H. J. (1989) *Biochemistry 28*, 9913-9918.

134. Licata, S., Saponiero, A., Mordente, A., and Minotti, G. (2000) *Chemical Research In Toxicology 13*, 414-420.

135. Liepinsh, E., Leupin, W., and Otting, G. (1994) *Nucleic Acids Research 22*, 2249-2254.

136. Lu, M., Guo, Q., and Kallenbach, N. R. (1992) *Biochemistry 31*, 2455-2459.

137. Maillet, J. B., Lachet, V., and Coveney, P. V. (1999) *Physical Chemistry Chemical Physics 1*, 5277-5290.

138. Makarov, V. L., Lejnine, S., Bedoyan, J., and Langmore, J. P. (1993) *Cell 73*, 775-787.

139. Marathias, V. M. and Bolton, P. H. (1999) *Biochemistry 38*, 4355-4364.

140. Marcu, K. B., Bossone, S. A., and Patel, A. J. (1992) *Annual Review of Biochemistry 61*, 809-860.

141. Medicines Control Agency, 1. N. E. L. L. S. 5. Medicines Control Agency. Market Towers, 1 Nine Elms Lane, London.SW8 5NQ. 2002. Ref Type: Internet Communication

142. Martin, E. J., Blaney, J. M., Siani, M. A., Spellmeyer, D. C., Wong, A. K., and Moos, W. H. (1995) *Journal of Medicinal Chemistry 38*, 1431-1436.

143. Matsugami, A., Okuizumi, T., Uesugi, S., and Katahira, M. (2003) *Journal of Biological Chemistry 278*, 28147-28153.

144. Mayer, C. and Timsit, Y. (2001) *Cellular and Molecular Biology 47*, 815-822.

145. Mayer-Jung, C., Moras, D., and Timsit, Y. (1998) *Embo Journal 17*, 2709-2718.

146. Mccall, M., Brown, T., and Kennard, O. (1985) *Journal of Molecular Biology 183*, 385-396.

147. Mergny, J. L., Phan, A. T., and Lacroix, L. (1998) *FEBS Letters 435*, 74-78.

148. Michel, F. and Westhof, E. (1990) *Journal of Molecular Biology 216*, 585-610.

149. Miller, M., Harrison, R. W., Wlodawer, A., Appella, E., and Sussman, J. L. (1988) *Nature 334*, 85-86.

150. Moody, E. M. and Bevilacqua, P. C. (2003) *Journal of the American Chemical Society 125*, 16285-16293.

151. Moody, E. M. and Bevilacqua, P. C. (2003) *Journal of the American Chemical Society 125*, 2032-2033.

152. Morden, K. M., Chu, Y. G., Martin, F. H., and Tinoco, I. Jr. (1983) *Biochemistry 22*, 5557-5563.

153. Muyldermans, S., Dejonge, J., Wyns, L., and Travers, A. A. (1994) *Nucleic Acids Research 22*, 5635-5639.

154. Nakatani, K., Akimitsu, O., and Saito, I. (1999) *Angewandte Chemie : International Edition 38*, 3378-3381.

155. Nakatani, K., Sando, S., and Saito, I. (2000) *Journal of the American Chemical Society 122*, 2172-2177.

156. Nelson, J. W. and Tinoco, I. (1985) *Biochemistry 24*, 6416-6421.

157. Obmolova, G., Ban, C., Hsieh, P., and Yang, W. (2000) *Nature 407*, 703-710.

158. Okada, Y., Streisinger, G., Owen, J. E., Newton, G., Tsugita, A., and Inouye, M. (1972) *Nature 236*, 338-341.

159. Ollis, D. L., Brick, P., Hamlin, R., Xuong, N. G., and Steitz, T. A. (1985) *Nature 313*, 762-766.

160. Olson, W. K., Bansal, M., Burley, S. K., Dickerson, R. E., Gerstein, M., Harvey, S. C., Heinemann, U., Lu, X. J., Neidle, S., Shakked, Z., Sklenar, H., Suzuki, M., Tung, C. S., Westhof, E., Wolberger, C., and Berman, H. M. (2001) *Journal of Molecular Biology 313*, 229-237.

161. Ortho Biotech Products A.P. DOXIL Prescribing Information. Ortho Biotech Products A.P., Raritan.NJ 08869-0670. 2002.
Ref Type: Internet Communication

162. Parkinson, G. N., Lee, M. P. H., and Neidle, S. (2002) *Nature 417*, 876-880.

163. Patel, D. J., Kozlowski, S. A., Marky, L. A., Rice, J. A., Broka, C., Itakura, K., and Breslauer, K. J. (1982) *Biochemistry 21*, 445-451.

164. Patel, P. K., Bhavesh, N. S., and Hosur, R. V. (2000) *Biochemical and Biophysical Research Communications 278*, 833-838.

165. Patel, P. K., Bhavesh, N. S., and Hosur, R. V. (2000) *Biochemical and Biophysical Research Communications 270*, 967-971.

166. Patel, P. K. and Hosur, R. V. (1999) *Nucleic Acids Research 27*, 2457-2464.

167. Patel, P. K., Koti, A. S. R., and Hosur, R. V. (1999) *Nucleic Acids Research 27*, 3836-3843.

168. Pelengaris, S., Rudolph, B., and Littlewood, T. (2000) *Current Opinion in Genetics & Development 10*, 100-105.

169. Pharmaceutical Research and Manufacturers of America. Why do prescription drugs cost so much? 2000. PhRMA.
Ref Type: Pamphlet

170. Phillips, K., Dauter, Z., Murchie, A. I. H., Lilley, D. M. J., and Luisi, B. (1997) *Journal of Molecular Biology 273*, 171-182.

171. Piantini, U., Sorensen, O. W., and Ernst, R. R. (1982) *Journal of the American Chemical Society 104*, 6800-6801.

172. Pilch, D. S., Xu, Z. T., Sun, Q., Lavoie, E. J., Liu, L. F., and Breslauer, K. J. (1997) *Proceedings of the National Academy of Sciences of the United States of America 94*, 13565-13570.

173. Piotto, M., Saudek, V., and Sklenar, V. (1992) *Journal of Biomolecular Nmr 2*, 661-665.

174. Poddevin, B., Meguenni, S., Elias, I., Vasseur, M., and Blumenfeld, M. (1994) *Antisense Research and Development 4*, 147-154.

175. Postel, E. H., Berberich, S. J., Rooney, J. W., and Kaetzel, D. M. (2000) *Journal of Bioenergetics and Biomembranes 32*, 277-284.

176. Prescott, D. M. (1994) *Microbiological Reviews 58*, 233-267.

177. Rance, M., Sorensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., and Wuthrich, K. (1983) *Biochemical and Biophysical Research Communications 117*, 479-485.

178. Read, M., Harrison, R. J., Romagnoli, B., Tanious, F. A., Gowan, S. H., Reszka, A. P., Wilson, W. D., Kelland, L. R., and Neidle, S. (2001) *Proceedings of the National Academy of Sciences of the United States of America 98*, 4844-4849.

179. Remin, M. and Shugar, D. (1972) *Biochem. Biophys. Res. Commun. 48*, 636-642.

180. Rhodes, D. and Giraldo, R. (1995) *Current Opinion in Structural Biology 5*, 311-322.

181. Robinson, H., Liaw, Y. C., Vandermarel, G. A., Vanboom, J. H., and Wang, A. H. J. (1990) *Nucleic Acids Research 18*, 4851-4858.

182. Robinson, H., Liaw, Y. C., Vandermarel, G. A., Vanboom, J. H., and Wang, A. H. J. (1990) *Nucleic Acids Research 18*, 4851-4858.

183. Robinson, H., Yang, D. Z., and Wang, A. H. J. (1994) *Gene 149*, 179-188.

184. Sahasrabudhe, A., Lawrence, L., Epa, V. C., Varghese, J. N., Colman, P. M., and McKimm-Breschkin, J. L. (1998) *Virology 247*, 14-21.

185. Sakata, T., Hiroaki, H., Oda, Y., Tanaka, T., Ikehara, M., and Uesugi, S. (1990) *Nucleic Acids Research 18*, 3831-3839.

186. Sakatsume, O., Tsutsui, H., Wang, Y. F., Gao, H., Tang, X. R., Yamauchi, T., Murata, T., Itakura, K., and Yokoyama, K. K. (1996) *Journal of Biological Chemistry 271*, 31322-31333.

187. Schaffitzel, C., Berger, I., Postberg, J., Hanes, J., Lipps, H. J., and Pluckthun, A. (2001) *Proceedings of the National Academy of Sciences of the United States of America 98*, 8572-8577.

188. Schneider, B., Cohen, D. M., Schleifer, L., Srinivasan, A. R., Olson, W. K., and Berman, H. M. (1993) *Biophysical Journal 65*, 2291-2303.

189. Schofield, M. J., Brownewell, F. E., Nayak, S., Du, C. W., Kool, E. T., and Hsieh, P. (2001) *Journal of Biological Chemistry 276*, 45505-45508.

190. Schofield, M. J., Nayak, S., Scott, T. H., Du, C. W., and Hsieh, P. (2001) *Journal of Biological Chemistry 276*, 28291-28299.

191. Schroeder, M. C., Kraker, A. J., Moore, C. W., Kiely, J. S., Dewitt, S. H., and Czarnik, A. W. (1994) *Abstracts of Papers of the American Chemical Society 208*, 239-MEDI.

192. Schultze, P., Macaya, R. F., and Feigon, J. (1994) *Journal of Molecular Biology 235*, 1532-1547.

193. Searle, M. S. (1993) *Progress in Nuclear Magnetic Resonance Spectroscopy 25*, 403-480.

194. Searle, M. S., Hall, J. G., Denny, W. A., and Wakelin, L. P. G. (1989) *Biochemical Journal 259*, 433-441.

195. Sebestyen, F., Dibo, G., Kovacs, A., and Furka, A. (1993) *Bioorganic & Medicinal Chemistry Letters 3*, 413-418.

196. Sen, D. and Gilbert, W. (1988) *Nature 334*, 364-366.

197. Shan, K., Lincoff, A. M., and Young, J. B. (1996) *Annals Of Internal Medicine 125*, 47-58.

198. Shimizu, A. and Honjo, T. (1984) *Cell 36*, 801-803.

199. Shore, D. (1994) *Trends in Genetics 10*, 408-412.

200. Siddiqui-Jain, A., Grand, C. L., Bearss, D. J., and Hurley, L. H. (2002) *Proceedings of the National Academy of Sciences of the United States of America 99*, 11593-11598.

201. Siddiqui-Jain, A., Grand, C. L., Bearss, D. J., and Hurley, L. H. (2002) *Proceedings of the National Academy of Sciences of the United States of America 99*, 11593-11598.

202. Sim, S. P., Gatto, B., Yu, C., Liu, A. A., Li, T. K., Pilch, D. S., Lavoie, E. J., and Liu, L. F. (1997) *Biochemistry 36*, 13285-13291.

203. Simonsson, T. (2001) *Biological Chemistry 382*, 621-628.

204. Simonsson, T., Pecinka, P., and Kubista, M. (1998) *Nucleic Acids Research 26*, 1167-1172.

205. Simonsson, T., Pecinka, P., and Kubista, M. (1998) *Nucleic Acids Research 26*, 1167-1172.

206. Simonsson, T., Pribylova, M., and Vorlickova, M. (2000) *Biochemical and Biophysical Research Communications 279*, 306.

207. Sitkoff, D. and Case, D. A. (1998) *Progress in Nuclear Magnetic Resonance Spectroscopy 32*, 165-190.

208. Skulnick, H. I., Johnson, P. D., Aristoff, P. A., Morris, J. K., Lovasz, K. D., Howe, W. J., Watenpaugh, K. D., Janakiraman, M. N., Anderson, D. J., Reischer, R. J., Schwartz, T. M., Banitt, L. S., Tomich, P. K., Lynn, J. C., Horng, M. M., Chong, K. T., Hinshaw, R. R., Dolak, L. A., Seest, E. P., Schwende, F. J., Rush, B. D., Howard, G. M., Toth, L. N., Wilkinson, K. R., Kakuk, T. J., Johnson, C. W., Cole, S. L., Zaya, R. M., Zipp, G. L., Possert, P. L., Dalga, R. J., Zhong, W. Z., Williams, M. G., and Romines, K. R. (1997) *Journal of Medicinal Chemistry 40*, 1149-1164.

209. Skulnick, H. I., Johnson, P. D., Howe, W. J., Tomich, P. K., Chong, K. T., Watenpaugh, K. D., Janakiraman, M. N., Dolak, L. A., McGrath, J. P., Lynn, J. C., Horng, M. M., Hinshaw, R. R., Zipp, G. L., Ruwart, M. J., Schwende, F. J., Zhong, W. Z., Padbury, G. E., Dalga, R. J., Shiou, L. H., Possert, P. L., Rush, B. D., Wilkinson, K. F., Howard, G. M., Toth, L. N., Williams, M. G., Kakuk, T. J., Cole, S. L., Zaya, R. M., Lovasz, K. D., Morris, J. K., Romines, K. R., Thaisrivongs, S., and Aristoff, P. A. (1995) *Journal of Medicinal Chemistry 38*, 4968-4971.

210. Smith, C. K., Brannigan, J. A., and Moore, M. H. (1996) *Journal of Molecular Biology 263*, 237-258.

211. Smith, C. K., Davies, G. J., Dodson, E. J., and Moore, M. H. (1995) *Biochemistry 34*, 415-425.

212. Smith, C. K., Davies, G. J., Dodson, E. J., and Moore, M. H. (1995) *Biochemistry 34*, 415-425.

213. Smith, F. W. and Feigon, J. (1993) *Biochemistry 32*, 8682-8692.

214. Smith, F. W., Lau, F. W., and Feigon, J. (1994) *Proceedings of the National Academy of Sciences of the United States of America 91*, 10546-10550.

215. Spencer, C. A. and Groudine, M. (1991) *Advances in Cancer Research 56*, 1-48.

216. Stein, E. G., Rice, L. M., and Brunger, A. T. (1997) *Journal of Magnetic Resonance 124*, 154-164.

217. Strahan, G. D., Shafer, R. H., and Keniry, M. A. (1994) *Nucleic Acids Research 22*, 5447-5455.

218. Streisinger, G. and Owen, J. E. (1985) *Genetics 109*, 633-659.

219. Sun, D. Y., Thompson, B., Cathers, B. E., Salazar, M., Kerwin, S. M., Trent, J. O., Jenkins, T. C., Neidle, S., and Hurley, L. H. (1997) *Journal of Medicinal Chemistry 40*, 2113-2116.

220. Sundquist, W. I. and Klug, A. (1989) *Nature 342*, 825-829.

221. Szadkowski, M. and Jiricny, J. (2002) *Genes Chromosomes & Cancer 33*, 36-46.

222. Thaisrivongs, S., Skulnick, H. I., Turner, S. R., Strohbach, J. W., Tommasi, R. A., Johnson, P. D., Aristoff, P. A., Judge, T. M., Gammill, R. B., Morris, J. K., Romines, K. R., Chrusciel, R. A., Hinshaw, R. R., Chong, K. T., Tarpley, W. G., Poppe, S. M., Slade, D. E., Lynn, J. C., Horng, M. M., Tomich, P. K., Seest, E. P., Dolak, L. A., Howe, W. J., Howard, G. M., Schwende, F. J., Toth, L. N., Padbury, G. E., Wilson, G. J., Shiou, L. H., Zipp, G. L., Wilkinson, K. F., Rush, B. D., Ruwart, M. J., Koeplinger, K. A., Zhao, Z. Y., Cole, S., Zaya, R. M., Kakuk, T. J., Janakiraman, M. N., and Watenpaugh, K. D. (1996) *Journal of Medicinal Chemistry 39*, 4349-4353.

223. Tommerup, H., Dousmanis, A., and Delange, T. (1994) *Molecular and Cellular Biology 14*, 5777-5785.

224. Tomonaga, T. and Levens, D. (1996) *Proceedings of the National Academy of Sciences of the United States of America 93*, 5830-5835.

225. Tornier, C., Bessone, S., Varlet, I., Rudolph, C., Darmon, M., and Fleck, O. (2001) *Genetics 158*, 65-75.

226. Tuerk, C., Gauss, P., Thermes, C., Groebe, D. R., Gayle, M., Guild, N., Stormo, G., Daubentoncarafa, Y., Uhlenbeck, O. C., Tinoco, I., Brody, E. N., and Gold, L. (1988) *Proceedings of the National Academy of Sciences of the United States of America 85*, 1364-1368.

227. U.S.National Library of Medicine, 8. R. P. B. M. 2. MEDLINEplus. U.S.National Library of Medicine, 8600 Rockville Pike, Bethesda.MD 20894. 2002.

Ref Type: Internet Communication

228. van Drie, J. A., Roher, D. C., Blinn, J. R., and Gao, H. (2003) in *Modern Methods of Drug Discovery* (Hillisch, A. and Hilgenfeld, R., Eds.) pp 203-222, Birkhauser Verlag, Basel.

229. Varani, G. (1995) *Annual Review of Biophysics and Biomolecular Structure 24*, 379-404.

230. Varani, L., Gunderson, S. I., Mattaj, I. W., Kay, L. E., Neuhaus, D., and Varani, G. (2000) *Nature Structural Biology 7*, 329-335.

231. Varghese, J. N., Epa, V. C., and Colman, P. M. (1995) *Protein Science 4*, 1081-1087.

232. Venczel, E. A. and Sen, D. (1993) *Biochemistry 32*, 6220-6228.

233. Wang, Y., Delossantos, C., Gao, X. L., Greene, K., Live, D., and Patel, D. J. (1991) *Journal of Molecular Biology 222*, 819-832.

234. Wang, Y. and Patel, D. J. (1993) *Journal of Molecular Biology 234*, 1171-1183.

235. Ward, A. C. (1995) *Virus Genes 10*, 253-260.

236. Weber, I. T., Miller, M., Jaskolski, M., Leis, J., Skalka, A. M., and Wlodawer, A. (1989) *Science 243*, 928-931.

237. Wellinger, R. J., Wolf, A. J., and Zakian, V. A. (1993) *Cell 72*, 51-60.

238. Wheelhouse, R. T., Sun, D. K., Han, H. Y., Han, F. X. G., and Hurley, L. H. (1998) *Journal of the American Chemical Society 120*, 3261-3262.

239. Wijmenga, S. S., Kruithof, M., and Hilbers, C. W. (1997) *Journal of Biomolecular NMR 10*, 337-350.

240. Williams, H. E. L., Colgrave, M. L., and Searle, M. S. (2002) *European Journal of Biochemistry 269*, 1726-1733.

241. Williams, H. E. L. and Searle, M. S. (1999) *Journal of Molecular Biology 290*, 699-716.

242. Williams, H. E. L. and Searle, M. S. (1999) *Journal of Molecular Biology 290*, 699-716.

243. Williams, L. D., Egli, M., Qi, G., Bash, P., Vandermarel, G. A., Vanboom, J. H., Rich, A., and Frederick, C. A. (1990) *Proceedings of the National Academy of Sciences of the United States of America 87*, 2225-2229.

244. Williamson, J. R. (1993) *Current Opinion in Structural Biology 3*, 357-362.

245. Wlodawer, A., Miller, M., Jaskolski, M., Sathyanarayana, B. K., Baldwin, E., Weber, I. T., Selk, L. M., Clawson, L., Schneider, J., and Kent, S. B. H. (1989) *Science 245*, 616-621.

246. Woodson, S. A. and Crothers, D. M. (1988) *Biochemistry 27*, 8904-8914.

247. Woodson, S. A. and Crothers, D. M. (1988) *Biochemistry 27*, 3130-3141.

248. Wu, H. M. and Crothers, D. M. (1984) *Nature 308*, 509-513.

249. Wuthrich, K. (1986) *NMR of Proteins and Nucleic Acids* Wiley, New York.

250. Wuthrich, K. (1986) *NMR of Proteins and Nucleic Acids* Wiley, New York.

251. Yoshizawa, S., Kawai, G., Watanabe, K., Miura, K., and Hirao, I. (1997) *Biochemistry 36*, 4761-4767.

252. Young, P. R. and Kallenbach, N. R. (1981) *Journal of Molecular Biology 145*, 785-813.

253. Zahler, A. M., Williamson, J. R., Cech, T. R., and Prescott, D. M. (1991) *Nature 350*, 718-720.

254. Zakian, V. A. (1995) *Science 270*, 1601-1607.

255. Zhang, X. L., Mar, V., Zhou, W., Harrington, L., and Robinson, M. O. (1999) *Genes & Development 13*, 2388-2399.

256. Zhang, X. L. and Patel, D. J. (1990) *Biochemistry 29*, 9451-9466.

257. Zhu, J. and Wartell, R. M. (1999) *Biochemistry 38*, 15986-15993.

258. Zhu, L. M., Chou, S. H., Xu, J. D., and Reid, B. R. (1995) *Nature Structural Biology 2*, 1012-1017.

259. Zipp, G. L., Shiou, L., Possert, P. L., Dalga, R. J., Ruwart, M. J., Rush, B. D., Wilkinson, K. F., Schwende, F. J., Zhong, W. Z., Williams, M. G., Padbury, G. E., Johnson, P. D., Skulnick, H. I., Aristoff, P. A., and Thaisrivongs, S. (1996) *Abstracts of Papers of the American Chemical Society 211*, 11-BTEC.

