2002

Electrospray mass spectrometry and nuclear magnetic resonance spectroscopy of DNA-interactive agents

Michelle Lisa Colgrave
University of Wollongong

ELECTROSPRAY MASS SPECTROMETRY AND NUCLEAR MAGNETIC RESONANCE SPECTROSCOPY OF DNA-INTERACTIVE AGENTS

A thesis submitted in fulfillment of the requirements for the award of the degree

DOCTOR OF PHILOSOPHY

from

The University of Wollongong

by

Michelle Lisa Colgrave, BSc(Hons)

Department of Chemistry
June, 2002

CERTIFICATION

I, Michelle Lisa Colgrave, declare that this thesis, submitted in fulfillment of the requirements for the award of Doctor of Philosophy, in the Department of Chemistry, University of Wollongong, is wholly my own work unless otherwise referenced or acknowledged. The document has not been submitted for qualifications at any other academic institution.

Michelle Lisa Colgrave

18 June, 2001

ACKNOWLEDGEMENTS

- *Firstly, I would like to thank Prof. Margaret Sheil and Dr. Mark Searle for giving me the opportunity to work under their supervision on this fascinating area of research and also for their guidance and patience throughout.*

- *I would like to extend my most sincere thanks to all of the members of the two research groups in which I was privileged to have worked. These are the Biomolecular Mass Spectrometry Laboratory at the University of Wollongong, Australia, and Mark Searle's NMR Spectroscopy group within the School of Chemistry at the University of Nottingham, UK.*

- *Special thanks, in particular, go to Dr. Jennifer Beck in Wollongong and Dr. Huw Williams in Nottingham for their help and expertise.*

- *Thanks also go out to all the technical staff at both institutions without whose help I would never have got anything to work.*

- *Additionally, I would like to thank the ARC for the funding of my scholarship, the Boehringer Ingelheim Fonds Institute for the travel allowance to the UK and the Universities of Wollongong and Nottingham for additional funding to attend conferences in Barcelona, Spain and Sheffield, UK.*

PUBLICATIONS

2002

Williams, H.E.L.; Colgrave, M.L. and Searle, M.S.
Drug recognition of a single-strand break: nogalamycin intercalation between two co-axially stacked hairpins
European J. Biochem. (2002) **269(6)**, 1726-1733

Colgrave, M.L.; Beck, J.L., Sheil, M.M. and Searle, M.S.
Electrospray ionisation mass spectrometric detection of weak non-covalent interactiuons in nogalamycin-DNA complexes
Chemical Communications (2002) **6**, 556-557

2001

Colgrave, M.L.; Iannitti-Tito, P.; Wickham, G. and Sheil, M.M.
Rapid determination of sequence selectivity and stability of alkylated oligonucleotide adducts by electrospray tandem mass spectrometry
Biochemistry (2001) submitted

Beck, J.L.; Colgrave, M.L.; Ralph, S.R. and Sheil, M.M.
Electrospray ionisation mass spectrometry of oligonucleotide complexes with drugs, metals and proteins
Mass Spectrometry Reviews (2001) **20(2)**, 61-87

Colgrave, M.L.; Williams, H.E.L. and Searle, M.S.
Binding of anthracycline antibiotic nogalamycin to the site of a DNA single strand break engineered between two co-axially stacked hairpins
Chemical Communications (2001) **4**, 315-316

Colgrave, M.L.; Williams, H.E.L. and Searle, M.S.
Drug recognition of a DNA single strand break: nogalamycin intercalation between two co-axially stacked hairpins containing stabilising 5'-GAA loops
NACON V (9^{th}-12^{th} April, 2001) Sheffield, UK

Beck, J.L.; Colgrave, M.L.; Kapur, A.; Iannitti-Tito, P; Ralph, S. R.; Sheil, M.M.; Weimann, A. and Wickham, G.
Electrospray and tandem mass spectrometry of drug-DNA complexes
Advances in Mass Spectrometry (2001) vol 15, John Wiley & Sons, Chichester (Emilio Gulpi, Ed.) pp 175-192