Appendices

Appendix 1: ^{1}H Chemical Shift Data

Appendix 1.1: DNA ^{1}H chemical shifts of the complex of nogalamycin with d(TGATCA)$_2$ at 298 K.

	T1	G2	A3	T4	C5	A6
H1'	5.930	5.137	6.155	5.976	5.902	6.517
H2'1	2.636	2.512	2.459	1.897	2.317	2.488
H2'2	2.872	2.666	2.689	2.352	2.620	2.545
H3'	4.846	---	4.956	4.861	4.997	4.736
H4'	4.208	---	4.305	4.050	4.238	4.354
H5'1	3.884	4.136	4.133	---	---	4.202
H5'2	3.948	4.234	4.172	---	---	---
H5	---	---	---	---	5.688	---
H6	7.504	---	---	7.077	7.688	---
H8	---	7.948	8.036	---	---	8.467
Me	1.485	---	---	1.073	---	---

Appendix 1.2: DNA ^{1}H chemical shifts of nogalamycin-d(CTGATCAG)$_2$ at 298 K.

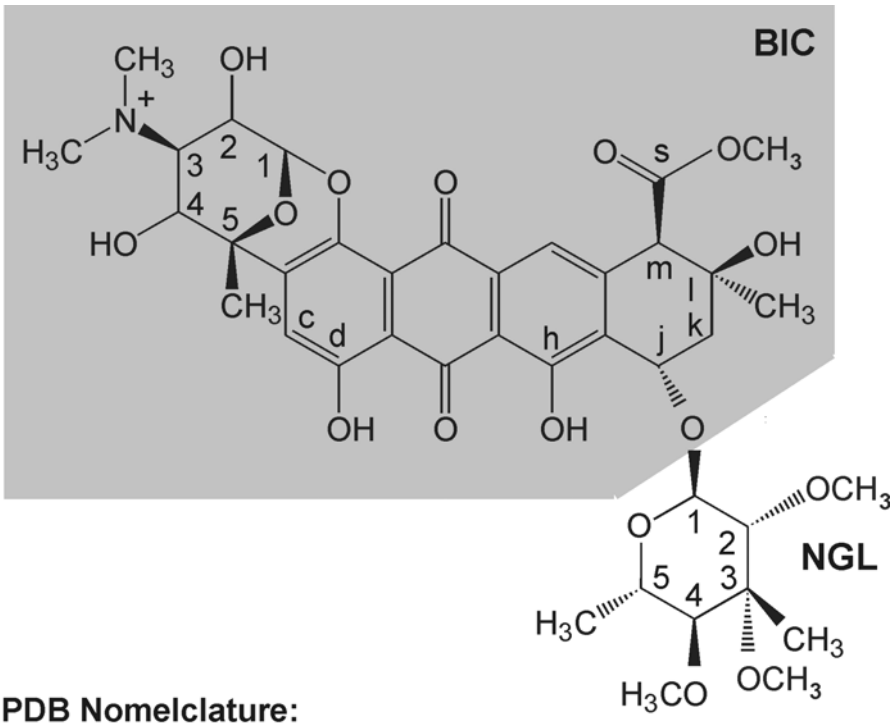
	C1	T2	G3	A4	T5	C6	A7	G8
H1'	5.755	6.190	5.202	6.117	5.958	5.881	6.323	6.084
H2'1	2.126	2.662	2.519	2.446	1.897	2.295	2.443	2.319
H2'2	2.482	2.760	2.594	2.703	2.359	2.605	2.816	2.475
H3'	4.562	5.086	4.936	4.928	4.862	5.013	5.062	4.665
H4'	4.016	4.226	4.509	4.279	4.047	4.227	4.526	4.216
H5'1	3.794	4.149	4.080	4.182	4.185	4.060	4.239	3.962
H5'2	3.847	---	4.359	---	4.301	4.104	4.280	4.142
H5	5.692	---	---	---	---	5.686	---	---
H6	7.776	7.387	---	---	7.072	7.690	---	---
H8	---	---	7.881	8.005	---	---	8.428	7.700
Me	---	1.436	---	---	1.094	---	---	---

Appendix 1.3: DNA ^{1}H chemical shifts of nogalamycin-d(CCTGATCAGG)$_2$ at 298 K.

	C1	C2	T3	G4	A5	T6
H1'	5.727	5.900	6.122	5.185	6.103	5.939
H2'1	2.194	2.088	2.603	2.499	2.445	1.882
H2'2	2.454	2.503	2.751	2.595	2.691	2.343
H3'	4.615	4.734	5.038	4.936	4.926	4.856
H4'	4.057	4.136	4.362	4.489	4.252	4.029
H5'1	3.764	4.087	4.177	4.032	4.168	4.179
H5'2	3.801	---	4.235	4.366	---	4.279
H5	5.840	5.471	---	---	---	---
H6	7.678	7.589	7.268	---	---	7.060
H8	---	---	---	7.872	7.992	---
Me	---	---	1.404	---	---	1.081

	C7	A8	G9	G10
H1'	5.843	6.268	6.103	5.651
H2'1	2.262	2.433	2.364	2.357
H2'2	2.593	2.765	2.480	2.592
H3'	4.999	5.084	4.652	4.651
H4'	4.220	4.530	3.948	4.221
H5'1	4.033	4.189	---	---
H5'2	4.181	4.228	---	---
H5	5.672	---	---	---
H6	7.678	---	---	---
H8	---	8.384	7.678	7.735
Me	---	---	---	---

Due to restrictions imposed by the PDB format, the following naming convention is used for nogalamycin throughout the appendices:



PDB Nomelclature:

Residue	Atom	Position	Up (1) or Down (2)
NGL	H	5	2

Appendix 1.4: Nogalamycin ^{1}H chemical shifts of the complex of nogalamycin with $d(TGATCA)_2$ at 298 K.

	BIC7	NGL8
H11	5.669	---
H12	---	5.310
H21	4.215	3.036
H22	---	3.669
H31	---	1.464
H32	4.681	3.443
H41	3.957	3.581
H42	---	3.183
H51	1.631	3.822
H52	---	1.411
HC	6.927	---
HJ1	4.718	---
HK1	2.914	---
HK2	1.917	---
HL2	1.459	---
HM2	4.495	---
HO	6.649	---
HS2	3.787	---
M31	2.891	---
M32	2.973	---

Appendix 1.5: Nogalamycin ^{1}H chemical shifts of the complex of nogalamycin with $d(CTGATCAG)_2$ at 298 K.

	BIC9	NGL10
H11	5.777	---
H12	---	5.144
H21	---	3.051
H22	---	3.631
H31	---	1.398
H32	---	3.579
H41	---	3.180
H42	---	3.417
H51	1.660	3.819
H52	---	1.459
HC	6.939	---
HJ1	4.634	---
HK1	2.902	---
HK2	1.911	---
HL2	1.501	---
HM2	4.500	---
HO	6.669	---
HS2	3.794	---
M31	2.928	---
M32	2.980	---

Appendix 1.6: Nogalamycin ^{1}H chemical shifts of the complex of nogalamycin with d(CCTGATCAGG)$_2$ at 298 K.

	BIC11	NGL12
H11	5.747	---
H12	---	5.099
H21	4.210	3.027
H22	---	3.605
H31	---	1.386
H32	4.653	3.399
H41	---	3.161
H42	---	3.566
H51	1.643	3.787
H52	---	1.440
HC	6.920	---
HJ1	4.574	---
HK1	2.911	---
HK2	2.947	---
HL2	1.600	---
HM2	4.225	---
HO	6.630	---
HS2	3.948	---
M31	2.910	---
M32	2.990	---

Appendix 1.7: DNA ^{1}H chemical shifts of the complex of nogalamycin with d(GCTACGAAGTGC) at 288 K.

	G1	C2	T3	A4	C5	G6
H1'	5.970	6.388	6.665	6.347	5.991	---
H2'1	2.680	2.352	2.540	2.434	1.704	2.527
H2'2	2.794	2.449	2.645	2.730	2.264	2.666
H3'	4.877	5.039	4.987	5.080	4.839	---
H4'	4.263	4.284	4.449	4.470	4.305	4.425
H5'1	3.755	4.216	---	4.269	4.116	---
H5'2	---	---	---	4.357	---	---
H5	---	5.450	---	---	4.875	---
H6	---	7.474	7.828	---	7.032	---
H8	7.964	---	---	8.545	---	8.079
Me	---	---	2.022	---	---	---

	A7	A8	G9	T10	G11	C12
H1'	5.978	6.274	5.515	6.132	5.577	5.931
H2'1	3.029	2.874	2.591	2.655	2.564	2.013
H2'2	3.375	---	---	2.694	---	2.180
H3'	4.572	4.826	4.933	4.980	5.000	4.423
H4'	2.043	---	4.392	---	4.621	4.252
H5'1	2.268	---	4.185	4.305	4.342	3.900
H5'2	---	---	---	4.349	---	4.012
H5	---	---	---	---	---	5.225
H6	---	---	---	7.188	---	7.432
H8	8.100	7.992	7.971	---	7.999	---
Me	---	---	---	1.110	---	---

Appendix 1.8: DNA ^{1}H chemical shifts of the complex of nogalamycin with d(GCCCGTAGTGC) at 298 K.

	G1	C2	C3	C4	G5	T6
H1'	5.693	6.067	6.025	5.934	5.737	5.599
H2'1	2.413	2.039	1.614	1.731	2.393	1.595
H2'2	2.496	2.394	2.339	2.135	2.558	1.908
H3'	4.68	4.89	4.638	4.686	4.746	4.264
H4'	4.049	4.17	4.216	4.147	4.34	---
H5'1	3.624	3.957	4.028	3.798	3.936	3.107
H5'2	---	4.03	4.108	3.849	4.023	3.242
H5	---	5.308	5.684	5.487	---	---
H6	---	7.51	7.726	7.209	---	6.965
H8	7.796	---	---	---	8.038	---
Me	---	---	---	---	---	1.225

	A7	G8	T9	G10	C11
H1'	6.072	5.338	5.694	5.593	5.852
H2'1	2.57	2.272	2.372	2.371	1.779
H2'2	2.734	2.411	2.428	2.512	1.993
H3'	4.651	---	4.872	4.735	4.221
H4'	4.174	4.214	4.084	4.503	3.783
H5'1	3.59	3.97	3.962	3.977	3.633
H5'2	3.77	4.085	3.981	4.195	---
H5	---	---	---	---	4.969
H6	---	---	6.802	---	7.127
H8	7.821	7.747	---	7.404	---
Me	---	---	0.643	---	---

Appendix 1.9: Nogalamycin ^{1}H chemical shifts of the complex of nogalamycin with d(GCTACGAAGTGC) at 288 K.

	BIC13	NGL14
H11	5.754	---
H12	---	5.117
H21	4.318	3.161
H22	---	3.786
H31	---	1.289
H32	4.742	3.201
H41	3.988	3.512
H42	---	3.518
H51	1.654	---
H52	---	1.392
HC	7.024	---
HJ1	4.437	---
HK1	2.000	---
HK2	2.949	---
HL2	1.522	---
HM2	4.342	---
HO	6.881	---
HS2	3.876	---
M32	3.118	---
M31	2.979	---

Appendix 1.10: Nogalamycin ^{1}H chemical shifts of the complex of nogalamycin with d(GCCCGTAGTGC) at 298 K.

	BIC12	NGL13
H11	5.580	---
H12	---	---
H21	4.141	2.641
H22	---	3.745
H31	---	1.258
H32	4.451	3.393
H41	3.832	3.640
H42	---	3.779
H51	1.427	1.824
H52	---	---
HC	6.750	---
HD	11.639	---
HH	11.457	---
HJ1	4.821	---
HK1	1.990	---
HK2	3.079	---
HL2	1.505	---
HM2	4.503	---
HO	6.797	---
M31	2.845	---
M32	2.894	---

Appendix 1.11: DNA ^{1}H chemical shifts of d(TAGGAGGT)$_4$ at 298 K.

	T1	A2	G3	G4	A5	G6	G7	T8
H1'	5.846	6.155	5.969	6.014	6.048	6.056	6.160	5.999
H2'1	1.676	2.825	2.520	2.246	2.672	2.665	2.476	2.122
H2'2	2.198	---	2.928	2.958	2.835	2.753	2.622	---
H3'	4.544	4.998	4.968	5.021	5.163	5.096	4.938	---
H4'	3.487	---	4.499	4.413	4.443	4.585	4.478	---
H5'1	3.940	---	4.169	4.185	4.291	4.244	4.199	---
H5'2	4.050	---	4.237	4.299	4.412	4.309	4.244	---
H1	---	---	11.537	---	---	---	10.729	---
H2	---	---	---	11.300	---	---	---	---
H6	7.122	---	---	---	---	---	---	7.283
H8	---	8.225	7.912	7.146	8.257	8.054	7.648	---
Me	1.528	---	---	---	---	---	---	1.598

Appendix 1.12: DNA ^{1}H chemical shifts of d(TGGAGGC)$_4$ at 298 K.

	T1	G2	G3	A4	G5	G6	C7
H1'	5.918	6.063	6.034	6.082	6.119	6.167	5.969
H2'1	2.220	2.612	2.365	2.657	2.528	2.529	2.081
H2'2	2.540	3.024	2.956	2.780	2.802	2.686	2.137
H3'	4.720	5.031	5.047	5.163	5.124	4.946	4.969
H4'	4.104	4.453	4.468	4.371	4.651	4.523	4.465
H5'1	3.796	4.123	4.221	4.298	4.243	4.062	4.018
H5'2	---	---	4.328	---	4.293	4.234	4.227
H1	---	11.568	11.238	---	10.496	10.777	---
H2	---	---	---	9.362	---	---	---
H21	---	5.803	6.067	---	5.589	---	---
H22	---	9.925	9.358	---	9.363	---	---
H5	---	---	---	---	---	---	5.561
H6	7.522	---	---	---	---	---	7.452
H8	---	8.104	7.315	8.252	8.076	7.670	---
Me	1.369	---	---	---	---	---	---

Appendix 1.13: DNA ^{1}H chemical shifts of d(TGGGAGGC)$_4$ at 298 K.

	T1	G2	G3	G4	A5	G6	G7	C8
H1'	5.925	6.078	6.056	5.988	6.020	6.074	6.141	6.024
H2'1	2.148	2.713	2.414	2.294	2.637	2.712	2.491	2.082
H2'2	2.427	3.007	2.886	2.964	2.790	2.786	2.626	2.164
H3'	4.702	5.000	4.993	5.053	5.156	5.090	4.934	4.449
H4'	4.067	4.401	4.543	4.496	4.495	4.612	4.489	4.177
H5'1	3.695	4.060	4.205	4.286	4.282	4.228	4.180	4.031
H5'2	3.717	4.120	4.307	4.369	4.372	4.292	4.257	4.056
H5	---	---	---	---	---	---	---	5.552
H6	7.420	---	---	---	---	---	---	7.465
H8	---	8.004	7.626	7.246	8.238	8.028	7.649	---
Me	1.407	---	---	---	---	---	---	---

Appendix 2: Distance Restraints

Appendix 2.1: Nogalamycin$_2$-d(TGATCA)$_2$

Identifier A	Residue A	Atom A	Identifier B	Residue B	Atom B	Distance
1	DT	H1'	13	BIC	HO	3.37
1	DT	H1'	1	DT	H6	3.67
1	DT	H2'1	1	DT	H1'	2.75
1	DT	H2'2	13	BIC	HO	2.73
1	DT	H2'2	1	DT	H1'	2.85
1	DT	H2'2	1	DT	H3'	3.02
1	DT	H2'2	1	DT	H5'1	3.42
1	DT	H3'	1	DT	H6	2.63
1	DT	H6	13	BIC	HO	3.80
1	DT	H6	1	DT	H2'1	3.20
1	DT	H6	1	DT	H2'2	2.46
1	DT	H7	13	BIC	H11	2.46
1	DT	H7	1	DT	H6	3.10
1	DT	H7	2	DG	H4'	2.47
2	DG	H1'	13	BIC	HO	3.81
2	DG	H1'	2	DG	H2'2	3.53
2	DG	H1'	2	DG	H4'	3.93
2	DG	H1'	2	DG	H5'1	2.22
2	DG	H1'	2	DG	H8	3.83
2	DG	H1'	3	DA	H8	2.54
2	DG	H2'1	2	DG	H8	2.49
2	DG	H2'2	2	DG	H8	2.65
2	DG	H3'	2	DG	H1'	3.41
2	DG	H3'	2	DG	H8	3.54
2	DG	H4'	13	BIC	HO	4.18
2	DG	H4'	2	DG	H1'	4.25
2	DG	H4'	2	DG	H3'	3.10
2	DG	H4'	2	DG	H8	3.73
2	DG	H5'1	2	DG	H1'	3.36
2	DG	H5'2	2	DG	H3'	2.78
2	DG	H8	1	DT	H5'2	0.66
2	DG	H8	3	DA	H8	2.14
3	DA	H1'	14	NGL	H52	2.48
3	DA	H1'	3	DA	H4'	3.86
3	DA	H1'	3	DA	H5'1	3.90
3	DA	H1'	4	DT	H6	3.05
3	DA	H2'1	3	DA	H2'2	1.69
3	DA	H2'1	3	DA	H8	2.02
3	DA	H2'2	3	DA	H8	2.91
3	DA	H2'2	4	DT	H6	3.25
3	DA	H3'	3	DA	H2'1	2.90
3	DA	H3'	3	DA	H2'2	3.54
3	DA	H3'	3	DA	H4'	2.70

Identifier A	Residue A	Atom A	Identifier B	Residue B	Atom B	Distance
3	DA	H3'	3	DA	H5'1	2.41
3	DA	H3'	3	DA	H8	3.06
4	DT	H1'	4	DT	H3'	3.57
4	DT	H1'	4	DT	H4'	3.90
4	DT	H1'	4	DT	H6	3.85
4	DT	H1'	5	DC	H6	3.26
4	DT	H2'1	14	NGL	H51	3.10
4	DT	H2'1	4	DT	H1'	3.56
4	DT	H2'1	4	DT	H2'2	2.78
4	DT	H2'1	4	DT	H6	2.54
4	DT	H2'1	5	DC	H6	2.77
4	DT	H2'2	4	DT	H1'	2.59
4	DT	H3'	4	DT	H2'1	3.94
4	DT	H3'	4	DT	H4'	3.71
4	DT	H3'	4	DT	H6	4.00
4	DT	H4'	4	DT	H3'	3.19
4	DT	H5'1	4	DT	H3'	1.38
4	DT	H5'2	4	DT	H3'	1.54
4	DT	H6	3	DA	H2'1	3.57
4	DT	H6	3	DA	H2'2	2.77
4	DT	H6	4	DT	H2'2	3.69
4	DT	H6	4	DT	H3'	3.56
4	DT	H7	3	DA	H2'1	3.88
4	DT	H7	3	DA	H2'2	3.93
4	DT	H7	3	DA	H8	3.42
4	DT	H7	4	DT	H6	3.12
5	DC	H2'1	5	DC	H1'	2.61
5	DC	H2'1	5	DC	H2'2	1.31
5	DC	H2'1	5	DC	H6	2.25
5	DC	H2'1	6	DA	H8	3.32
5	DC	H2'2	5	DC	H1'	3.21
5	DC	H2'2	5	DC	H2'1	2.56
5	DC	H2'2	5	DC	H3'	2.29
5	DC	H2'2	6	DA	H8	3.74
5	DC	H3'	5	DC	H2'1	4.08
5	DC	H3'	5	DC	H5'1	3.33
5	DC	H5	4	DT	H6	3.37
5	DC	H6	5	DC	H2'2	2.14
6	DA	H1'	6	DA	H3'	4.08
6	DA	H2'1	6	DA	H1'	4.21
6	DA	H2'1	6	DA	H2'2	1.67
6	DA	H2'1	6	DA	H8	2.64
6	DA	H2'2	6	DA	H1'	3.10
6	DA	H2'2	6	DA	H8	4.03
6	DA	H3'	6	DA	H4'	1.69
6	DA	H4'	6	DA	H8	2.10
13	BIC	H51	13	BIC	HC	1.90

Identifier A	Residue A	Atom A	Identifier B	Residue B	Atom B	Distance
13	BIC	HK2	13	BIC	HJ1	2.14
13	BIC	HK2	14	NGL	H51	2.58
13	BIC	HK2	1	DT	H7	4.29
14	NGL	H21	14	NGL	H12	2.94
14	NGL	H21	14	NGL	H22	1.71
14	NGL	H21	14	NGL	H31	3.12
14	NGL	H21	14	NGL	H32	1.04
14	NGL	H22	13	BIC	HM2	2.97
14	NGL	H22	14	NGL	H12	2.85
14	NGL	H22	5	DC	H1'	2.83
14	NGL	H22	6	DA	H8	3.61
14	NGL	H31	13	BIC	HJ1	1.97
14	NGL	H31	14	NGL	H12	2.50
14	NGL	H31	14	NGL	H41	1.49
14	NGL	H31	14	NGL	H51	2.11
14	NGL	H31	4	DT	H2'1	1.49
14	NGL	H32	14	NGL	H21	2.41
14	NGL	H32	14	NGL	H22	1.83
14	NGL	H32	14	NGL	H42	1.61
14	NGL	H32	14	NGL	H51	4.41
14	NGL	H32	4	DT	H1'	2.71
14	NGL	H41	14	NGL	H32	1.37
14	NGL	H41	14	NGL	H51	4.26
14	NGL	H42	14	NGL	H12	2.75
14	NGL	H51	14	NGL	H42	2.39
14	NGL	H52	13	BIC	HJ1	3.43
14	NGL	H52	14	NGL	H21	2.92
14	NGL	H52	14	NGL	H41	4.11
14	NGL	H52	14	NGL	H51	2.05
14	NGL	H52	2	DG	H1'	2.48
14	NGL	H52	2	DG	H4'	2.65
14	NGL	H52	3	DA	H2'1	2.06
14	NGL	H52	4	DT	H2'1	1.78
14	NGL	H52	4	DT	H5'1	2.81
7	DT	H1'	15	BIC	HO	3.37
7	DT	H1'	7	DT	H6	3.67
7	DT	H2'1	7	DT	H1'	2.75
7	DT	H2'2	15	BIC	HO	2.73
7	DT	H2'2	7	DT	H1'	2.85
7	DT	H2'2	7	DT	H3'	3.02
7	DT	H2'2	7	DT	H5'1	3.42
7	DT	H3'	7	DT	H6	2.63
7	DT	H6	15	BIC	HO	3.80
7	DT	H6	7	DT	H2'1	3.20
7	DT	H6	7	DT	H2'2	2.46
7	DT	H7	15	BIC	H11	2.46
7	DT	H7	7	DT	H6	3.10

Identifier A	Residue A	Atom A	Identifier B	Residue B	Atom B	Distance
7	DT	H7	8	DG	H4'	2.47
8	DG	H1'	15	BIC	HO	3.81
8	DG	H1'	8	DG	H2'2	3.53
8	DG	H1'	8	DG	H4'	3.93
8	DG	H1'	8	DG	H5'1	2.22
8	DG	H1'	8	DG	H8	3.83
8	DG	H1'	9	DA	H8	2.54
8	DG	H2'1	8	DG	H8	2.49
8	DG	H2'2	8	DG	H8	2.65
8	DG	H3'	8	DG	H1'	3.41
8	DG	H3'	8	DG	H8	3.54
8	DG	H4'	15	BIC	HO	4.18
8	DG	H4'	8	DG	H1'	4.25
8	DG	H4'	8	DG	H3'	3.10
8	DG	H4'	8	DG	H8	3.73
8	DG	H5'1	8	DG	H1'	3.36
8	DG	H5'2	8	DG	H3'	2.78
8	DG	H8	7	DT	H5'2	0.66
8	DG	H8	9	DA	H8	2.14
9	DA	H1'	16	NGL	H52	2.48
9	DA	H1'	9	DA	H4'	3.86
9	DA	H1'	9	DA	H5'1	3.90
9	DA	H1'	10	DT	H6	3.05
9	DA	H2'1	9	DA	H2'2	1.69
9	DA	H2'1	9	DA	H8	2.02
9	DA	H2'2	9	DA	H8	2.91
9	DA	H2'2	10	DT	H6	3.25
9	DA	H3'	9	DA	H2'1	2.90
9	DA	H3'	9	DA	H2'2	3.54
9	DA	H3'	9	DA	H4'	2.70
9	DA	H3'	9	DA	H5'1	2.41
9	DA	H3'	9	DA	H8	3.06
10	DT	H1'	10	DT	H3'	3.57
10	DT	H1'	10	DT	H4'	3.90
10	DT	H1'	10	DT	H6	3.85
10	DT	H1'	11	DC	H6	3.26
10	DT	H2'1	16	NGL	H51	3.10
10	DT	H2'1	10	DT	H1'	3.56
10	DT	H2'1	10	DT	H2'2	2.78
10	DT	H2'1	10	DT	H6	2.54
10	DT	H2'1	11	DC	H6	2.77
10	DT	H2'2	10	DT	H1'	2.59
10	DT	H3'	10	DT	H2'1	3.94
10	DT	H3'	10	DT	H4'	3.71
10	DT	H3'	10	DT	H6	4.00
10	DT	H4'	10	DT	H3'	3.19
10	DT	H5'1	10	DT	H3'	1.38

Identifier A	Residue A	Atom A	Identifier B	Residue B	Atom B	Distance
10	DT	H5'2	10	DT	H3'	1.54
10	DT	H6	9	DA	H2'1	3.57
10	DT	H6	9	DA	H2'2	2.77
10	DT	H6	10	DT	H2'2	3.69
10	DT	H6	10	DT	H3'	3.56
10	DT	H7	9	DA	H2'1	3.88
10	DT	H7	9	DA	H2'2	3.93
10	DT	H7	9	DA	H8	3.42
10	DT	H7	10	DT	H6	3.12
11	DC	H2'1	11	DC	H1'	2.61
11	DC	H2'1	11	DC	H2'2	1.31
11	DC	H2'1	11	DC	H6	2.25
11	DC	H2'1	12	DA	H8	3.32
11	DC	H2'2	11	DC	H1'	3.21
11	DC	H2'2	11	DC	H2'1	2.56
11	DC	H2'2	11	DC	H3'	2.29
11	DC	H2'2	12	DA	H8	3.74
11	DC	H3'	11	DC	H2'1	4.08
11	DC	H3'	11	DC	H5'1	3.33
11	DC	H5	10	DT	H6	3.37
11	DC	H6	11	DC	H2'2	2.14
12	DA	H1'	12	DA	H3'	4.08
12	DA	H2'1	12	DA	H1'	4.21
12	DA	H2'1	12	DA	H2'2	1.67
12	DA	H2'1	12	DA	H8	2.64
12	DA	H2'2	12	DA	H1'	3.10
12	DA	H2'2	12	DA	H8	4.03
12	DA	H3'	12	DA	H4'	1.69
12	DA	H4'	12	DA	H8	2.10
15	BIC	H51	15	BIC	HC	1.90
15	BIC	HK2	15	BIC	HJ1	2.14
15	BIC	HK2	16	NGL	H51	2.58
15	BIC	HK2	7	DT	H7	4.29
16	NGL	H21	16	NGL	H12	2.94
16	NGL	H21	16	NGL	H22	1.71
16	NGL	H21	16	NGL	H31	3.12
16	NGL	H21	16	NGL	H32	1.04
16	NGL	H22	15	BIC	HM2	2.97
16	NGL	H22	16	NGL	H12	2.85
16	NGL	H22	11	DC	H1'	2.83
16	NGL	H22	12	DA	H8	3.61
16	NGL	H31	15	BIC	HJ1	1.97
16	NGL	H31	16	NGL	H12	2.50
16	NGL	H31	16	NGL	H41	1.49
16	NGL	H31	16	NGL	H51	2.11
16	NGL	H31	10	DT	H2'1	1.49
16	NGL	H32	16	NGL	H21	2.41

Identifier A	Residue A	Atom A	Identifier B	Residue B	Atom B	Distance
16	NGL	H32	16	NGL	H22	1.83
16	NGL	H32	16	NGL	H42	1.61
16	NGL	H32	16	NGL	H51	4.41
16	NGL	H32	10	DT	H1'	2.71
16	NGL	H41	16	NGL	H32	1.37
16	NGL	H41	16	NGL	H51	4.26
16	NGL	H42	16	NGL	H12	2.75
16	NGL	H51	16	NGL	H42	2.39
16	NGL	H52	15	BIC	HJ1	3.43
16	NGL	H52	16	NGL	H21	2.92
16	NGL	H52	16	NGL	H41	4.11
16	NGL	H52	16	NGL	H51	2.05
16	NGL	H52	8	DG	H1'	2.48
16	NGL	H52	8	DG	H4'	2.65
16	NGL	H52	9	DA	H2'1	2.06
16	NGL	H52	10	DT	H2'1	1.78
16	NGL	H52	10	DT	H5'1	2.81

Appendix 2.2: Nogalamycin$_2$-d(CCTGATCAGG)$_2$

Identifier A	Residue A	Atom A	Identifier B	Residue B	Atom B	Distance
1	DC	H1'	1	DC	H6	2.86
1	DC	H1'	2	DC	H6	3.01
1	DC	H2'1	1	DC	H2'2	2.32
1	DC	H2'1	1	DC	H3'	2.96
1	DC	H2'1	1	DC	H5	3.05
1	DC	H2'1	1	DC	H6	2.54
1	DC	H2'1	2	DC	H6	2.22
1	DC	H2'2	1	DC	H2'1	2.41
1	DC	H2'2	1	DC	H5	2.99
1	DC	H3'	1	DC	H2'1	6.71
1	DC	H3'	1	DC	H6	3.52
1	DC	H3'	2	DC	H6	3.47
1	DC	H5	1	DC	H6	3.62
1	DC	H5'1	1	DC	H3'	4.89
1	DC	H5'1	1	DC	H6	4.62
1	DC	H5'2	1	DC	H3'	4.41
1	DC	H5'2	1	DC	H6	4.74
2	DC	H1'	2	DC	H2'2	3.09
2	DC	H1'	2	DC	H4'	4.21
2	DC	H1'	2	DC	H6	3.83
2	DC	H1'	3	DT	H6	4.65
2	DC	H2'1	2	DC	H1'	4.05
2	DC	H2'1	2	DC	H2'2	2.33
2	DC	H2'1	2	DC	H6	2.72
2	DC	H2'1	3	DT	H6	2.46
2	DC	H2'2	2	DC	H1'	2.85
2	DC	H2'2	2	DC	H2'1	2.19
2	DC	H2'2	2	DC	H4'	3.42
2	DC	H3'	2	DC	H2'1	4.96
2	DC	H3'	2	DC	H6	3.75
2	DC	H5	21	BIC	H11	3.03
2	DC	H6	1	DC	H2'2	2.45
2	DC	H6	2	DC	H5	2.69
3	DT	H1'	21	BIC	HO	3.37
3	DT	H1'	3	DT	H3'	5.22
3	DT	H1'	3	DT	H6	3.67
3	DT	H1'	4	DG	H5'1	7.18
3	DT	H2'1	3	DT	H1'	2.75
3	DT	H2'1	4	DG	H4'	7.33
3	DT	H2'2	21	BIC	HO	2.73
3	DT	H2'2	3	DT	H1'	2.85
3	DT	H2'2	3	DT	H3'	3.02
3	DT	H2'2	3	DT	H5'1	3.42
3	DT	H3'	21	BIC	HO	9.29
3	DT	H3'	3	DT	H6	2.63

Identifier A	Residue A	Atom A	Identifier B	Residue B	Atom B	Distance
3	DT	H4'	21	BIC	HO	5.23
3	DT	H5'2	3	DT	H2'2	4.95
3	DT	H6	21	BIC	HO	3.80
3	DT	H6	2	DC	H2'2	2.99
3	DT	H6	3	DT	H2'1	3.20
3	DT	H6	3	DT	H2'2	2.46
3	DT	H7	21	BIC	H11	2.46
3	DT	H7	2	DC	H2'1	2.47
3	DT	H7	2	DC	H2'2	4.22
3	DT	H7	2	DC	H3'	4.67
3	DT	H7	2	DC	H5	6.88
3	DT	H7	3	DT	H6	3.10
3	DT	H7	4	DG	H4'	2.47
4	DG	H1'	21	BIC	HO	3.81
4	DG	H1'	4	DG	H2'1	4.60
4	DG	H1'	4	DG	H2'2	3.53
4	DG	H1'	4	DG	H4'	3.93
4	DG	H1'	4	DG	H5'1	2.22
4	DG	H1'	4	DG	H8	3.83
4	DG	H1'	5	DA	H8	2.54
4	DG	H2'1	4	DG	H8	2.49
4	DG	H2'2	4	DG	H5'2	5.45
4	DG	H2'2	4	DG	H8	2.65
4	DG	H3'	4	DG	H1'	3.41
4	DG	H3'	4	DG	H8	3.54
4	DG	H4'	21	BIC	HO	4.18
4	DG	H4'	4	DG	H1'	4.25
4	DG	H4'	4	DG	H3'	3.10
4	DG	H4'	4	DG	H8	3.73
4	DG	H4'	5	DA	H8	6.46
4	DG	H5'1	3	DT	H1'	6.21
4	DG	H5'1	4	DG	H1'	3.36
4	DG	H5'1	5	DA	H8	6.56
4	DG	H5'2	4	DG	H3'	2.78
4	DG	H8	3	DT	H5'2	0.66
4	DG	H8	5	DA	H8	2.14
5	DA	H1'	22	NGL	H52	2.48
5	DA	H1'	5	DA	H3'	4.92
5	DA	H1'	5	DA	H4'	3.86
5	DA	H1'	5	DA	H5'1	3.90
5	DA	H1'	5	DA	H8	4.62
5	DA	H1'	6	DT	H6	3.05
5	DA	H2'1	22	NGL	H52	5.60
5	DA	H2'1	5	DA	H2'2	1.69
5	DA	H2'1	5	DA	H8	2.02
5	DA	H2'2	5	DA	H8	2.91
5	DA	H2'2	6	DT	H6	3.25

Identifier A	Residue A	Atom A	Identifier B	Residue B	Atom B	Distance
5	DA	H3'	5	DA	H2'1	2.90
5	DA	H3'	5	DA	H2'2	3.54
5	DA	H3'	5	DA	H4'	2.70
5	DA	H3'	5	DA	H5'1	2.41
5	DA	H3'	5	DA	H8	3.06
5	DA	H3'	6	DT	H6	5.05
5	DA	H8	5	DA	H4'	6.16
5	DA	H8	5	DA	H5'1	8.97
6	DT	H1'	6	DT	H3'	3.57
6	DT	H1'	6	DT	H4'	3.90
6	DT	H1'	6	DT	H6	3.85
6	DT	H1'	7	DC	H6	3.26
6	DT	H2'1	22	NGL	H51	3.10
6	DT	H2'1	6	DT	H1'	3.56
6	DT	H2'1	6	DT	H2'2	2.78
6	DT	H2'1	6	DT	H4'	7.62
6	DT	H2'1	6	DT	H6	2.54
6	DT	H2'1	7	DC	H5	4.93
6	DT	H2'1	7	DC	H6	2.77
6	DT	H2'2	6	DT	H1'	2.59
6	DT	H3'	6	DT	H2'1	3.94
6	DT	H3'	6	DT	H4'	3.71
6	DT	H3'	6	DT	H6	4.00
6	DT	H4'	21	BIC	HO	2.30
6	DT	H4'	6	DT	H2'2	8.24
6	DT	H4'	6	DT	H3'	3.19
6	DT	H5'1	6	DT	H3'	1.38
6	DT	H5'2	6	DT	H3'	1.54
6	DT	H6	5	DA	H2'1	3.57
6	DT	H6	5	DA	H2'2	2.77
6	DT	H6	5	DA	H3'	5.25
6	DT	H6	5	DA	H8	6.26
6	DT	H6	6	DT	H2'2	3.69
6	DT	H6	6	DT	H3'	3.56
6	DT	H6	7	DC	H6	5.60
6	DT	H7	5	DA	H1'	6.68
6	DT	H7	5	DA	H2'1	3.88
6	DT	H7	5	DA	H2'2	3.93
6	DT	H7	5	DA	H3'	5.69
6	DT	H7	5	DA	H8	3.42
6	DT	H7	6	DT	H6	3.12
6	DT	H7	7	DC	H5	8.63
7	DC	H1'	8	DA	H8	4.98
7	DC	H2'1	7	DC	H1'	2.61
7	DC	H2'1	7	DC	H2'2	1.31
7	DC	H2'1	7	DC	H5	6.02
7	DC	H2'1	7	DC	H6	2.25

Identifier A	Residue A	Atom A	Identifier B	Residue B	Atom B	Distance
7	DC	H2'1	8	DA	H8	3.32
7	DC	H2'2	7	DC	H1'	3.21
7	DC	H2'2	7	DC	H2'1	2.56
7	DC	H2'2	7	DC	H3'	2.29
7	DC	H2'2	8	DA	H8	3.74
7	DC	H3'	7	DC	H2'1	4.08
7	DC	H3'	7	DC	H5'1	3.33
7	DC	H4'	7	DC	H1'	6.39
7	DC	H5	6	DT	H1'	7.65
7	DC	H5	6	DT	H6	3.37
7	DC	H6	7	DC	H2'2	2.14
8	DA	H1'	8	DA	H3'	4.08
8	DA	H2'1	8	DA	H1'	4.21
8	DA	H2'1	8	DA	H2'2	1.67
8	DA	H2'1	8	DA	H8	2.64
8	DA	H2'2	8	DA	H1'	3.10
8	DA	H2'2	8	DA	H8	4.03
8	DA	H2'2	9	DG	H8	2.17
8	DA	H3'	8	DA	H4'	1.69
8	DA	H4'	8	DA	H1'	4.97
8	DA	H4'	8	DA	H8	2.10
8	DA	H8	7	DC	H3'	6.41
8	DA	H8	8	DA	H1'	4.63
8	DA	H8	8	DA	H3'	5.97
9	DG	H1'	9	DG	H2'1	3.13
9	DG	H2'1	10	DG	H8	2.00
9	DG	H2'1	9	DG	H3'	3.48
9	DG	H3'	9	DG	H2'2	3.89
9	DG	H4'	9	DG	H3'	5.19
9	DG	H4'	9	DG	H3'	8.48
9	DG	H8	9	DG	H2'1	2.39
10	DG	H2'1	10	DG	H1'	4.73
10	DG	H2'2	10	DG	H1'	3.49
10	DG	H3'	10	DG	H4'	2.50
10	DG	H3'	10	DG	H8	2.88
10	DG	H3'	21	BIC	HC	8.47
10	DG	H8	9	DG	H2'2	2.03
11	DC	H1'	11	DC	H6	2.86
11	DC	H1'	12	DC	H6	3.01
11	DC	H2'1	11	DC	H2'2	2.32
11	DC	H2'1	11	DC	H3'	2.96
11	DC	H2'1	11	DC	H5	3.05
11	DC	H2'1	11	DC	H6	2.54
11	DC	H2'1	12	DC	H6	2.22
11	DC	H2'2	11	DC	H2'1	2.41
11	DC	H2'2	11	DC	H5	2.99
11	DC	H3'	11	DC	H2'1	6.71

Identifier A	Residue A	Atom A	Identifier B	Residue B	Atom B	Distance
11	DC	H3'	11	DC	H6	3.52
11	DC	H3'	12	DC	H6	3.47
11	DC	H5	11	DC	H6	3.62
11	DC	H5'1	11	DC	H3'	4.89
11	DC	H5'1	11	DC	H6	4.62
11	DC	H5'2	11	DC	H3'	4.41
11	DC	H5'2	11	DC	H6	4.74
12	DC	H1'	12	DC	H2'2	3.09
12	DC	H1'	12	DC	H4'	4.21
12	DC	H1'	12	DC	H6	3.83
12	DC	H1'	13	DT	H6	4.65
12	DC	H2'1	12	DC	H1'	4.05
12	DC	H2'1	12	DC	H2'2	2.33
12	DC	H2'1	12	DC	H6	2.72
12	DC	H2'1	13	DT	H6	2.46
12	DC	H2'2	12	DC	H1'	2.85
12	DC	H2'2	12	DC	H2'1	2.19
12	DC	H2'2	12	DC	H4'	3.42
12	DC	H3'	12	DC	H2'1	4.96
12	DC	H3'	12	DC	H6	3.75
12	DC	H5	21	BIC	H11	3.03
12	DC	H6	11	DC	H2'2	2.45
12	DC	H6	12	DC	H5	2.69
13	DT	H1'	21	BIC	HO	3.37
13	DT	H1'	13	DT	H3'	5.22
13	DT	H1'	13	DT	H6	3.67
13	DT	H1'	14	DG	H5'1	7.18
13	DT	H2'1	13	DT	H1'	2.75
13	DT	H2'1	14	DG	H4'	7.33
13	DT	H2'2	21	BIC	HO	2.73
13	DT	H2'2	13	DT	H1'	2.85
13	DT	H2'2	13	DT	H3'	3.02
13	DT	H2'2	13	DT	H5'1	3.42
13	DT	H3'	21	BIC	HO	9.29
13	DT	H3'	13	DT	H6	2.63
13	DT	H4'	21	BIC	HO	5.23
13	DT	H5'2	13	DT	H2'2	4.95
13	DT	H6	21	BIC	HO	3.80
13	DT	H6	12	DC	H2'2	2.99
13	DT	H6	13	DT	H2'1	3.20
13	DT	H6	13	DT	H2'2	2.46
13	DT	H7	21	BIC	H11	2.46
13	DT	H7	12	DC	H2'1	2.47
13	DT	H7	12	DC	H2'2	4.22
13	DT	H7	12	DC	H3'	4.67
13	DT	H7	12	DC	H5	6.88
13	DT	H7	13	DT	H6	3.10