2000

Colgrave, M.L.; Wickham, G. and Sheil, M.M.
ESI-MS/MS investigations into the sequence selectivity of DNA-interactive ligands: the pyrindamycins
15^{th} IMSC Conference (26^{th} Aug-1^{st} Sept, 2000) Barcelona, Spain

1999

Colgrave, M. L.; Boschenok, J.; Wickham, G. and Sheil, M. M.
Investigation of pyrindamycin A binding to oligonucleotides
ANZSMS 17 (31^{st} Jan-4^{th} Feb, 1999) Thredbo, Australia

ABSTRACT

The binding of the antitumour antibiotics pyrindamycin A (duocarmycin C_2) and pyrindamycin B (duocarmycin C_1) to self-complementary oligonucleotides has been studied using electrospray ionisation mass spectrometry (ESI-MS) and tandem mass spectrometry (MS/MS). The pyrindamycins bind via non-covalent interactions in the minor groove of DNA with subsequent alkylation of the N3 of adenine (particularly within 5′-AAA and 5′-TTA sequences). The relative stability of adducts with different oligonucleotides can be inferred from the ESI mass spectra. MS/MS spectra of alkylated adducts enable facile identification of the site of alkylation. We show here that the site(s) of alkylation by the pyrindamycins are strongly influenced by the location of the preferred high affinity binding sites within these short oligonucleotides. Alkylation of guanine by pyrindamycin A is shown to occur only upon binding to single-stranded oligonucleotides, providing further confirmation of the importance of the initial non-covalent interaction in conferring sequence selectivity. These data clearly demonstrate value of using ESI-MS/MS to pre-screen ligand-oligonucleotide complexes prior to performing more detailed structural studies.

In the second part of this work, the binding of nogalamycin to unusual DNA structures was explored by nuclear magnetic resonance (NMR) spectroscopy. The first of these was an oligonucleotide 5′-d(GCGAAGCACGAAGT)-3′, which was designed to form a double hairpin or loop sequence. Two further oligonucleotides, 5′-d(GTGCGAAGCTAC)-3′ and 5′-d(GCTACGAAGTGC)-3′ were designed to incorporate a mispaired thymine, which may then form a bulge. The folding of the hairpins was evidenced by anomalous chemical shift values for the sugar H4', H5' and H5" protons (characteristic of the adenine involved in the loop) and the presence of imino protons (protected from exchange by hydrogen bonding interactions in the stem).

Nogalamycin formed 1:1 complexes with each of the oligonucleotides with the ligand bound at the TpG step. Upon ligand binding, changes in chemical shift values for the DNA resonances were observed, with the largest differences noted for the nucleotides in the

intercalation site. Nogalamycin bound to 5′-d(GCGAAGCACGAAGT)-3′ at the nicked site and stabilised the double hairpin structure through co-axial stacking interactions. The intercalation of nogalamycin to 5′-d(GTGCGAAGCTAC)-3′ gave rise to an equilibrium in solution between a frame-shifted conformation and bulged conformation. The binding of nogalamycin to 5′-d(GCTACGAAGTGC)-3′ stabilised the bulge positioned at T_3. Finally, ESI-MS and melting temperature data indicate that the relative stability of the complexes formed with nogalamycin is in the order: 5′-d(GCGAAGCACGAAGT)-3′-Ng > 5′-d(GCTACGAAGTGC)-3′-Ng > 5′-d(GTGCGAAGCTAC)-3′-Ng.