Identifier A	Residue A	Atom A	Identifier B	Residue B	Atom B	Distance
13	DT	H7	14	DG	H4'	2.47
14	DG	H1'	21	BIC	HO	3.81
14	DG	H1'	14	DG	H2'1	4.60
14	DG	H1'	14	DG	H2'2	3.53
14	DG	H1'	14	DG	H4'	3.93
14	DG	H1'	14	DG	H5'1	2.22
14	DG	H1'	14	DG	H8	3.83
14	DG	H1'	15	DA	H8	2.54
14	DG	H2'1	14	DG	H8	2.49
14	DG	H2'2	14	DG	H5'2	5.45
14	DG	H2'2	14	DG	H8	2.65
14	DG	H3'	14	DG	H1'	3.41
14	DG	H3'	14	DG	H8	3.54
14	DG	H4'	21	BIC	HO	4.18
14	DG	H4'	14	DG	H1'	4.25
14	DG	H4'	14	DG	H3'	3.10
14	DG	H4'	14	DG	H8	3.73
14	DG	H4'	15	DA	H8	6.46
14	DG	H5'1	13	DT	H1'	6.21
14	DG	H5'1	14	DG	H1'	3.36
14	DG	H5'1	15	DA	H8	6.56
14	DG	H5'2	14	DG	H3'	2.78
14	DG	H8	13	DT	H5'2	0.66
14	DG	H8	15	DA	H8	2.14
15	DA	H1'	24	NGL	H12	2.48
15	DA	H1'	15	DA	H3'	4.92
15	DA	H1'	15	DA	H4'	3.86
15	DA	H1'	15	DA	H5'1	3.90
15	DA	H1'	15	DA	H8	4.62
15	DA	H1'	16	DT	H6	3.05
15	DA	H2'1	24	NGL	H12	5.60
15	DA	H2'1	15	DA	H2'2	1.69
15	DA	H2'1	15	DA	H8	2.02
15	DA	H2'2	15	DA	H8	2.91
15	DA	H2'2	16	DT	H6	3.25
15	DA	H3'	15	DA	H2'1	2.90
15	DA	H3'	15	DA	H2'2	3.54
15	DA	H3'	15	DA	H4'	2.70
15	DA	H3'	15	DA	H5'1	2.41
15	DA	H3'	15	DA	H8	3.06
15	DA	H3'	16	DT	H6	5.05
15	DA	H8	15	DA	H4'	6.16
15	DA	H8	15	DA	H5'1	8.97
16	DT	H1'	16	DT	H3'	3.57
16	DT	H1'	16	DT	H4'	3.90
16	DT	H1'	16	DT	H6	3.85
16	DT	H1'	17	DC	H6	3.26

Identifier A	Residue A	Atom A	Identifier B	Residue B	Atom B	Distance
16	DT	H2'1	24	NGL	H12	3.10
16	DT	H2'1	16	DT	H1'	3.56
16	DT	H2'1	16	DT	H2'2	2.78
16	DT	H2'1	16	DT	H4'	7.62
16	DT	H2'1	16	DT	H6	2.54
16	DT	H2'1	17	DC	H5	4.93
16	DT	H2'1	17	DC	H6	2.77
16	DT	H2'2	16	DT	H1'	2.59
16	DT	H3'	16	DT	H2'1	3.94
16	DT	H3'	16	DT	H4'	3.71
16	DT	H3'	16	DT	H6	4.00
16	DT	H4'	21	BIC	HO	2.30
16	DT	H4'	16	DT	H2'2	8.24
16	DT	H4'	16	DT	H3'	3.19
16	DT	H5'1	16	DT	H3'	1.38
16	DT	H5'2	16	DT	H3'	1.54
16	DT	H6	15	DA	H2'1	3.57
16	DT	H6	15	DA	H2'2	2.77
16	DT	H6	15	DA	H3'	5.25
16	DT	H6	15	DA	H8	6.26
16	DT	H6	16	DT	H2'2	3.69
16	DT	H6	16	DT	H3'	3.56
16	DT	H6	17	DC	H6	5.60
16	DT	H7	15	DA	H1'	6.68
16	DT	H7	15	DA	H2'1	3.88
16	DT	H7	15	DA	H2'2	3.93
16	DT	H7	15	DA	H3'	5.69
16	DT	H7	15	DA	H8	3.42
16	DT	H7	16	DT	H6	3.12
16	DT	H7	17	DC	H5	8.63
17	DC	H1'	18	DA	H8	4.98
17	DC	H2'1	17	DC	H1'	2.61
17	DC	H2'1	17	DC	H2'2	1.31
17	DC	H2'1	17	DC	H5	6.02
17	DC	H2'1	17	DC	H6	2.25
17	DC	H2'1	18	DA	H8	3.32
17	DC	H2'2	17	DC	H1'	3.21
17	DC	H2'2	17	DC	H2'1	2.56
17	DC	H2'2	17	DC	H3'	2.29
17	DC	H2'2	18	DA	H8	3.74
17	DC	H3'	17	DC	H2'1	4.08
17	DC	H3'	17	DC	H5'1	3.33
17	DC	H4'	17	DC	H1'	6.39
17	DC	H5	16	DT	H1'	7.65
17	DC	H5	16	DT	H6	3.37
17	DC	H6	17	DC	H2'2	2.14
18	DA	H1'	18	DA	H3'	4.08

Identifier A	Residue A	Atom A	Identifier B	Residue B	Atom B	Distance
18	DA	H2'1	18	DA	H1'	4.21
18	DA	H2'1	18	DA	H2'2	1.67
18	DA	H2'1	18	DA	H8	2.64
18	DA	H2'2	18	DA	H1'	3.10
18	DA	H2'2	18	DA	H8	4.03
18	DA	H2'2	19	DG	H8	2.17
18	DA	H3'	18	DA	H4'	1.69
18	DA	H4'	18	DA	H1'	4.97
18	DA	H4'	18	DA	H8	2.10
18	DA	H8	17	DC	H3'	6.41
18	DA	H8	18	DA	H1'	4.63
18	DA	H8	18	DA	H3'	5.97
19	DG	H1'	19	DG	H2'1	3.13
19	DG	H2'1	20	DG	H8	2.00
19	DG	H2'1	19	DG	H3'	3.48
19	DG	H3'	19	DG	H2'2	3.89
19	DG	H4'	19	DG	H3'	5.19
19	DG	H4'	19	DG	H3'	8.48
19	DG	H8	19	DG	H2'1	2.39
20	DG	H2'1	20	DG	H1'	4.73
20	DG	H2'2	20	DG	H1'	3.49
20	DG	H3'	20	DG	H4'	2.50
20	DG	H3'	20	DG	H8	2.88
20	DG	H3'	23	BIC	HC	8.47
20	DG	H8	19	DG	H2'2	2.03
21	BIC	H11	21	BIC	H51	8.03
21	BIC	H11	21	BIC	HS2	5.72
21	BIC	H11	3	DT	H6	7.58
21	BIC	H51	21	BIC	H11	4.77
21	BIC	H51	21	BIC	HC	1.90
21	BIC	H51	9	DG	H3'	1.24
21	BIC	HK2	21	BIC	HJ1	2.14
21	BIC	HK2	22	NGL	H51	2.58
21	BIC	HK2	3	DT	H7	4.29
21	BIC	HO	22	NGL	H51	8.61
21	BIC	HO	3	DT	H4'	8.12
21	BIC	HO	4	DG	H8	6.39
21	BIC	HS2	21	BIC	H11	7.10
21	BIC	HS2	21	BIC	HC	4.73
22	NGL	H21	22	NGL	H22	2.94
22	NGL	H21	22	NGL	H22	1.71
22	NGL	H21	22	NGL	H31	3.12
22	NGL	H21	22	NGL	H32	1.04
22	NGL	H22	21	BIC	HM2	2.97
22	NGL	H22	22	NGL	H22	2.85
22	NGL	H22	7	DC	H1'	2.83
22	NGL	H22	8	DA	H4'	5.08

Identifier A	Residue A	Atom A	Identifier B	Residue B	Atom B	Distance
22	NGL	H22	8	DA	H8	3.61
22	NGL	H31	21	BIC	HJ1	1.97
22	NGL	H31	22	NGL	H22	2.50
22	NGL	H31	22	NGL	H41	1.49
22	NGL	H31	22	NGL	H51	2.11
22	NGL	H31	5	DA	H3'	5.65
22	NGL	H31	6	DT	H2'1	1.49
22	NGL	H32	22	NGL	H22	6.39
22	NGL	H32	22	NGL	H21	2.41
22	NGL	H32	22	NGL	H22	1.83
22	NGL	H32	22	NGL	H42	1.61
22	NGL	H32	22	NGL	H51	4.41
22	NGL	H32	5	DA	H1'	8.81
22	NGL	H32	6	DT	H1'	2.71
22	NGL	H32	6	DT	H1'	4.73
22	NGL	H41	22	NGL	H32	1.37
22	NGL	H41	22	NGL	H51	4.26
22	NGL	H42	21	BIC	HM2	2.18
22	NGL	H42	22	NGL	H22	2.75
22	NGL	H42	7	DC	H1'	3.34
22	NGL	H42	8	DA	H4'	6.11
22	NGL	H51	22	NGL	H42	2.39
22	NGL	H51	3	DT	H1'	3.74
22	NGL	H52	21	BIC	HJ1	3.43
22	NGL	H52	22	NGL	H22	4.64
22	NGL	H52	22	NGL	H21	2.92
22	NGL	H52	22	NGL	H41	4.11
22	NGL	H52	22	NGL	H51	2.05
22	NGL	H52	4	DG	H1'	2.48
22	NGL	H52	4	DG	H4'	2.65
22	NGL	H52	5	DA	H2'1	2.06
22	NGL	H52	5	DA	H2'2	5.74
22	NGL	H52	5	DA	H3'	5.51
22	NGL	H52	6	DT	H2'1	1.78
22	NGL	H52	6	DT	H5'1	2.81
22	NGL	H52	6	DT	H7	4.67
23	BIC	H21	23	BIC	H51	8.03
23	BIC	H21	23	BIC	HS2	5.72
23	BIC	H21	13	DT	H6	7.58
23	BIC	H51	23	BIC	H21	4.77
23	BIC	H51	23	BIC	HC	1.90
23	BIC	H51	19	DG	H3'	1.24
23	BIC	HK2	23	BIC	HJ1	2.14
23	BIC	HK2	24	NGL	H51	2.58
23	BIC	HK2	13	DT	H7	4.29
23	BIC	HO	24	NGL	H51	8.61
23	BIC	HO	13	DT	H4'	8.12

Identifier A	Residue A	Atom A	Identifier B	Residue B	Atom B	Distance
23	BIC	HO	14	DG	H8	6.39
23	BIC	HS2	23	BIC	H21	7.10
23	BIC	HS2	23	BIC	HC	4.73
24	NGL	H21	24	NGL	H22	2.94
24	NGL	H21	24	NGL	H22	1.71
24	NGL	H21	24	NGL	H31	3.12
24	NGL	H21	24	NGL	H32	1.04
24	NGL	H22	23	BIC	HM2	2.97
24	NGL	H22	24	NGL	H22	2.85
24	NGL	H22	17	DC	H1'	2.83
24	NGL	H22	18	DA	H4'	5.08
24	NGL	H22	18	DA	H8	3.61
24	NGL	H31	23	BIC	HJ1	1.97
24	NGL	H31	24	NGL	H22	2.50
24	NGL	H31	24	NGL	H41	1.49
24	NGL	H31	24	NGL	H51	2.11
24	NGL	H31	15	DA	H3'	5.65
24	NGL	H31	16	DT	H2'1	1.49
24	NGL	H32	24	NGL	H22	6.39
24	NGL	H32	24	NGL	H21	2.41
24	NGL	H32	24	NGL	H22	1.83
24	NGL	H32	24	NGL	H42	1.61
24	NGL	H32	24	NGL	H51	4.41
24	NGL	H32	15	DA	H1'	8.81
24	NGL	H32	16	DT	H1'	2.71
24	NGL	H32	16	DT	H1'	4.73
24	NGL	H41	24	NGL	H32	1.37
24	NGL	H41	24	NGL	H51	4.26
24	NGL	H42	23	BIC	HM2	2.18
24	NGL	H42	24	NGL	H22	2.75
24	NGL	H42	17	DC	H1'	3.34
24	NGL	H42	18	DA	H4'	6.11
24	NGL	H51	24	NGL	H42	2.39
24	NGL	H51	13	DT	H1'	3.74
24	NGL	H52	23	BIC	HJ1	3.43
24	NGL	H52	24	NGL	H22	4.64
24	NGL	H52	24	NGL	H21	2.92
24	NGL	H52	24	NGL	H41	4.11
24	NGL	H52	24	NGL	H51	2.05
24	NGL	H52	14	DG	H1'	2.48
24	NGL	H52	14	DG	H4'	2.65
24	NGL	H52	15	DA	H2'1	2.06
24	NGL	H52	15	DA	H2'2	5.74
24	NGL	H52	15	DA	H3'	5.51
24	NGL	H52	16	DT	H2'1	1.78
24	NGL	H52	16	DT	H5'1	2.81
24	NGL	H52	16	DT	H7	4.67

Appendix 2.3: Nogalamycin-d(GCTACGAAGTGC)

Identifier A	Residue A	Atom A	Identifier B	Residue B	Atom B	Distance
1	DG	H1'	2	DC	H2'1	2.60
2	DC	H2'1	2	DC	H2'2	1.85
2	DC	H3'	2	DC	H2'2	1.87
2	DC	H42	2	DC	H41	1.74
2	DC	H5	2	DC	H41	2.83
2	DC	H5	2	DC	H42	2.57
3	DC	H2'1	3	DC	H1'	3.06
3	DC	H2'1	3	DC	H2'2	1.92
3	DC	H2'1	3	DC	H6	3.07
3	DC	H2'2	3	DC	H1'	2.27
3	DC	H3'	3	DC	H2'1	2.41
3	DC	H3'	3	DC	H2'2	3.70
3	DC	H4'	2	DC	H5'2	2.81
3	DC	H5'2	3	DC	H1'	2.95
4	DC	H1'	5	DG	H5'2	2.78
4	DC	H2'1	4	DC	H2'2	1.78
4	DC	H2'1	5	DG	H8	3.68
4	DC	H2'2	4	DC	H1'	2.38
4	DC	H2'2	4	DC	H6	3.08
4	DC	H2'2	5	DG	H8	2.98
4	DC	H3'	4	DC	H2'2	3.18
4	DC	H4'	4	DC	H5'2	1.07
4	DC	H42	4	DC	H41	1.73
4	DC	H5	4	DC	H41	3.58
4	DC	H5	4	DC	H42	2.44
4	DC	H5	4	DC	H6	2.60
5	DG	H1'	5	DG	H8	4.41
5	DG	H2'1	5	DG	H1'	2.94
5	DG	H2'1	5	DG	H2'2	2.18
5	DG	H2'1	5	DG	H8	3.41
5	DG	H2'1	6	DT	H6	2.84
5	DG	H2'2	5	DG	H1'	3.50
5	DG	H2'2	5	DG	H8	2.36
5	DG	H3'	5	DG	H2'1	2.53
5	DG	H3'	5	DG	H2'2	2.49
6	DT	H1'	6	DT	H2'2	2.11
6	DT	H2'1	6	DT	H2'2	1.70
6	DT	H2'1	6	DT	H3'	3.03
6	DT	H2'1	6	DT	H6	2.31
6	DT	H2'2	6	DT	H3'	3.51
6	DT	H2'2	6	DT	H5'1	2.97
6	DT	H2'2	6	DT	H5'2	2.75
6	DT	H2'2	6	DT	H6	3.48
6	DT	H3'	6	DT	H5'2	2.68
6	DT	H7	5	DG	H2'1	3.74

Identifier A	Residue A	Atom A	Identifier B	Residue B	Atom B	Distance
7	DA	H1'	7	DA	H8	3.22
7	DA	H2'1	7	DA	H1'	2.81
7	DA	H2'1	7	DA	H2'2	2.09
7	DA	H2'1	7	DA	H8	2.36
7	DA	H2'2	7	DA	H1'	2.40
7	DA	H3'	7	DA	H2'1	2.83
7	DA	H3'	7	DA	H2'2	2.92
7	DA	H5'1	7	DA	H8	3.25
8	DG	H1'	8	DG	H2'1	2.70
8	DG	H1'	8	DG	H8	4.90
8	DG	H1'	9	DT	H6	3.74
8	DG	H2'1	8	DG	H2'2	3.90
8	DG	H2'1	8	DG	H8	2.41
8	DG	H2'2	8	DG	H1'	4.26
9	DT	H3'	9	DT	H6	3.43
9	DT	H5'2	9	DT	H6	4.65
9	DT	H7	8	DG	H2'1	4.33
9	DT	H7	8	DG	H2'2	3.65
10	DG	H1'	10	DG	H2'2	2.34
10	DG	H1'	10	DG	H8	3.46
10	DG	H2'1	10	DG	H2'2	2.60
10	DG	H2'2	10	DG	H8	2.69
11	DC	H1'	11	DC	H2'1	3.13
11	DC	H2'1	11	DC	H2'2	2.12
11	DC	H2'2	11	DC	H1'	2.55
11	DC	H2'2	11	DC	H6	4.26
11	DC	H3'	11	DC	H2'1	2.44
11	DC	H3'	11	DC	H2'2	2.78
11	DC	H3'	11	DC	H6	3.62
11	DC	H4'	11	DC	H6	3.60
11	DC	H42	11	DC	H41	1.71
11	DC	H5	11	DC	H42	2.51
11	DC	H5	11	DC	H6	2.97
12	BIC	H21	12	BIC	H11	3.49
12	BIC	H51	12	BIC	H41	0.63
12	BIC	HJ1	13	NGL	H21	2.53
12	BIC	HK2	13	NGL	H41	3.86
12	BIC	HL2	11	DC	H4'	2.88
12	BIC	HL2	13	NGL	H22	3.80
13	NGL	H21	13	NGL	H22	5.15
13	NGL	H22	13	NGL	H51	2.26
13	NGL	H31	12	BIC	HK2	2.70
13	NGL	H31	13	NGL	H32	3.31
13	NGL	H31	13	NGL	H51	4.58
13	NGL	H42	2	DC	H5'2	2.75
13	NGL	H51	13	NGL	H21	2.32

Appendix 2.4: Nogalamycin-d(GCCCGTAGTGC)

Identifier A	Residue A	Atom A	Identifier B	Residue B	Atom B	Distance
1	DG	H1'	1	DG	H8	4.46
1	DG	H1'	2	DC	H6	3.82
1	DG	H2'1	1	DG	H1'	3.73
1	DG	H2'1	1	DG	H5'1	3.51
1	DG	H2'1	2	DC	H5	4.38
2	DC	H6	1	DG	H2'1	4.90
1	DG	H2'2	1	DG	H1'	2.87
1	DG	H3'	1	DG	H2'2	3.45
1	DG	H8	1	DG	H2'2	3.93
1	DG	H2'2	2	DC	H5	3.59
2	DC	H6	1	DG	H2'2	3.14
1	DG	H3'	1	DG	H8	6.15
1	DG	H1'	1	DG	H4'	3.42
1	DG	H5'1	1	DG	H3'	3.14
1	DG	H5'1	1	DG	H4'	2.86
1	DG	H5'1	1	DG	H8	3.68
10	DT	H1'	10	DT	H6	5.03
10	DT	H1'	13	BIC	HO	4.51
10	DT	H3'	10	DT	H6	4.23
10	DT	H7	10	DT	H6	3.25
10	DT	H7	13	BIC	H11	3.28
13	BIC	H51	10	DT	H7	3.35
10	DT	H7	9	DG	H1'	6.24
10	DT	H7	9	DG	H2'1	3.52
10	DT	H7	9	DG	H8	3.89
11	DG	H1'	11	DG	H8	5.11
11	DG	H1'	12	DC	H6	3.27
11	DG	H1'	13	BIC	HO	4.04
11	DG	H2'1	11	DG	H1'	2.85
11	DG	H2'1	12	DC	H5	3.11
12	DC	H6	11	DG	H2'1	2.87
11	DG	H3'	11	DG	H8	4.33
11	DG	H4'	11	DG	H1'	3.69
11	DG	H3'	11	DG	H4'	2.99
11	DG	H4'	13	BIC	HO	4.80
11	DG	H5'1	11	DG	H1'	3.28
11	DG	H5'1	11	DG	H8	4.37
12	DC	H1'	12	DC	H6	4.08
3	DT	H7	1	DG	H1'	4.02
12	DC	H2'1	12	DC	H1'	3.41
12	DC	H2'1	12	DC	H2'2	2.23
12	DC	H2'1	12	DC	H5'1	3.88
12	DC	H2'1	12	DC	H6	2.96
12	DC	H2'2	12	DC	H3'	3.69
12	DC	H2'2	12	DC	H5'1	3.74

Identifier A	Residue A	Atom A	Identifier B	Residue B	Atom B	Distance
12	DC	H6	12	DC	H2'2	4.58
12	DC	H3'	12	DC	H6	3.29
12	DC	H4'	12	DC	H6	4.63
12	DC	H5'1	12	DC	H1'	3.65
12	DC	H5'1	12	DC	H3'	3.25
12	DC	H5'1	12	DC	H4'	2.59
12	DC	H5'2	12	DC	H6	3.89
12	DC	H5	12	DC	H6	2.73
13	BIC	H11	10	DT	H6	4.30
13	BIC	H11	13	BIC	H21	2.60
13	BIC	H32	13	BIC	H21	4.03
13	BIC	H32	13	BIC	H11	4.40
13	BIC	H41	13	BIC	H11	4.30
13	BIC	H41	13	BIC	H32	2.19
13	BIC	H41	13	BIC	HC	4.59
13	BIC	H51	13	BIC	H11	4.32
13	BIC	H51	13	BIC	H41	2.57
13	BIC	HK1	13	BIC	HJ1	2.52
13	BIC	HK1	14	NGL	H12	4.45
13	BIC	HK1	14	NGL	H22	2.60
13	BIC	HJ1	13	BIC	HK2	2.54
14	NGL	H12	13	BIC	HK2	4.42
13	BIC	HK2	14	NGL	H22	3.51
13	BIC	HL2	11	DG	H1'	3.00
13	BIC	HL2	11	DG	H4'	3.28
13	BIC	HL2	12	DC	H1'	3.65
13	BIC	HL2	12	DC	H5'1	3.10
13	BIC	HK1	13	BIC	HL2	3.02
13	BIC	HL2	13	BIC	HK2	3.66
13	BIC	HL2	13	BIC	HM2	2.48
13	BIC	HL2	13	BIC	HO	4.49
13	BIC	HL2	14	NGL	H21	5.47
13	BIC	HL2	14	NGL	H22	2.97
13	BIC	HM2	13	BIC	HO	2.61
13	BIC	HS2	10	DT	H1'	3.94
4	DA	H1'	14	NGL	H12	3.51
13	BIC	HJ1	14	NGL	H21	4.12
14	NGL	H12	14	NGL	H21	2.80
14	NGL	H22	14	NGL	H21	3.20
14	NGL	H41	14	NGL	H21	2.33
2	DC	H1'	14	NGL	H21	4.79
14	NGL	H22	14	NGL	H12	4.08
14	NGL	H31	12	DC	H1'	2.70
14	NGL	H31	12	DC	H2'2	3.93
14	NGL	H31	12	DC	H5'1	3.23
14	NGL	H31	13	BIC	HK1	4.55
14	NGL	H31	14	NGL	H12	4.39

Identifier A	Residue A	Atom A	Identifier B	Residue B	Atom B	Distance
14	NGL	H31	14	NGL	H22	2.44
14	NGL	H31	14	NGL	H41	3.62
14	NGL	H41	14	NGL	H12	2.55
14	NGL	H41	2	DC	H4'	4.45
14	NGL	H41	4	DA	H8	3.54
14	NGL	H52	13	BIC	HK1	4.32
13	BIC	HK2	14	NGL	H52	4.40
14	NGL	H52	14	NGL	H12	5.00
14	NGL	H52	14	NGL	H41	2.81
14	NGL	H52	14	NGL	H42	2.44
14	NGL	H52	14	NGL	H42	2.59
14	NGL	H52	14	NGL	H42	2.59
14	NGL	H52	14	NGL	H42	2.75
14	NGL	H52	4	DA	H4'	3.72
2	DC	H1'	2	DC	H6	4.46
2	DC	H2'1	2	DC	H1'	3.47
2	DC	H3'	2	DC	H2'1	2.78
2	DC	H2'1	2	DC	H5'1	3.94
2	DC	H6	2	DC	H2'1	2.52
2	DC	H2'2	14	NGL	H41	3.63
2	DC	H2'2	2	DC	H4'	4.26
2	DC	H2'2	2	DC	H5'1	4.23
2	DC	H6	2	DC	H2'2	3.43
2	DC	H1'	2	DC	H3'	4.54
2	DC	H3'	2	DC	H6	3.50
2	DC	H4'	2	DC	H1'	3.59
2	DC	H5'1	2	DC	H6	4.59
2	DC	H5'1	3	DT	H6	4.00
2	DC	H5	1	DG	H8	4.10
2	DC	H5	2	DC	H6	2.69
3	DT	H1'	3	DT	H6	3.66
4	DA	H8	3	DT	H1'	5.10
3	DT	H6	3	DT	H2'1	2.75
3	DT	H2'2	3	DT	H6	3.80
3	DT	H4'	3	DT	H1'	4.07
3	DT	H3'	3	DT	H4'	3.26
3	DT	H7	2	DC	H3'	5.52
3	DT	H7	2	DC	H6	4.29
3	DT	H7	3	DT	H6	3.30
4	DA	H1'	4	DA	H8	4.80
4	DA	H2'1	4	DA	H2'2	2.27
4	DA	H2'1	4	DA	H8	2.85
5	DC	H6	4	DA	H2'1	4.73
4	DA	H2'2	4	DA	H1'	2.79
4	DA	H3'	4	DA	H2'2	3.42
4	DA	H2'2	4	DA	H8	4.33
5	DC	H5'1	4	DA	H2'2	3.83

Identifier A	Residue A	Atom A	Identifier B	Residue B	Atom B	Distance
5	DC	H6	4	DA	H2'2	3.45
4	DA	H3'	4	DA	H8	4.15
5	DC	H2'1	5	DC	H2'2	2.17
5	DC	H2'2	5	DC	H3'	3.54
5	DC	H6	5	DC	H2'2	4.56
5	DC	H1'	5	DC	H4'	3.64
5	DC	H5'1	5	DC	H3'	3.25
6	DG	H4'	7	DA	H2'1	3.58
7	DA	H2'1	7	DA	H2'2	2.07
7	DA	H3'	7	DA	H2'1	4.11
7	DA	H2'2	7	DA	H3'	3.33
7	DA	H4'	7	DA	H2'2	3.31
7	DA	H4'	7	DA	H3'	3.33
7	DA	H5'1	7	DA	H3'	2.68
7	DA	H8	7	DA	H5'1	2.66
8	DA	H1'	9	DG	H8	4.92
8	DA	H2'1	8	DA	H1'	3.11
8	DA	H3'	8	DA	H2'1	3.20
9	DG	H1'	10	DT	H6	4.09
9	DG	H2'1	9	DG	H1'	2.92
9	DG	H2'1	9	DG	H4'	3.96
9	DG	H4'	9	DG	H1'	3.60

Appendix 2.5: d(TAGGAGGT)$_4$

Identifier A	Residue A	Atom A	Identifier B	Residue B	Atom B	Distance
1	DT	H2'1	1	DT	H2'2	2.03
1	DT	H3'	1	DT	H2'1	2.77
1	DT	H2'2	1	DT	H3'	2.72
1	DT	H5'1	1	DT	H2'2	5.03
1	DT	H4'	1	DT	H3'	2.57
1	DT	H5'1	1	DT	H4'	2.83
1	DT	H5'1	1	DT	H3'	2.80
2	DA	H2'1	2	DA	H1'	2.58
2	DA	H2'1	2	DA	H3'	2.25
3	DG	H2'1	3	DG	H1'	3.21
3	DG	H2'1	3	DG	H2'2	1.82
3	DG	H4'	3	DG	H2'1	6.08
3	DG	H5'1	3	DG	H2'1	5.35
3	DG	H5'2	3	DG	H2'1	5.94
3	DG	H2'2	3	DG	H1'	2.47
3	DG	H2'2	3	DG	H3'	2.60
3	DG	H3'	3	DG	H5'2	2.70
4	DG	H2'1	4	DG	H1'	3.03
4	DG	H2'1	4	DG	H2'2	1.84
4	DG	H2'1	4	DG	H3'	2.13
4	DG	H4'	4	DG	H2'1	5.61
4	DG	H2'2	4	DG	H1'	2.24
4	DG	H2'2	4	DG	H3'	2.55
4	DG	H4'	4	DG	H2'2	5.01
4	DG	H1'	4	DG	H3'	4.30
4	DG	H5'1	4	DG	H3'	3.64
4	DG	H4'	4	DG	H5'1	2.68
4	DG	H5'2	4	DG	H3'	2.62
5	DA	H2'1	5	DA	H1'	2.26
5	DA	H2'1	5	DA	H2'2	1.85
5	DA	H2'1	5	DA	H3'	2.63
5	DA	H2'2	5	DA	H1'	2.99
5	DA	H2'2	5	DA	H3'	2.40
5	DA	H5'2	5	DA	H2'2	5.67
5	DA	H1'	5	DA	H3'	4.38
5	DA	H3'	5	DA	H4'	4.77
5	DA	H1'	5	DA	H5'1	3.72
5	DA	H5'1	5	DA	H3'	2.63
6	DG	H2'2	6	DG	H1'	2.31
6	DG	H2'2	6	DG	H3'	2.57
6	DG	H4'	6	DG	H2'2	6.50
6	DG	H5'1	6	DG	H2'2	5.33
6	DG	H3'	6	DG	H1'	4.75
6	DG	H4'	6	DG	H1'	3.58
6	DG	H4'	6	DG	H3'	2.70

Identifier A	Residue A	Atom A	Identifier B	Residue B	Atom B	Distance
6	DG	H5'1	6	DG	H1'	3.84
6	DG	H5'1	6	DG	H3'	3.36
6	DG	H5'2	6	DG	H3'	2.69
6	DG	H4'	6	DG	H5'2	2.51
7	DG	H2'1	7	DG	H1'	2.74
7	DG	H2'1	7	DG	H2'2	1.88
7	DG	H2'2	7	DG	H1'	2.33
7	DG	H2'2	7	DG	H3'	2.66
7	DG	H4'	7	DG	H2'2	5.71
7	DG	H4'	7	DG	H1'	3.04
8	DT	H2'1	8	DT	H1'	2.09
2	DA	H8	2	DA	H1'	8.02
3	DG	H8	2	DA	H1'	9.38
3	DG	H8	3	DG	H1'	5.77
3	DG	H1'	4	DG	H8	4.46
3	DG	H8	3	DG	H2'1	3.01
3	DG	H2'1	4	DG	H8	4.26
3	DG	H8	3	DG	H2'2	4.27
4	DG	H8	4	DG	H1'	5.96
4	DG	H2'1	5	DA	H8	3.51
4	DG	H2'2	5	DA	H8	3.01
5	DA	H2'1	5	DA	H8	3.91
5	DA	H2'2	5	DA	H8	2.68
5	DA	H3'	5	DA	H8	4.25
6	DG	H8	6	DG	H1'	3.93
7	DG	H8	6	DG	H1'	4.82
6	DG	H8	6	DG	H2'1	2.94
6	DG	H8	6	DG	H2'2	3.75
7	DG	H8	7	DG	H1'	5.47
7	DG	H2'1	7	DG	H8	2.75
8	DT	H6	7	DG	H2'1	7.20
7	DG	H8	7	DG	H2'2	3.55
8	DT	H6	8	DT	H1'	4.34
8	DT	H6	8	DT	H2'1	3.29
9	DT	H2'1	9	DT	H2'2	2.03
9	DT	H3'	9	DT	H2'1	2.77
9	DT	H2'2	9	DT	H3'	2.72
9	DT	H5'1	9	DT	H2'2	5.03
9	DT	H4'	9	DT	H3'	2.57
9	DT	H5'1	9	DT	H4'	2.83
9	DT	H5'1	9	DT	H3'	2.80
10	DA	H2'1	10	DA	H1'	2.58
10	DA	H2'1	10	DA	H3'	2.25
11	DG	H2'1	11	DG	H1'	3.21
11	DG	H2'1	11	DG	H2'2	1.82
11	DG	H4'	11	DG	H2'1	6.08
11	DG	H5'1	11	DG	H2'1	5.35

Identifier A	Residue A	Atom A	Identifier B	Residue B	Atom B	Distance
11	DG	H5'2	11	DG	H2'1	5.94
11	DG	H2'2	11	DG	H1'	2.47
11	DG	H2'2	11	DG	H3'	2.60
11	DG	H3'	11	DG	H5'2	2.70
12	DG	H2'1	12	DG	H1'	3.03
12	DG	H2'1	12	DG	H2'2	1.84
12	DG	H2'1	12	DG	H3'	2.13
12	DG	H4'	12	DG	H2'1	5.61
12	DG	H2'2	12	DG	H1'	2.24
12	DG	H2'2	12	DG	H3'	2.55
12	DG	H4'	12	DG	H2'2	5.01
12	DG	H1'	12	DG	H3'	4.30
12	DG	H5'1	12	DG	H3'	3.64
12	DG	H4'	12	DG	H5'1	2.68
12	DG	H5'2	12	DG	H3'	2.62
13	DA	H2'1	13	DA	H1'	2.26
13	DA	H2'1	13	DA	H2'2	1.85
13	DA	H2'1	13	DA	H3'	2.63
13	DA	H2'2	13	DA	H1'	2.99
13	DA	H2'2	13	DA	H3'	2.40
13	DA	H5'2	13	DA	H2'2	5.67
13	DA	H1'	13	DA	H3'	4.38
13	DA	H3'	13	DA	H4'	4.77
13	DA	H1'	13	DA	H5'1	3.72
13	DA	H5'1	13	DA	H3'	2.63
14	DG	H2'2	14	DG	H1'	2.31
14	DG	H2'2	14	DG	H3'	2.57
14	DG	H4'	14	DG	H2'2	6.50
14	DG	H5'1	14	DG	H2'2	5.33
14	DG	H3'	14	DG	H1'	4.75
14	DG	H4'	14	DG	H1'	3.58
14	DG	H4'	14	DG	H3'	2.70
14	DG	H5'1	14	DG	H1'	3.84
14	DG	H5'1	14	DG	H3'	3.36
14	DG	H5'2	14	DG	H3'	2.69
14	DG	H4'	14	DG	H5'2	2.51
15	DG	H2'1	15	DG	H1'	2.74
15	DG	H2'1	15	DG	H2'2	1.88
15	DG	H2'2	15	DG	H1'	2.33
15	DG	H2'2	15	DG	H3'	2.66
15	DG	H4'	15	DG	H2'2	5.71
15	DG	H4'	15	DG	H1'	3.04
16	DT	H2'1	16	DT	H1'	2.09
10	DA	H8	10	DA	H1'	8.02
11	DG	H8	10	DA	H1'	9.38
11	DG	H8	11	DG	H1'	5.77
11	DG	H1'	12	DG	H8	4.46

Identifier A	Residue A	Atom A	Identifier B	Residue B	Atom B	Distance
11	DG	H8	11	DG	H2'1	3.01
11	DG	H2'1	12	DG	H8	4.26
11	DG	H8	11	DG	H2'2	4.27
12	DG	H8	12	DG	H1'	5.96
12	DG	H2'1	13	DA	H8	3.51
12	DG	H2'2	13	DA	H8	3.01
13	DA	H2'1	13	DA	H8	3.91
13	DA	H2'2	13	DA	H8	2.68
13	DA	H3'	13	DA	H8	4.25
14	DG	H8	14	DG	H1'	3.93
15	DG	H8	14	DG	H1'	4.82
14	DG	H8	14	DG	H2'1	2.94
14	DG	H8	14	DG	H2'2	3.75
15	DG	H8	15	DG	H1'	5.47
15	DG	H2'1	15	DG	H8	2.75
16	DT	H6	15	DG	H2'1	7.20
15	DG	H8	15	DG	H2'2	3.55
16	DT	H6	16	DT	H1'	4.34
16	DT	H6	16	DT	H2'1	3.29
17	DT	H2'1	17	DT	H2'2	2.03
17	DT	H3'	17	DT	H2'1	2.77
17	DT	H2'2	17	DT	H3'	2.72
17	DT	H5'1	17	DT	H2'2	5.03
17	DT	H4'	17	DT	H3'	2.57
17	DT	H5'1	17	DT	H4'	2.83
17	DT	H5'1	17	DT	H3'	2.80
18	DA	H2'1	18	DA	H1'	2.58
18	DA	H2'1	18	DA	H3'	2.25
19	DG	H2'1	19	DG	H1'	3.21
19	DG	H2'1	19	DG	H2'2	1.82
19	DG	H4'	19	DG	H2'1	6.08
19	DG	H5'1	19	DG	H2'1	5.35
19	DG	H5'2	19	DG	H2'1	5.94
19	DG	H2'2	19	DG	H1'	2.47
19	DG	H2'2	19	DG	H3'	2.60
19	DG	H3'	19	DG	H5'2	2.70
20	DG	H2'1	20	DG	H1'	3.03
20	DG	H2'1	20	DG	H2'2	1.84
20	DG	H2'1	20	DG	H3'	2.13
20	DG	H4'	20	DG	H2'1	5.61
20	DG	H2'2	20	DG	H1'	2.24
20	DG	H2'2	20	DG	H3'	2.55
20	DG	H4'	20	DG	H2'2	5.01
20	DG	H1'	20	DG	H3'	4.30
20	DG	H5'1	20	DG	H3'	3.64
20	DG	H4'	20	DG	H5'1	2.68
20	DG	H5'2	20	DG	H3'	2.62

Identifier A	Residue A	Atom A	Identifier B	Residue B	Atom B	Distance
21	DA	H2'1	21	DA	H1'	2.26
21	DA	H2'1	21	DA	H2'2	1.85
21	DA	H2'1	21	DA	H3'	2.63
21	DA	H2'2	21	DA	H1'	2.99
21	DA	H2'2	21	DA	H3'	2.40
21	DA	H5'2	21	DA	H2'2	5.67
21	DA	H1'	21	DA	H3'	4.38
21	DA	H3'	21	DA	H4'	4.77
21	DA	H1'	21	DA	H5'1	3.72
21	DA	H5'1	21	DA	H3'	2.63
22	DG	H2'2	22	DG	H1'	2.31
22	DG	H2'2	22	DG	H3'	2.57
22	DG	H4'	22	DG	H2'2	6.50
22	DG	H5'1	22	DG	H2'2	5.33
22	DG	H3'	22	DG	H1'	4.75
22	DG	H4'	22	DG	H1'	3.58
22	DG	H4'	22	DG	H3'	2.70
22	DG	H5'1	22	DG	H1'	3.84
22	DG	H5'1	22	DG	H3'	3.36
22	DG	H5'2	22	DG	H3'	2.69
22	DG	H4'	22	DG	H5'2	2.51
23	DG	H2'1	23	DG	H1'	2.74
23	DG	H2'1	23	DG	H2'2	1.88
23	DG	H2'2	23	DG	H1'	2.33
23	DG	H2'2	23	DG	H3'	2.66
23	DG	H4'	23	DG	H2'2	5.71
23	DG	H4'	23	DG	H1'	3.04
24	DT	H2'1	24	DT	H1'	2.09
18	DA	H8	18	DA	H1'	8.02
19	DG	H8	18	DA	H1'	9.38
19	DG	H8	19	DG	H1'	5.77
19	DG	H1'	20	DG	H8	4.46
19	DG	H8	19	DG	H2'1	3.01
19	DG	H2'1	20	DG	H8	4.26
19	DG	H8	19	DG	H2'2	4.27
20	DG	H8	20	DG	H1'	5.96
20	DG	H2'1	21	DA	H8	3.51
20	DG	H2'2	21	DA	H8	3.01
21	DA	H2'1	21	DA	H8	3.91
21	DA	H2'2	21	DA	H8	2.68
21	DA	H3'	21	DA	H8	4.25
22	DG	H8	22	DG	H1'	3.93
23	DG	H8	22	DG	H1'	4.82
22	DG	H8	22	DG	H2'1	2.94
22	DG	H8	22	DG	H2'2	3.75
23	DG	H8	23	DG	H1'	5.47
23	DG	H2'1	23	DG	H8	2.75

Identifier A	Residue A	Atom A	Identifier B	Residue B	Atom B	Distance
24	DT	H6	23	DG	H2'1	7.20
23	DG	H8	23	DG	H2'2	3.55
24	DT	H6	24	DT	H1'	4.34
24	DT	H6	24	DT	H2'1	3.29
25	DT	H2'1	25	DT	H2'2	2.03
25	DT	H3'	25	DT	H2'1	2.77
25	DT	H2'2	25	DT	H3'	2.72
25	DT	H5'1	25	DT	H2'2	5.03
25	DT	H4'	25	DT	H3'	2.57
25	DT	H5'1	25	DT	H4'	2.83
25	DT	H5'1	25	DT	H3'	2.80
26	DA	H2'1	26	DA	H1'	2.58
26	DA	H2'1	26	DA	H3'	2.25
27	DG	H2'1	27	DG	H1'	3.21
27	DG	H2'1	27	DG	H2'2	1.82
27	DG	H4'	27	DG	H2'1	6.08
27	DG	H5'1	27	DG	H2'1	5.35
27	DG	H5'2	27	DG	H2'1	5.94
27	DG	H2'2	27	DG	H1'	2.47
27	DG	H2'2	27	DG	H3'	2.60
27	DG	H3'	27	DG	H5'2	2.70
28	DG	H2'1	28	DG	H1'	3.03
28	DG	H2'1	28	DG	H2'2	1.84
28	DG	H2'1	28	DG	H3'	2.13
28	DG	H4'	28	DG	H2'1	5.61
28	DG	H2'2	28	DG	H1'	2.24
28	DG	H2'2	28	DG	H3'	2.55
28	DG	H4'	28	DG	H2'2	5.01
28	DG	H1'	28	DG	H3'	4.30
28	DG	H5'1	28	DG	H3'	3.64
28	DG	H4'	28	DG	H5'1	2.68
28	DG	H5'2	28	DG	H3'	2.62
29	DA	H2'1	29	DA	H1'	2.26
29	DA	H2'1	29	DA	H2'2	1.85
29	DA	H2'1	29	DA	H3'	2.63
29	DA	H2'2	29	DA	H1'	2.99
29	DA	H2'2	29	DA	H3'	2.40
29	DA	H5'2	29	DA	H2'2	5.67
29	DA	H1'	29	DA	H3'	4.38
29	DA	H3'	29	DA	H4'	4.77
29	DA	H1'	29	DA	H5'1	3.72
29	DA	H5'1	29	DA	H3'	2.63
30	DG	H2'2	30	DG	H1'	2.31
30	DG	H2'2	30	DG	H3'	2.57
30	DG	H4'	30	DG	H2'2	6.50
30	DG	H5'1	30	DG	H2'2	5.33
30	DG	H3'	30	DG	H1'	4.75

Identifier A	Residue A	Atom A	Identifier B	Residue B	Atom B	Distance
30	DG	H4'	30	DG	H1'	3.58
30	DG	H4'	30	DG	H3'	2.70
30	DG	H5'1	30	DG	H1'	3.84
30	DG	H5'1	30	DG	H3'	3.36
30	DG	H5'2	30	DG	H3'	2.69
30	DG	H4'	30	DG	H5'2	2.51
31	DG	H2'1	31	DG	H1'	2.74
31	DG	H2'1	31	DG	H2'2	1.88
31	DG	H2'2	31	DG	H1'	2.33
31	DG	H2'2	31	DG	H3'	2.66
31	DG	H4'	31	DG	H2'2	5.71
31	DG	H4'	31	DG	H1'	3.04
32	DT	H2'1	32	DT	H1'	2.09
26	DA	H8	26	DA	H1'	8.02
27	DG	H8	26	DA	H1'	9.38
27	DG	H8	27	DG	H1'	5.77
27	DG	H1'	28	DG	H8	4.46
27	DG	H8	27	DG	H2'1	3.01
27	DG	H2'1	28	DG	H8	4.26
27	DG	H8	27	DG	H2'2	4.27
28	DG	H8	28	DG	H1'	5.96
28	DG	H2'1	29	DA	H8	3.51
28	DG	H2'2	29	DA	H8	3.01
29	DA	H2'1	29	DA	H8	3.91
29	DA	H2'2	29	DA	H8	2.68
29	DA	H3'	29	DA	H8	4.25
30	DG	H8	30	DG	H1'	3.93
31	DG	H8	30	DG	H1'	4.82
30	DG	H8	30	DG	H2'1	2.94
30	DG	H8	30	DG	H2'2	3.75
31	DG	H8	31	DG	H1'	5.47
31	DG	H2'1	31	DG	H8	2.75
32	DT	H6	31	DG	H2'1	7.20
31	DG	H8	31	DG	H2'2	3.55
32	DT	H6	32	DT	H1'	4.34
32	DT	H6	32	DT	H2'1	3.29

Appendix 2.6: d(TGGAGGC)$_4$

Identifier A	Residue A	Atom A	Identifier B	Residue B	Atom B	Distance
2	DG	H8	1	DT	H1'	3.95
1	DT	H2'1	1	DT	H1'	3.35
1	DT	H2'1	1	DT	H2'2	2.13
1	DT	H2'1	1	DT	H3'	3.50
1	DT	H4'	1	DT	H2'1	3.68
1	DT	H2'1	1	DT	H5'1	4.07
1	DT	H2'1	1	DT	H6	2.78
1	DT	H2'1	2	DG	H8	3.22
1	DT	H2'2	1	DT	H1'	2.57
1	DT	H2'2	1	DT	H3'	3.50
1	DT	H2'2	1	DT	H4'	2.89
1	DT	H2'2	1	DT	H5'1	4.60
1	DT	H2'2	1	DT	H6	3.65
2	DG	H8	1	DT	H2'2	2.60
1	DT	H3'	1	DT	H1'	4.22
1	DT	H3'	1	DT	H6	4.17
1	DT	H3'	2	DG	H8	4.71
1	DT	H1'	1	DT	H4'	2.92
1	DT	H4'	1	DT	H3'	2.92
1	DT	H4'	1	DT	H6	4.34
1	DT	H5'1	1	DT	H1'	4.33
1	DT	H5'1	1	DT	H3'	3.56
1	DT	H5'1	1	DT	H4'	2.50
1	DT	H5'1	1	DT	H6	3.56
1	DT	H5'1	2	DG	H8	4.65
1	DT	H6	2	DG	H8	8.13
2	DG	H2'1	2	DG	H2'2	2.00
2	DG	H3'	2	DG	H2'1	2.44
2	DG	H2'1	2	DG	H4'	4.28
2	DG	H2'1	2	DG	H5'1	4.07
2	DG	H8	2	DG	H2'1	2.51
2	DG	H2'1	3	DG	H8	3.25
2	DG	H2'2	2	DG	H1'	2.50
2	DG	H3'	2	DG	H2'2	2.81
2	DG	H2'2	2	DG	H4'	4.25
2	DG	H2'2	2	DG	H5'1	4.10
2	DG	H2'2	2	DG	H8	3.21
3	DG	H8	2	DG	H2'2	2.44
2	DG	H3'	3	DG	H8	2.83
2	DG	H3'	2	DG	H4'	2.63
2	DG	H4'	2	DG	H8	4.60
2	DG	H4'	3	DG	H8	4.37
2	DG	H5'1	2	DG	H4'	2.58
2	DG	H8	2	DG	H5'1	3.32
2	DG	H5'1	3	DG	H8	6.20

Identifier A	Residue A	Atom A	Identifier B	Residue B	Atom B	Distance
4	DA	H8	3	DG	H1'	3.37
3	DG	H1'	3	DG	H2'1	2.71
3	DG	H2'1	3	DG	H2'2	1.97
3	DG	H3'	3	DG	H2'1	2.39
3	DG	H2'1	3	DG	H4'	4.00
3	DG	H2'1	3	DG	H5'1	4.64
3	DG	H2'1	3	DG	H5'2	4.44
3	DG	H2'1	3	DG	H8	2.40
3	DG	H2'1	4	DA	H8	2.94
3	DG	H2'2	3	DG	H1'	2.38
3	DG	H3'	3	DG	H2'2	2.43
3	DG	H2'2	3	DG	H4'	3.44
3	DG	H2'2	3	DG	H5'1	4.15
3	DG	H2'2	3	DG	H5'2	3.94
3	DG	H8	3	DG	H2'2	2.68
3	DG	H2'2	4	DA	H8	2.54
3	DG	H3'	4	DA	H8	4.20
3	DG	H4'	3	DG	H1'	2.60
3	DG	H4'	4	DA	H8	4.68
3	DG	H5'1	3	DG	H8	4.34
3	DG	H1'	3	DG	H5'2	3.06
3	DG	H5'2	3	DG	H8	4.41
3	DG	H8	2	DG	H8	6.05
3	DG	H8	4	DA	H8	3.82
4	DA	H8	4	DA	H1'	3.44
4	DA	H2'1	4	DA	H2'2	1.99
4	DA	H2'1	4	DA	H3'	4.05
4	DA	H2'1	4	DA	H4'	3.63
4	DA	H5'1	4	DA	H2'1	3.00
4	DA	H2'1	4	DA	H8	3.24
5	DG	H8	4	DA	H2'1	2.45
4	DA	H2'2	4	DA	H1'	2.76
4	DA	H2'2	4	DA	H3'	3.09
4	DA	H2'2	4	DA	H4'	4.14
4	DA	H2'2	4	DA	H8	2.57
4	DA	H3'	4	DA	H1'	3.97
4	DA	H3'	4	DA	H8	3.02
4	DA	H4'	4	DA	H1'	2.85
4	DA	H3'	4	DA	H4'	2.79
4	DA	H4'	4	DA	H8	4.49
5	DG	H8	4	DA	H4'	5.05
4	DA	H5'1	4	DA	H8	4.61
5	DG	H8	5	DG	H1'	3.52
6	DG	H8	5	DG	H1'	2.70
5	DG	H2'1	5	DG	H1'	3.87
5	DG	H2'1	5	DG	H2'2	3.00
5	DG	H2'1	6	DG	H8	2.38

Identifier A	Residue A	Atom A	Identifier B	Residue B	Atom B	Distance
5	DG	H2'2	5	DG	H1'	2.36
5	DG	H2'2	5	DG	H3'	3.28
5	DG	H2'2	5	DG	H4'	4.00
5	DG	H2'2	5	DG	H5'1	3.76
5	DG	H2'2	6	DG	H8	2.61
5	DG	H3'	5	DG	H1'	3.94
5	DG	H3'	5	DG	H8	3.38
5	DG	H3'	6	DG	H8	4.14
5	DG	H4'	5	DG	H1'	3.31
5	DG	H4'	5	DG	H8	4.25
5	DG	H4'	6	DG	H8	8.07
5	DG	H8	5	DG	H5'1	4.11
5	DG	H8	5	DG	H5'2	4.05
5	DG	H1'	5	DG	H5'2	4.21
5	DG	H5'2	5	DG	H4'	3.93
6	DG	H8	6	DG	H5'2	3.64
6	DG	H8	6	DG	H1'	3.39
6	DG	H1'	7	DC	H6	3.74
5	DG	H2'1	5	DG	H5'2	3.78
6	DG	H2'1	6	DG	H1'	2.86
6	DG	H2'1	6	DG	H2'2	2.10
6	DG	H3'	6	DG	H2'1	2.44
6	DG	H2'1	6	DG	H4'	3.64
6	DG	H2'1	6	DG	H5'1	4.20
6	DG	H2'1	6	DG	H5'2	4.01
6	DG	H2'1	7	DC	H5	4.26
6	DG	H2'1	7	DC	H6	3.26
6	DG	H2'2	6	DG	H1'	2.66
6	DG	H3'	6	DG	H2'2	2.82
6	DG	H2'2	6	DG	H4'	3.41
6	DG	H2'2	6	DG	H5'1	4.00
6	DG	H2'2	6	DG	H5'2	3.30
6	DG	H2'2	7	DC	H5	4.04
6	DG	H2'2	7	DC	H6	2.93
6	DG	H3'	6	DG	H1'	2.84
6	DG	H3'	6	DG	H8	3.33
6	DG	H4'	6	DG	H1'	3.34
6	DG	H4'	6	DG	H8	4.75
6	DG	H5'1	6	DG	H1'	3.81
6	DG	H5'2	6	DG	H5'1	1.97
5	DG	H8	6	DG	H8	3.95
7	DC	H1'	7	DC	H6	3.58
7	DC	H1'	7	DC	H2'1	2.54
7	DC	H4'	7	DC	H2'1	2.64
7	DC	H5	7	DC	H2'1	5.27
7	DC	H5'2	7	DC	H2'1	5.73
7	DC	H6	7	DC	H2'1	2.47

Identifier A	Residue A	Atom A	Identifier B	Residue B	Atom B	Distance
7	DC	H1'	7	DC	H2'2	2.51
7	DC	H4'	7	DC	H2'2	2.88
7	DC	H6	7	DC	H2'2	3.08
7	DC	H4'	7	DC	H1'	4.11
7	DC	H4'	7	DC	H6	3.26
7	DC	H5'1	7	DC	H1'	3.14
7	DC	H4'	7	DC	H5'1	3.09
7	DC	H5'2	7	DC	H5'1	2.33
7	DC	H6	7	DC	H5'1	4.38
7	DC	H5'2	7	DC	H1'	5.20
7	DC	H4'	7	DC	H5'2	2.42
7	DC	H6	7	DC	H5'2	4.42
7	DC	H5	6	DG	H1'	4.73
7	DC	H5	6	DG	H8	4.71
7	DC	H5	7	DC	H6	2.69
6	DG	H8	7	DC	H6	6.33
9	DG	H8	8	DT	H1'	3.95
8	DT	H2'1	8	DT	H1'	3.35
8	DT	H2'1	8	DT	H2'2	2.13
8	DT	H2'1	8	DT	H3'	3.50
8	DT	H4'	8	DT	H2'1	3.68
8	DT	H2'1	8	DT	H5'1	4.07
8	DT	H2'1	8	DT	H6	2.78
8	DT	H2'1	9	DG	H8	3.22
8	DT	H2'2	8	DT	H1'	2.57
8	DT	H2'2	8	DT	H3'	3.50
8	DT	H2'2	8	DT	H4'	2.89
8	DT	H2'2	8	DT	H5'1	4.60
8	DT	H2'2	8	DT	H6	3.65
9	DG	H8	8	DT	H2'2	2.60
8	DT	H3'	8	DT	H1'	4.22
8	DT	H3'	8	DT	H6	4.17
8	DT	H3'	9	DG	H8	4.71
8	DT	H1'	8	DT	H4'	2.92
8	DT	H4'	8	DT	H3'	2.92
8	DT	H4'	8	DT	H6	4.34
8	DT	H5'1	8	DT	H1'	4.33
8	DT	H5'1	8	DT	H3'	3.56
8	DT	H5'1	8	DT	H4'	2.50
8	DT	H5'1	8	DT	H6	3.56
8	DT	H5'1	9	DG	H8	4.65
8	DT	H6	9	DG	H8	8.13
9	DG	H2'1	9	DG	H2'2	2.00
9	DG	H3'	9	DG	H2'1	2.44
9	DG	H2'1	9	DG	H4'	4.28
9	DG	H2'1	9	DG	H5'1	4.07
9	DG	H8	9	DG	H2'1	2.51

Identifier A	Residue A	Atom A	Identifier B	Residue B	Atom B	Distance
9	DG	H2'1	10	DG	H8	3.25
9	DG	H2'2	9	DG	H1'	2.50
9	DG	H3'	9	DG	H2'2	2.81
9	DG	H2'2	9	DG	H4'	4.25
9	DG	H2'2	9	DG	H5'1	4.10
9	DG	H2'2	9	DG	H8	3.21
10	DG	H8	9	DG	H2'2	2.44
9	DG	H3'	10	DG	H8	2.83
9	DG	H3'	9	DG	H4'	2.63
9	DG	H4'	9	DG	H8	4.60
9	DG	H4'	10	DG	H8	4.37
9	DG	H5'1	9	DG	H4'	2.58
9	DG	H8	9	DG	H5'1	3.32
9	DG	H5'1	10	DG	H8	6.20
11	DA	H8	10	DG	H1'	3.37
10	DG	H1'	10	DG	H2'1	2.71
10	DG	H2'1	10	DG	H2'2	1.97
10	DG	H3'	10	DG	H2'1	2.39
10	DG	H2'1	10	DG	H4'	4.00
10	DG	H2'1	10	DG	H5'1	4.64
10	DG	H2'1	10	DG	H5'2	4.44
10	DG	H2'1	10	DG	H8	2.40
10	DG	H2'1	11	DA	H8	2.94
10	DG	H2'2	10	DG	H1'	2.38
10	DG	H3'	10	DG	H2'2	2.43
10	DG	H2'2	10	DG	H4'	3.44
10	DG	H2'2	10	DG	H5'1	4.15
10	DG	H2'2	10	DG	H5'2	3.94
10	DG	H8	10	DG	H2'2	2.68
10	DG	H2'2	11	DA	H8	2.54
10	DG	H3'	11	DA	H8	4.20
10	DG	H4'	10	DG	H1'	2.60
10	DG	H4'	11	DA	H8	4.68
10	DG	H5'1	10	DG	H8	4.34
10	DG	H1'	10	DG	H5'2	3.06
10	DG	H5'2	10	DG	H8	4.41
10	DG	H8	9	DG	H8	6.05
10	DG	H8	11	DA	H8	3.82
11	DA	H8	11	DA	H1'	3.44
11	DA	H2'1	11	DA	H2'2	1.99
11	DA	H2'1	11	DA	H3'	4.05
11	DA	H2'1	11	DA	H4'	3.63
11	DA	H5'1	11	DA	H2'1	3.00
11	DA	H2'1	11	DA	H8	3.24
12	DG	H8	11	DA	H2'1	2.45
11	DA	H2'2	11	DA	H1'	2.76
11	DA	H2'2	11	DA	H3'	3.09

Identifier A	Residue A	Atom A	Identifier B	Residue B	Atom B	Distance
11	DA	H2'2	11	DA	H4'	4.14
11	DA	H2'2	11	DA	H8	2.57
11	DA	H3'	11	DA	H1'	3.97
11	DA	H3'	11	DA	H8	3.02
11	DA	H4'	11	DA	H1'	2.85
11	DA	H3'	11	DA	H4'	2.79
11	DA	H4'	11	DA	H8	4.49
12	DG	H8	11	DA	H4'	5.05
11	DA	H5'1	11	DA	H8	4.61
12	DG	H8	12	DG	H1'	3.52
13	DG	H8	12	DG	H1'	2.70
12	DG	H2'1	12	DG	H1'	3.87
12	DG	H2'1	12	DG	H2'2	3.00
12	DG	H2'1	13	DG	H8	2.38
12	DG	H2'2	12	DG	H1'	2.36
12	DG	H2'2	12	DG	H3'	3.28
12	DG	H2'2	12	DG	H4'	4.00
12	DG	H2'2	12	DG	H5'1	3.76
12	DG	H2'2	13	DG	H8	2.61
12	DG	H3'	12	DG	H1'	3.94
12	DG	H3'	12	DG	H8	3.38
12	DG	H3'	13	DG	H8	4.14
12	DG	H4'	12	DG	H1'	3.31
12	DG	H4'	12	DG	H8	4.25
12	DG	H4'	13	DG	H8	8.07
12	DG	H8	12	DG	H5'1	4.11
12	DG	H8	12	DG	H5'2	4.05
12	DG	H1'	12	DG	H5'2	4.21
12	DG	H5'2	12	DG	H4'	3.93
13	DG	H8	13	DG	H5'2	3.64
13	DG	H8	13	DG	H1'	3.39
13	DG	H1'	14	DC	H6	3.74
12	DG	H2'1	12	DG	H5'2	3.78
13	DG	H2'1	13	DG	H1'	2.86
13	DG	H2'1	13	DG	H2'2	2.10
13	DG	H3'	13	DG	H2'1	2.44
13	DG	H2'1	13	DG	H4'	3.64
13	DG	H2'1	13	DG	H5'1	4.20
13	DG	H2'1	13	DG	H5'2	4.01
13	DG	H2'1	14	DC	H5	4.26
13	DG	H2'1	14	DC	H6	3.26
13	DG	H2'2	13	DG	H1'	2.66
13	DG	H3'	13	DG	H2'2	2.82
13	DG	H2'2	13	DG	H4'	3.41
13	DG	H2'2	13	DG	H5'1	4.00
13	DG	H2'2	13	DG	H5'2	3.30
13	DG	H2'2	14	DC	H5	4.04

Identifier A	Residue A	Atom A	Identifier B	Residue B	Atom B	Distance
13	DG	H2'2	14	DC	H6	2.93
13	DG	H3'	13	DG	H1'	2.84
13	DG	H3'	13	DG	H8	3.33
13	DG	H4'	13	DG	H1'	3.34
13	DG	H4'	13	DG	H8	4.75
13	DG	H5'1	13	DG	H1'	3.81
13	DG	H5'2	13	DG	H5'1	1.97
12	DG	H8	13	DG	H8	3.95
14	DC	H1'	14	DC	H6	3.58
14	DC	H1'	14	DC	H2'1	2.54
14	DC	H4'	14	DC	H2'1	2.64
14	DC	H5	14	DC	H2'1	5.27
14	DC	H5'2	14	DC	H2'1	5.73
14	DC	H6	14	DC	H2'1	2.47
14	DC	H1'	14	DC	H2'2	2.51
14	DC	H4'	14	DC	H2'2	2.88
14	DC	H6	14	DC	H2'2	3.08
14	DC	H4'	14	DC	H1'	4.11
14	DC	H4'	14	DC	H6	3.26
14	DC	H5'1	14	DC	H1'	3.14
14	DC	H4'	14	DC	H5'1	3.09
14	DC	H5'2	14	DC	H5'1	2.33
14	DC	H6	14	DC	H5'1	4.38
14	DC	H5'2	14	DC	H1'	5.20
14	DC	H4'	14	DC	H5'2	2.42
14	DC	H6	14	DC	H5'2	4.42
14	DC	H5	13	DG	H1'	4.73
14	DC	H5	13	DG	H8	4.71
14	DC	H5	14	DC	H6	2.69
13	DG	H8	14	DC	H6	6.33
16	DG	H8	15	DT	H1'	3.95
15	DT	H2'1	15	DT	H1'	3.35
15	DT	H2'1	15	DT	H2'2	2.13
15	DT	H2'1	15	DT	H3'	3.50
15	DT	H4'	15	DT	H2'1	3.68
15	DT	H2'1	15	DT	H5'1	4.07
15	DT	H2'1	15	DT	H6	2.78
15	DT	H2'1	16	DG	H8	3.22
15	DT	H2'2	15	DT	H1'	2.57
15	DT	H2'2	15	DT	H3'	3.50
15	DT	H2'2	15	DT	H4'	2.89
15	DT	H2'2	15	DT	H5'1	4.60
15	DT	H2'2	15	DT	H6	3.65
16	DG	H8	15	DT	H2'2	2.60
15	DT	H3'	15	DT	H1'	4.22
15	DT	H3'	15	DT	H6	4.17
15	DT	H3'	16	DG	H8	4.71

Identifier A	Residue A	Atom A	Identifier B	Residue B	Atom B	Distance
15	DT	H1'	15	DT	H4'	2.92
15	DT	H4'	15	DT	H3'	2.92
15	DT	H4'	15	DT	H6	4.34
15	DT	H5'1	15	DT	H1'	4.33
15	DT	H5'1	15	DT	H3'	3.56
15	DT	H5'1	15	DT	H4'	2.50
15	DT	H5'1	15	DT	H6	3.56
15	DT	H5'1	16	DG	H8	4.65
15	DT	H6	16	DG	H8	8.13
16	DG	H2'1	16	DG	H2'2	2.00
16	DG	H3'	16	DG	H2'1	2.44
16	DG	H2'1	16	DG	H4'	4.28
16	DG	H2'1	16	DG	H5'1	4.07
16	DG	H8	16	DG	H2'1	2.51
16	DG	H2'1	17	DG	H8	3.25
16	DG	H2'2	16	DG	H1'	2.50
16	DG	H3'	16	DG	H2'2	2.81
16	DG	H2'2	16	DG	H4'	4.25
16	DG	H2'2	16	DG	H5'1	4.10
16	DG	H2'2	16	DG	H8	3.21
17	DG	H8	16	DG	H2'2	2.44
16	DG	H3'	17	DG	H8	2.83
16	DG	H3'	16	DG	H4'	2.63
16	DG	H4'	16	DG	H8	4.60
16	DG	H4'	17	DG	H8	4.37
16	DG	H5'1	16	DG	H4'	2.58
16	DG	H8	16	DG	H5'1	3.32
16	DG	H5'1	17	DG	H8	6.20
18	DA	H8	17	DG	H1'	3.37
17	DG	H1'	17	DG	H2'1	2.71
17	DG	H2'1	17	DG	H2'2	1.97
17	DG	H3'	17	DG	H2'1	2.39
17	DG	H2'1	17	DG	H4'	4.00
17	DG	H2'1	17	DG	H5'1	4.64
17	DG	H2'1	17	DG	H5'2	4.44
17	DG	H2'1	17	DG	H8	2.40
17	DG	H2'1	18	DA	H8	2.94
17	DG	H2'2	17	DG	H1'	2.38
17	DG	H3'	17	DG	H2'2	2.43
17	DG	H2'2	17	DG	H4'	3.44
17	DG	H2'2	17	DG	H5'1	4.15
17	DG	H2'2	17	DG	H5'2	3.94
17	DG	H8	17	DG	H2'2	2.68
17	DG	H2'2	18	DA	H8	2.54
17	DG	H3'	18	DA	H8	4.20
17	DG	H4'	17	DG	H1'	2.60
17	DG	H4'	18	DA	H8	4.68

Identifier A	Residue A	Atom A	Identifier B	Residue B	Atom B	Distance
17	DG	H5'1	17	DG	H8	4.34
17	DG	H1'	17	DG	H5'2	3.06
17	DG	H5'2	17	DG	H8	4.41
17	DG	H8	16	DG	H8	6.05
17	DG	H8	18	DA	H8	3.82
18	DA	H8	18	DA	H1'	3.44
18	DA	H2'1	18	DA	H2'2	1.99
18	DA	H2'1	18	DA	H3'	4.05
18	DA	H2'1	18	DA	H4'	3.63
18	DA	H5'1	18	DA	H2'1	3.00
18	DA	H2'1	18	DA	H8	3.24
19	DG	H8	18	DA	H2'1	2.45
18	DA	H2'2	18	DA	H1'	2.76
18	DA	H2'2	18	DA	H3'	3.09
18	DA	H2'2	18	DA	H4'	4.14
18	DA	H2'2	18	DA	H8	2.57
18	DA	H3'	18	DA	H1'	3.97
18	DA	H3'	18	DA	H8	3.02
18	DA	H4'	18	DA	H1'	2.85
18	DA	H3'	18	DA	H4'	2.79
18	DA	H4'	18	DA	H8	4.49
19	DG	H8	18	DA	H4'	5.05
18	DA	H5'1	18	DA	H8	4.61
19	DG	H8	19	DG	H1'	3.52
20	DG	H8	19	DG	H1'	2.70
19	DG	H2'1	19	DG	H1'	3.87
19	DG	H2'1	19	DG	H2'2	3.00
19	DG	H2'1	20	DG	H8	2.38
19	DG	H2'2	19	DG	H1'	2.36
19	DG	H2'2	19	DG	H3'	3.28
19	DG	H2'2	19	DG	H4'	4.00
19	DG	H2'2	19	DG	H5'1	3.76
19	DG	H2'2	20	DG	H8	2.61
19	DG	H3'	19	DG	H1'	3.94
19	DG	H3'	19	DG	H8	3.38
19	DG	H3'	20	DG	H8	4.14
19	DG	H4'	19	DG	H1'	3.31
19	DG	H4'	19	DG	H8	4.25
19	DG	H4'	20	DG	H8	8.07
19	DG	H8	19	DG	H5'1	4.11
19	DG	H8	19	DG	H5'2	4.05
19	DG	H1'	19	DG	H5'2	4.21
19	DG	H5'2	19	DG	H4'	3.93
20	DG	H8	20	DG	H5'2	3.64
20	DG	H8	20	DG	H1'	3.39
20	DG	H1'	21	DC	H6	3.74
19	DG	H2'1	19	DG	H5'2	3.78

Identifier A	Residue A	Atom A	Identifier B	Residue B	Atom B	Distance
20	DG	H2'1	20	DG	H1'	2.86
20	DG	H2'1	20	DG	H2'2	2.10
20	DG	H3'	20	DG	H2'1	2.44
20	DG	H2'1	20	DG	H4'	3.64
20	DG	H2'1	20	DG	H5'1	4.20
20	DG	H2'1	20	DG	H5'2	4.01
20	DG	H2'1	21	DC	H5	4.26
20	DG	H2'1	21	DC	H6	3.26
20	DG	H2'2	20	DG	H1'	2.66
20	DG	H3'	20	DG	H2'2	2.82
20	DG	H2'2	20	DG	H4'	3.41
20	DG	H2'2	20	DG	H5'1	4.00
20	DG	H2'2	20	DG	H5'2	3.30
20	DG	H2'2	21	DC	H5	4.04
20	DG	H2'2	21	DC	H6	2.93
20	DG	H3'	20	DG	H1'	2.84
20	DG	H3'	20	DG	H8	3.33
20	DG	H4'	20	DG	H1'	3.34
20	DG	H4'	20	DG	H8	4.75
20	DG	H5'1	20	DG	H1'	3.81
20	DG	H5'2	20	DG	H5'1	1.97
19	DG	H8	20	DG	H8	3.95
21	DC	H1'	21	DC	H6	3.58
21	DC	H1'	21	DC	H2'1	2.54
21	DC	H4'	21	DC	H2'1	2.64
21	DC	H5	21	DC	H2'1	5.27
21	DC	H5'2	21	DC	H2'1	5.73
21	DC	H6	21	DC	H2'1	2.47
21	DC	H1'	21	DC	H2'2	2.51
21	DC	H4'	21	DC	H2'2	2.88
21	DC	H6	21	DC	H2'2	3.08
21	DC	H4'	21	DC	H1'	4.11
21	DC	H4'	21	DC	H6	3.26
21	DC	H5'1	21	DC	H1'	3.14
21	DC	H4'	21	DC	H5'1	3.09
21	DC	H5'2	21	DC	H5'1	2.33
21	DC	H6	21	DC	H5'1	4.38
21	DC	H5'2	21	DC	H1'	5.20
21	DC	H4'	21	DC	H5'2	2.42
21	DC	H6	21	DC	H5'2	4.42
21	DC	H5	20	DG	H1'	4.73
21	DC	H5	20	DG	H8	4.71
21	DC	H5	21	DC	H6	2.69
20	DG	H8	21	DC	H6	6.33
23	DG	H8	22	DT	H1'	3.95
22	DT	H2'1	22	DT	H1'	3.35
22	DT	H2'1	22	DT	H2'2	2.13

Identifier A	Residue A	Atom A	Identifier B	Residue B	Atom B	Distance
22	DT	H2'1	22	DT	H3'	3.50
22	DT	H4'	22	DT	H2'1	3.68
22	DT	H2'1	22	DT	H5'1	4.07
22	DT	H2'1	22	DT	H6	2.78
22	DT	H2'1	23	DG	H8	3.22
22	DT	H2'2	22	DT	H1'	2.57
22	DT	H2'2	22	DT	H3'	3.50
22	DT	H2'2	22	DT	H4'	2.89
22	DT	H2'2	22	DT	H5'1	4.60
22	DT	H2'2	22	DT	H6	3.65
23	DG	H8	22	DT	H2'2	2.60
22	DT	H3'	22	DT	H1'	4.22
22	DT	H3'	22	DT	H6	4.17
22	DT	H3'	23	DG	H8	4.71
22	DT	H1'	22	DT	H4'	2.92
22	DT	H4'	22	DT	H3'	2.92
22	DT	H4'	22	DT	H6	4.34
22	DT	H5'1	22	DT	H1'	4.33
22	DT	H5'1	22	DT	H3'	3.56
22	DT	H5'1	22	DT	H4'	2.50
22	DT	H5'1	22	DT	H6	3.56
22	DT	H5'1	23	DG	H8	4.65
22	DT	H6	23	DG	H8	8.13
23	DG	H2'1	23	DG	H2'2	2.00
23	DG	H3'	23	DG	H2'1	2.44
23	DG	H2'1	23	DG	H4'	4.28
23	DG	H2'1	23	DG	H5'1	4.07
23	DG	H8	23	DG	H2'1	2.51
23	DG	H2'1	24	DG	H8	3.25
23	DG	H2'2	23	DG	H1'	2.50
23	DG	H3'	23	DG	H2'2	2.81
23	DG	H2'2	23	DG	H4'	4.25
23	DG	H2'2	23	DG	H5'1	4.10
23	DG	H2'2	23	DG	H8	3.21
24	DG	H8	23	DG	H2'2	2.44
23	DG	H3'	24	DG	H8	2.83
23	DG	H3'	23	DG	H4'	2.63
23	DG	H4'	23	DG	H8	4.60
23	DG	H4'	24	DG	H8	4.37
23	DG	H5'1	23	DG	H4'	2.58
23	DG	H8	23	DG	H5'1	3.32
23	DG	H5'1	24	DG	H8	6.20
25	DA	H8	24	DG	H1'	3.37
24	DG	H1'	24	DG	H2'1	2.71
24	DG	H2'1	24	DG	H2'2	1.97
24	DG	H3'	24	DG	H2'1	2.39
24	DG	H2'1	24	DG	H4'	4.00

Identifier A	Residue A	Atom A	Identifier B	Residue B	Atom B	Distance
24	DG	H2'1	24	DG	H5'1	4.64
24	DG	H2'1	24	DG	H5'2	4.44
24	DG	H2'1	24	DG	H8	2.40
24	DG	H2'1	25	DA	H8	2.94
24	DG	H2'2	24	DG	H1'	2.38
24	DG	H3'	24	DG	H2'2	2.43
24	DG	H2'2	24	DG	H4'	3.44
24	DG	H2'2	24	DG	H5'1	4.15
24	DG	H2'2	24	DG	H5'2	3.94
24	DG	H8	24	DG	H2'2	2.68
24	DG	H2'2	25	DA	H8	2.54
24	DG	H3'	25	DA	H8	4.20
24	DG	H4'	24	DG	H1'	2.60
24	DG	H4'	25	DA	H8	4.68
24	DG	H5'1	24	DG	H8	4.34
24	DG	H1'	24	DG	H5'2	3.06
24	DG	H5'2	24	DG	H8	4.41
24	DG	H8	23	DG	H8	6.05
24	DG	H8	25	DA	H8	3.82
25	DA	H8	25	DA	H1'	3.44
25	DA	H2'1	25	DA	H2'2	1.99
25	DA	H2'1	25	DA	H3'	4.05
25	DA	H2'1	25	DA	H4'	3.63
25	DA	H5'1	25	DA	H2'1	3.00
25	DA	H2'1	25	DA	H8	3.24
26	DG	H8	25	DA	H2'1	2.45
25	DA	H2'2	25	DA	H1'	2.76
25	DA	H2'2	25	DA	H3'	3.09
25	DA	H2'2	25	DA	H4'	4.14
25	DA	H2'2	25	DA	H8	2.57
25	DA	H3'	25	DA	H1'	3.97
25	DA	H3'	25	DA	H8	3.02
25	DA	H4'	25	DA	H1'	2.85
25	DA	H3'	25	DA	H4'	2.79
25	DA	H4'	25	DA	H8	4.49
26	DG	H8	25	DA	H4'	5.05
25	DA	H5'1	25	DA	H8	4.61
26	DG	H8	26	DG	H1'	3.52
27	DG	H8	26	DG	H1'	2.70
26	DG	H2'1	26	DG	H1'	3.87
26	DG	H2'1	26	DG	H2'2	3.00
26	DG	H2'1	27	DG	H8	2.38
26	DG	H2'2	26	DG	H1'	2.36
26	DG	H2'2	26	DG	H3'	3.28
26	DG	H2'2	26	DG	H4'	4.00
26	DG	H2'2	26	DG	H5'1	3.76
26	DG	H2'2	27	DG	H8	2.61

Identifier A	Residue A	Atom A	Identifier B	Residue B	Atom B	Distance
26	DG	H3'	26	DG	H1'	3.94
26	DG	H3'	26	DG	H8	3.38
26	DG	H3'	27	DG	H8	4.14
26	DG	H4'	26	DG	H1'	3.31
26	DG	H4'	26	DG	H8	4.25
26	DG	H4'	27	DG	H8	8.07
26	DG	H8	26	DG	H5'1	4.11
26	DG	H8	26	DG	H5'2	4.05
26	DG	H1'	26	DG	H5'2	4.21
26	DG	H5'2	26	DG	H4'	3.93
27	DG	H8	27	DG	H5'2	3.64
27	DG	H8	27	DG	H1'	3.39
27	DG	H1'	28	DC	H6	3.74
26	DG	H2'1	26	DG	H5'2	3.78
27	DG	H2'1	27	DG	H1'	2.86
27	DG	H2'1	27	DG	H2'2	2.10
27	DG	H3'	27	DG	H2'1	2.44
27	DG	H2'1	27	DG	H4'	3.64
27	DG	H2'1	27	DG	H5'1	4.20
27	DG	H2'1	27	DG	H5'2	4.01
27	DG	H2'1	28	DC	H5	4.26
27	DG	H2'1	28	DC	H6	3.26
27	DG	H2'2	27	DG	H1'	2.66
27	DG	H3'	27	DG	H2'2	2.82
27	DG	H2'2	27	DG	H4'	3.41
27	DG	H2'2	27	DG	H5'1	4.00
27	DG	H2'2	27	DG	H5'2	3.30
27	DG	H2'2	28	DC	H5	4.04
27	DG	H2'2	28	DC	H6	2.93
27	DG	H3'	27	DG	H1'	2.84
27	DG	H3'	27	DG	H8	3.33
27	DG	H4'	27	DG	H1'	3.34
27	DG	H4'	27	DG	H8	4.75
27	DG	H5'1	27	DG	H1'	3.81
27	DG	H5'2	27	DG	H5'1	1.97
26	DG	H8	27	DG	H8	3.95
28	DC	H1'	28	DC	H6	3.58
28	DC	H1'	28	DC	H2'1	2.54
28	DC	H4'	28	DC	H2'1	2.64
28	DC	H5	28	DC	H2'1	5.27
28	DC	H5'2	28	DC	H2'1	5.73
28	DC	H6	28	DC	H2'1	2.47
28	DC	H1'	28	DC	H2'2	2.51
28	DC	H4'	28	DC	H2'2	2.88
28	DC	H6	28	DC	H2'2	3.08
28	DC	H4'	28	DC	H1'	4.11
28	DC	H4'	28	DC	H6	3.26

Identifier A	Residue A	Atom A	Identifier B	Residue B	Atom B	Distance
28	DC	H5'1	28	DC	H1'	3.14
28	DC	H4'	28	DC	H5'1	3.09
28	DC	H5'2	28	DC	H5'1	2.33
28	DC	H6	28	DC	H5'1	4.38
28	DC	H5'2	28	DC	H1'	5.20
28	DC	H4'	28	DC	H5'2	2.42
28	DC	H6	28	DC	H5'2	4.42
28	DC	H5	27	DG	H1'	4.73
28	DC	H5	27	DG	H8	4.71
28	DC	H5	28	DC	H6	2.69
27	DG	H8	28	DC	H6	6.33

Appendix 2.7: d(TGGGAGGC)$_4$

Identifier A	Residue A	Atom A	Identifier B	Residue B	Atom B	Distance
1	DT	H1'	1	DT	H2'2	2.62
1	DT	H1'	1	DT	H4'	3.18
1	DT	H2'1	1	DT	H2'2	2.01
1	DT	H2'2	1	DT	H3'	2.60
1	DT	H2'2	2	DG	H8	1.87
1	DT	H4'	1	DT	H3'	2.78
2	DG	H2'1	2	DG	H2'2	1.80
2	DG	H2'1	2	DG	H3'	2.38
2	DG	H2'1	2	DG	H8	1.84
2	DG	H2'2	2	DG	H1'	2.38
2	DG	H2'2	2	DG	H3'	2.58
2	DG	H2'2	2	DG	H8	2.10
2	DG	H2'2	3	DG	H8	1.88
2	DG	H3'	2	DG	H4'	2.81
2	DG	H4'	2	DG	H1'	3.63
2	DG	H4'	2	DG	H5'1	2.85
2	DG	H4'	2	DG	H5'2	2.33
2	DG	H5'1	2	DG	H3'	2.86
2	DG	H5'2	2	DG	H3'	3.02
3	DG	H2'1	3	DG	H1'	3.25
3	DG	H2'1	3	DG	H2'2	1.85
3	DG	H2'1	3	DG	H3'	2.36
3	DG	H2'1	3	DG	H8	1.84
3	DG	H2'1	4	DG	H8	2.32
3	DG	H2'2	3	DG	H1'	2.35
3	DG	H2'2	3	DG	H3'	2.57
3	DG	H2'2	3	DG	H8	2.10
3	DG	H2'2	4	DG	H8	1.80
3	DG	H3'	3	DG	H5'2	3.24
3	DG	H4'	3	DG	H1'	3.36
3	DG	H4'	3	DG	H3'	2.93
3	DG	H5'1	3	DG	H1'	2.69
3	DG	H5'1	3	DG	H3'	2.67
3	DG	H5'1	3	DG	H4'	2.26
3	DG	H5'2	3	DG	H4'	2.54
3	DG	H8	3	DG	H3'	3.40
4	DG	H2'1	4	DG	H1'	2.84
4	DG	H2'1	4	DG	H2'2	1.87
4	DG	H2'1	4	DG	H3'	2.23
4	DG	H2'1	4	DG	H8	1.90
4	DG	H2'1	5	DA	H8	1.83
4	DG	H2'2	4	DG	H1'	2.26
4	DG	H2'2	4	DG	H3'	2.41
4	DG	H2'2	4	DG	H3'	2.51
4	DG	H2'2	4	DG	H8	1.95

Identifier A	Residue A	Atom A	Identifier B	Residue B	Atom B	Distance
4	DG	H2'2	5	DA	H8	1.89
4	DG	H4'	4	DG	H1'	3.39
4	DG	H4'	4	DG	H3'	2.63
4	DG	H5'1	4	DG	H1'	4.40
4	DG	H5'1	4	DG	H4'	1.89
4	DG	H5'2	4	DG	H1'	2.42
4	DG	H5'2	4	DG	H3'	2.64
4	DG	H8	3	DG	H1'	3.00
4	DG	H8	4	DG	H3'	2.93
5	DA	H1'	5	DA	H3'	4.13
5	DA	H2'1	5	DA	H1'	2.31
5	DA	H2'1	5	DA	H2'2	1.89
5	DA	H2'1	5	DA	H3'	2.80
5	DA	H2'1	5	DA	H8	2.28
5	DA	H2'1	6	DG	H8	1.81
5	DA	H2'2	5	DA	H1'	2.82
5	DA	H2'2	5	DA	H3'	2.25
5	DA	H2'2	5	DA	H8	1.82
5	DA	H3'	5	DA	H8	2.01
5	DA	H5'1	5	DA	H3'	2.80
5	DA	H5'2	5	DA	H1'	3.02
5	DA	H5'2	5	DA	H3'	2.63
6	DG	H1'	6	DG	H3'	3.77
6	DG	H1'	7	DG	H8	2.00
6	DG	H2'1	6	DG	H3'	2.15
6	DG	H2'1	6	DG	H8	1.89
6	DG	H2'1	7	DG	H8	1.88
6	DG	H2'2	6	DG	H1'	2.21
6	DG	H2'2	6	DG	H3'	2.50
6	DG	H2'2	7	DG	H8	1.85
6	DG	H3'	7	DG	H8	2.88
6	DG	H4'	6	DG	H1'	3.35
6	DG	H4'	6	DG	H3'	2.61
6	DG	H5'1	6	DG	H1'	3.39
6	DG	H5'1	6	DG	H3'	2.94
6	DG	H5'1	6	DG	H4'	2.22
6	DG	H5'2	6	DG	H1'	3.07
6	DG	H5'2	6	DG	H3'	2.40
6	DG	H5'2	6	DG	H4'	2.20
6	DG	H8	5	DA	H1'	2.88
6	DG	H8	6	DG	H1'	3.18
6	DG	H8	6	DG	H2'2	1.88
6	DG	H8	6	DG	H3'	2.58
7	DG	H2'1	7	DG	H1'	2.71
7	DG	H2'1	7	DG	H2'2	1.73
7	DG	H2'1	7	DG	H3'	2.21
7	DG	H2'1	7	DG	H8	1.88

Identifier A	Residue A	Atom A	Identifier B	Residue B	Atom B	Distance
7	DG	H2'1	8	DC	H6	2.00
7	DG	H2'2	7	DG	H1'	2.15
7	DG	H2'2	7	DG	H3'	2.36
7	DG	H2'2	7	DG	H8	1.90
7	DG	H2'2	8	DC	H6	1.82
7	DG	H3'	7	DG	H1'	4.32
7	DG	H3'	7	DG	H4'	2.82
7	DG	H3'	7	DG	H8	2.69
7	DG	H4'	7	DG	H1'	3.03
7	DG	H5'1	7	DG	H1'	3.31
7	DG	H5'2	7	DG	H3'	2.50
7	DG	H8	7	DG	H1'	2.83
8	DC	H1'	8	DC	H6	2.83
8	DC	H2'1	8	DC	H1'	3.22
8	DC	H2'1	8	DC	H6	1.80
8	DC	H2'2	8	DC	H1'	2.66
8	DC	H2'2	8	DC	H6	1.85
8	DC	H3'	8	DC	H2'1	2.74
8	DC	H3'	8	DC	H2'2	3.12
8	DC	H6	8	DC	H3'	2.39
9	DT	H1'	9	DT	H2'2	2.62
9	DT	H1'	9	DT	H4'	3.18
9	DT	H2'1	9	DT	H2'2	2.01
9	DT	H2'2	9	DT	H3'	2.60
9	DT	H2'2	10	DG	H8	1.87
9	DT	H4'	9	DT	H3'	2.78
10	DG	H2'1	10	DG	H2'2	1.80
10	DG	H2'1	10	DG	H3'	2.38
10	DG	H2'1	10	DG	H8	1.84
10	DG	H2'2	10	DG	H1'	2.38
10	DG	H2'2	10	DG	H3'	2.58
10	DG	H2'2	10	DG	H8	2.10
10	DG	H2'2	11	DG	H8	1.88
10	DG	H3'	10	DG	H4'	2.81
10	DG	H4'	10	DG	H1'	3.63
10	DG	H4'	10	DG	H5'1	2.85
10	DG	H4'	10	DG	H5'2	2.33
10	DG	H5'1	10	DG	H3'	2.86
10	DG	H5'2	10	DG	H3'	3.02
11	DG	H2'1	11	DG	H1'	3.25
11	DG	H2'1	11	DG	H2'2	1.85
11	DG	H2'1	11	DG	H3'	2.36
11	DG	H2'1	11	DG	H8	1.84
11	DG	H2'1	12	DG	H8	2.32
11	DG	H2'2	11	DG	H1'	2.35
11	DG	H2'2	11	DG	H3'	2.57
11	DG	H2'2	11	DG	H8	2.10

Identifier A	Residue A	Atom A	Identifier B	Residue B	Atom B	Distance
11	DG	H2'2	12	DG	H8	1.80
11	DG	H3'	11	DG	H5'2	3.24
11	DG	H4'	11	DG	H1'	3.36
11	DG	H4'	11	DG	H3'	2.93
11	DG	H5'1	11	DG	H1'	2.69
11	DG	H5'1	11	DG	H3'	2.67
11	DG	H5'1	11	DG	H4'	2.26
11	DG	H5'2	11	DG	H4'	2.54
11	DG	H8	11	DG	H3'	3.40
12	DG	H2'1	12	DG	H1'	2.84
12	DG	H2'1	12	DG	H2'2	1.87
12	DG	H2'1	12	DG	H3'	2.23
12	DG	H2'1	12	DG	H8	1.90
12	DG	H2'1	13	DA	H8	1.83
12	DG	H2'2	12	DG	H1'	2.26
12	DG	H2'2	12	DG	H3'	2.41
12	DG	H2'2	12	DG	H3'	2.51
12	DG	H2'2	12	DG	H8	1.95
12	DG	H2'2	13	DA	H8	1.89
12	DG	H4'	12	DG	H1'	3.39
12	DG	H4'	12	DG	H3'	2.63
12	DG	H5'1	12	DG	H1'	4.40
12	DG	H5'1	12	DG	H4'	1.89
12	DG	H5'2	12	DG	H1'	2.42
12	DG	H5'2	12	DG	H3'	2.64
12	DG	H8	11	DG	H1'	3.00
12	DG	H8	12	DG	H3'	2.93
13	DA	H1'	13	DA	H3'	4.13
13	DA	H2'1	13	DA	H1'	2.31
13	DA	H2'1	13	DA	H2'2	1.89
13	DA	H2'1	13	DA	H3'	2.80
13	DA	H2'1	13	DA	H8	2.28
13	DA	H2'1	14	DG	H8	1.81
13	DA	H2'2	13	DA	H1'	2.82
13	DA	H2'2	13	DA	H3'	2.25
13	DA	H2'2	13	DA	H8	1.82
13	DA	H3'	13	DA	H8	2.01
13	DA	H5'1	13	DA	H3'	2.80
13	DA	H5'2	13	DA	H1'	3.02
13	DA	H5'2	13	DA	H3'	2.63
14	DG	H1'	14	DG	H3'	3.77
14	DG	H1'	15	DG	H8	2.00
14	DG	H2'1	14	DG	H3'	2.15
14	DG	H2'1	14	DG	H8	1.89
14	DG	H2'1	15	DG	H8	1.88
14	DG	H2'2	14	DG	H1'	2.21
14	DG	H2'2	14	DG	H3'	2.50

Identifier A	Residue A	Atom A	Identifier B	Residue B	Atom B	Distance
14	DG	H2'2	15	DG	H8	1.85
14	DG	H3'	15	DG	H8	2.88
14	DG	H4'	14	DG	H1'	3.35
14	DG	H4'	14	DG	H3'	2.61
14	DG	H5'1	14	DG	H1'	3.39
14	DG	H5'1	14	DG	H3'	2.94
14	DG	H5'1	14	DG	H4'	2.22
14	DG	H5'2	14	DG	H1'	3.07
14	DG	H5'2	14	DG	H3'	2.40
14	DG	H5'2	14	DG	H4'	2.20
14	DG	H8	13	DA	H1'	2.88
14	DG	H8	14	DG	H1'	3.18
14	DG	H8	14	DG	H2'2	1.88
14	DG	H8	14	DG	H3'	2.58
15	DG	H2'1	15	DG	H1'	2.71
15	DG	H2'1	15	DG	H2'2	1.73
15	DG	H2'1	15	DG	H3'	2.21
15	DG	H2'1	15	DG	H8	1.88
15	DG	H2'1	16	DC	H6	2.00
15	DG	H2'2	15	DG	H1'	2.15
15	DG	H2'2	15	DG	H3'	2.36
15	DG	H2'2	15	DG	H8	1.90
15	DG	H2'2	16	DC	H6	1.82
15	DG	H3'	15	DG	H1'	4.32
15	DG	H3'	15	DG	H4'	2.82
15	DG	H3'	15	DG	H8	2.69
15	DG	H4'	15	DG	H1'	3.03
15	DG	H5'1	15	DG	H1'	3.31
15	DG	H5'2	15	DG	H3'	2.50
15	DG	H8	15	DG	H1'	2.83
16	DC	H1'	16	DC	H6	2.83
16	DC	H2'1	16	DC	H1'	3.22
16	DC	H2'1	16	DC	H6	1.80
16	DC	H2'2	16	DC	H1'	2.66
16	DC	H2'2	16	DC	H6	1.85
16	DC	H3'	16	DC	H2'1	2.74
16	DC	H3'	16	DC	H2'2	3.12
16	DC	H6	16	DC	H3'	2.39
17	DT	H1'	17	DT	H2'2	2.62
17	DT	H1'	17	DT	H4'	3.18
17	DT	H2'1	17	DT	H2'2	2.01
17	DT	H2'2	17	DT	H3'	2.60
17	DT	H2'2	18	DG	H8	1.87
17	DT	H4'	17	DT	H3'	2.78
18	DG	H2'1	18	DG	H2'2	1.80
18	DG	H2'1	18	DG	H3'	2.38
18	DG	H2'1	18	DG	H8	1.84

Identifier A	Residue A	Atom A	Identifier B	Residue B	Atom B	Distance
18	DG	H2'2	18	DG	H1'	2.38
18	DG	H2'2	18	DG	H3'	2.58
18	DG	H2'2	18	DG	H8	2.10
18	DG	H2'2	19	DG	H8	1.88
18	DG	H3'	18	DG	H4'	2.81
18	DG	H4'	18	DG	H1'	3.63
18	DG	H4'	18	DG	H5'1	2.85
18	DG	H4'	18	DG	H5'2	2.33
18	DG	H5'1	18	DG	H3'	2.86
18	DG	H5'2	18	DG	H3'	3.02
19	DG	H2'1	19	DG	H1'	3.25
19	DG	H2'1	19	DG	H2'2	1.85
19	DG	H2'1	19	DG	H3'	2.36
19	DG	H2'1	19	DG	H8	1.84
19	DG	H2'1	20	DG	H8	2.32
19	DG	H2'2	19	DG	H1'	2.35
19	DG	H2'2	19	DG	H3'	2.57
19	DG	H2'2	19	DG	H8	2.10
19	DG	H2'2	20	DG	H8	1.80
19	DG	H3'	19	DG	H5'2	3.24
19	DG	H4'	19	DG	H1'	3.36
19	DG	H4'	19	DG	H3'	2.93
19	DG	H5'1	19	DG	H1'	2.69
19	DG	H5'1	19	DG	H3'	2.67
19	DG	H5'1	19	DG	H4'	2.26
19	DG	H5'2	19	DG	H4'	2.54
19	DG	H8	19	DG	H3'	3.40
20	DG	H2'1	20	DG	H1'	2.84
20	DG	H2'1	20	DG	H2'2	1.87
20	DG	H2'1	20	DG	H3'	2.23
20	DG	H2'1	20	DG	H8	1.90
20	DG	H2'1	21	DA	H8	1.83
20	DG	H2'2	20	DG	H1'	2.26
20	DG	H2'2	20	DG	H3'	2.41
20	DG	H2'2	20	DG	H3'	2.51
20	DG	H2'2	20	DG	H8	1.95
20	DG	H2'2	21	DA	H8	1.89
20	DG	H4'	20	DG	H1'	3.39
20	DG	H4'	20	DG	H3'	2.63
20	DG	H5'1	20	DG	H1'	4.40
20	DG	H5'1	20	DG	H4'	1.89
20	DG	H5'2	20	DG	H1'	2.42
20	DG	H5'2	20	DG	H3'	2.64
20	DG	H8	19	DG	H1'	3.00
20	DG	H8	20	DG	H3'	2.93
21	DA	H1'	21	DA	H3'	4.13
21	DA	H2'1	21	DA	H1'	2.31

Identifier A	Residue A	Atom A	Identifier B	Residue B	Atom B	Distance
21	DA	H2'1	21	DA	H2'2	1.89
21	DA	H2'1	21	DA	H3'	2.80
21	DA	H2'1	21	DA	H8	2.28
21	DA	H2'1	22	DG	H8	1.81
21	DA	H2'2	21	DA	H1'	2.82
21	DA	H2'2	21	DA	H3'	2.25
21	DA	H2'2	21	DA	H8	1.82
21	DA	H3'	21	DA	H8	2.01
21	DA	H5'1	21	DA	H3'	2.80
21	DA	H5'2	21	DA	H1'	3.02
21	DA	H5'2	21	DA	H3'	2.63
22	DG	H1'	22	DG	H3'	3.77
22	DG	H1'	23	DG	H8	2.00
22	DG	H2'1	22	DG	H3'	2.15
22	DG	H2'1	22	DG	H8	1.89
22	DG	H2'1	23	DG	H8	1.88
22	DG	H2'2	22	DG	H1'	2.21
22	DG	H2'2	22	DG	H3'	2.50
22	DG	H2'2	23	DG	H8	1.85
22	DG	H3'	23	DG	H8	2.88
22	DG	H4'	22	DG	H1'	3.35
22	DG	H4'	22	DG	H3'	2.61
22	DG	H5'1	22	DG	H1'	3.39
22	DG	H5'1	22	DG	H3'	2.94
22	DG	H5'1	22	DG	H4'	2.22
22	DG	H5'2	22	DG	H1'	3.07
22	DG	H5'2	22	DG	H3'	2.40
22	DG	H5'2	22	DG	H4'	2.20
22	DG	H8	21	DA	H1'	2.88
22	DG	H8	22	DG	H1'	3.18
22	DG	H8	22	DG	H2'2	1.88
22	DG	H8	22	DG	H3'	2.58
23	DG	H2'1	23	DG	H1'	2.71
23	DG	H2'1	23	DG	H2'2	1.73
23	DG	H2'1	23	DG	H3'	2.21
23	DG	H2'1	23	DG	H8	1.88
23	DG	H2'1	24	DC	H6	2.00
23	DG	H2'2	23	DG	H1'	2.15
23	DG	H2'2	23	DG	H3'	2.36
23	DG	H2'2	23	DG	H8	1.90
23	DG	H2'2	24	DC	H6	1.82
23	DG	H3'	23	DG	H1'	4.32
23	DG	H3'	23	DG	H4'	2.82
23	DG	H3'	23	DG	H8	2.69
23	DG	H4'	23	DG	H1'	3.03
23	DG	H5'1	23	DG	H1'	3.31
23	DG	H5'2	23	DG	H3'	2.50

Identifier A	Residue A	Atom A	Identifier B	Residue B	Atom B	Distance
23	DG	H8	23	DG	H1'	2.83
24	DC	H1'	24	DC	H6	2.83
24	DC	H2'1	24	DC	H1'	3.22
24	DC	H2'1	24	DC	H6	1.80
24	DC	H2'2	24	DC	H1'	2.66
24	DC	H2'2	24	DC	H6	1.85
24	DC	H3'	24	DC	H2'1	2.74
24	DC	H3'	24	DC	H2'2	3.12
24	DC	H6	24	DC	H3'	2.39
25	DT	H1'	25	DT	H2'2	2.62
25	DT	H1'	25	DT	H4'	3.18
25	DT	H2'1	25	DT	H2'2	2.01
25	DT	H2'2	25	DT	H3'	2.60
25	DT	H2'2	26	DG	H8	1.87
25	DT	H4'	25	DT	H3'	2.78
26	DG	H2'1	26	DG	H2'2	1.80
26	DG	H2'1	26	DG	H3'	2.38
26	DG	H2'1	26	DG	H8	1.84
26	DG	H2'2	26	DG	H1'	2.38
26	DG	H2'2	26	DG	H3'	2.58
26	DG	H2'2	26	DG	H8	2.10
26	DG	H2'2	27	DG	H8	1.88
26	DG	H3'	26	DG	H4'	2.81
26	DG	H4'	26	DG	H1'	3.63
26	DG	H4'	26	DG	H5'1	2.85
26	DG	H4'	26	DG	H5'2	2.33
26	DG	H5'1	26	DG	H3'	2.86
26	DG	H5'2	26	DG	H3'	3.02
27	DG	H2'1	27	DG	H1'	3.25
27	DG	H2'1	27	DG	H2'2	1.85
27	DG	H2'1	27	DG	H3'	2.36
27	DG	H2'1	27	DG	H8	1.84
27	DG	H2'1	28	DG	H8	2.32
27	DG	H2'2	27	DG	H1'	2.35
27	DG	H2'2	27	DG	H3'	2.57
27	DG	H2'2	27	DG	H8	2.10
27	DG	H2'2	28	DG	H8	1.80
27	DG	H3'	27	DG	H5'2	3.24
27	DG	H4'	27	DG	H1'	3.36
27	DG	H4'	27	DG	H3'	2.93
27	DG	H5'1	27	DG	H1'	2.69
27	DG	H5'1	27	DG	H3'	2.67
27	DG	H5'1	27	DG	H4'	2.26
27	DG	H5'2	27	DG	H4'	2.54
27	DG	H8	27	DG	H3'	3.40
28	DG	H2'1	28	DG	H1'	2.84
28	DG	H2'1	28	DG	H2'2	1.87

Identifier A	Residue A	Atom A	Identifier B	Residue B	Atom B	Distance
28	DG	H2'1	28	DG	H3'	2.23
28	DG	H2'1	28	DG	H8	1.90
28	DG	H2'1	29	DA	H8	1.83
28	DG	H2'2	28	DG	H1'	2.26
28	DG	H2'2	28	DG	H3'	2.41
28	DG	H2'2	28	DG	H3'	2.51
28	DG	H2'2	28	DG	H8	1.95
28	DG	H2'2	29	DA	H8	1.89
28	DG	H4'	28	DG	H1'	3.39
28	DG	H4'	28	DG	H3'	2.63
28	DG	H5'1	28	DG	H1'	4.40
28	DG	H5'1	28	DG	H4'	1.89
28	DG	H5'2	28	DG	H1'	2.42
28	DG	H5'2	28	DG	H3'	2.64
28	DG	H8	27	DG	H1'	3.00
28	DG	H8	28	DG	H3'	2.93
29	DA	H1'	29	DA	H3'	4.13
29	DA	H2'1	29	DA	H1'	2.31
29	DA	H2'1	29	DA	H2'2	1.89
29	DA	H2'1	29	DA	H3'	2.80
29	DA	H2'1	29	DA	H8	2.28
29	DA	H2'1	30	DG	H8	1.81
29	DA	H2'2	29	DA	H1'	2.82
29	DA	H2'2	29	DA	H3'	2.25
29	DA	H2'2	29	DA	H8	1.82
29	DA	H3'	29	DA	H8	2.01
29	DA	H5'1	29	DA	H3'	2.80
29	DA	H5'2	29	DA	H1'	3.02
29	DA	H5'2	29	DA	H3'	2.63
30	DG	H1'	30	DG	H3'	3.77
30	DG	H1'	31	DG	H8	2.00
30	DG	H2'1	30	DG	H3'	2.15
30	DG	H2'1	30	DG	H8	1.89
30	DG	H2'1	31	DG	H8	1.88
30	DG	H2'2	30	DG	H1'	2.21
30	DG	H2'2	30	DG	H3'	2.50
30	DG	H2'2	31	DG	H8	1.85
30	DG	H3'	31	DG	H8	2.88
30	DG	H4'	30	DG	H1'	3.35
30	DG	H4'	30	DG	H3'	2.61
30	DG	H5'1	30	DG	H1'	3.39
30	DG	H5'1	30	DG	H3'	2.94
30	DG	H5'1	30	DG	H4'	2.22
30	DG	H5'2	30	DG	H1'	3.07
30	DG	H5'2	30	DG	H3'	2.40
30	DG	H5'2	30	DG	H4'	2.20
30	DG	H8	29	DA	H1'	2.88

Identifier A	Residue A	Atom A	Identifier B	Residue B	Atom B	Distance
30	DG	H8	30	DG	H1'	3.18
30	DG	H8	30	DG	H2'2	1.88
30	DG	H8	30	DG	H3'	2.58
31	DG	H2'1	31	DG	H1'	2.71
31	DG	H2'1	31	DG	H2'2	1.73
31	DG	H2'1	31	DG	H3'	2.21
31	DG	H2'1	31	DG	H8	1.88
31	DG	H2'1	32	DC	H6	2.00
31	DG	H2'2	31	DG	H1'	2.15
31	DG	H2'2	31	DG	H3'	2.36
31	DG	H2'2	31	DG	H8	1.90
31	DG	H2'2	32	DC	H6	1.82
31	DG	H3'	31	DG	H1'	4.32
31	DG	H3'	31	DG	H4'	2.82
31	DG	H3'	31	DG	H8	2.69
31	DG	H4'	31	DG	H1'	3.03
31	DG	H5'1	31	DG	H1'	3.31
31	DG	H5'2	31	DG	H3'	2.50
31	DG	H8	31	DG	H1'	2.83
32	DC	H1'	32	DC	H6	2.83
32	DC	H2'1	32	DC	H1'	3.22
32	DC	H2'1	32	DC	H6	1.80
32	DC	H2'2	32	DC	H1'	2.66
32	DC	H2'2	32	DC	H6	1.85
32	DC	H3'	32	DC	H2'1	2.74
32	DC	H3'	32	DC	H2'2	3.12
32	DC	H6	32	DC	H3'	2.39