ABBREVIATIONS & SYMBOLS

1D	one-dimensional
2D	two-dimensional
A	adenine
C	cytosine
COSY	correlation spectroscopy
dA	deoxyadenosine
dC	deoxycytosine
dG	deoxyguanidine
DNA	deoxyribonucleic acid
DQF	double quantum filtered
dT	deoxythymine
EDTA	ethylenediaminetetra acetic acid
FID	free induction decay
FT NMR	Fourier transformed nuclear magnetic resonance spectroscopy
G	guanine
Ng	nogalamycin
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
NOESY	nuclear Overhauser effect spectroscopy
Nogalamycin	2,6-epoxy-2H-naphthacenol[1,2-β]oxocin-14-carboxylic acid, 11-[(6-deoxy-3-C-methyl-2,3,4-tri-0-methyl-α-L-mannopyranosyl)oxy]-4-(dimethylamino)-3,4,5,6,9,11,12,13,14,16-decahydro-3,5,8,10,13-pentahydroxy-6,13-dimethyl-9,16-dioxo-, methyl ester
ppm	parts per million
RNA	ribonucleic acid
T	thymine
t1	first time domain
T1	longtitudinal relaxation (spin-lattice)
t2	second time domain
T2	transverse relaxation (spin-spin)
TOCSY	total correlation spectroscopy
TSP	trisilylpropanoate
wg	water gradient attenuation by tailored excitation

TABLE OF CONTENTS

Chapter Three: Sequence Selectivity, Stability, the Pyrindamycins and Mass Spectrometry

---

Chapter Four: The Novel DNA Structures and NMR

Chapter Five: Bulge Sequences, a Variation on the Hairpin

LIST OF SCHEMES

LIST OF FIGURES

Chapter Three

Chapter Four

Chapter Five

LIST OF TABLES

Chapter Five

Chapter Five

CHAPTER ONE: *DNA and the DNA-interactive ligands*

1.0 Introduction

Nucleic acids were discovered in 1868 when Friedrich Miescher isolated the nuclei from human pus cells. (Miescher, 1879) Subsequently, it was shown that nucleic acids were a major component of cells and tissues, however, at that time there was no understanding of the function of these nuclear materials. Hydrolysis of the nucleic acids from thymus glands was found to yield the nucleobases adenine (A), guanine (G), cytosine (C) and thymine (T), a deoxypentose sugar and phosphoric acid, whilst from yeast, uracil (U) instead of thymine and a pentose sugar in the place of a deoxypentose were identified. These sugars were later determined to be deoxyribose and ribose forming the basis of the names deoxyribonucleic acid (DNA) and ribonucleic acid (RNA). (Adams *et al.*, 1992) It was hypothesized, from the approximately equimolar proportions of bases present, that DNA consisted of repeating units of the four bases, *i.e.* the purines (adenine and guanine) and the pyrimidine (cytosine and thymine). Better quantitative methods and work by Chargaff (Chargaff, 1955) in the early 1950's helped to disprove this theory. Regularities were discovered in the composition of DNA, such as that the sum of the amino bases (A and C) equaled the sum of the keto bases (G and T) and that adenine and thymine were present in equivalent amounts, as were guanine and cytosine. This information was combined with X-ray diffraction data (Astbury, 1947; Pauling & Corey, 1953; Wilkins *et al.*, 1953; Franklin & Gosling, 1953) in the proposed double-helical structure composed of specifically hydrogen-bonded base pairs (Watson-Crick base pairs). (Watson & Crick, 1953) Proposals for the mechanism of DNA replication were then forthcoming, in which one strand of DNA would act as a template for the other. It was demonstrated that upon infection of bacteria with bacteriophage, there was synthesis of RNA complementary to the DNA (transcription),

and that this RNA associated with bacterial ribosomes to produce bacteriophage protein (translation). (Brenner *et al.*, 1961)

Many proteins interact with DNA owing to the essential roles that DNA plays in cellular replication and transcription. (Beck *et al.*, 2001) The disruption of DNA or these protein-DNA interactions may have a dramatic and possibly fatal effect on the cell and cellular processes. A large number of bacteria exploit this through the production of compounds that inhibit normal cellular operation. These compounds are known as antibiotics and may possess antitumour, anti-viral or antimicrobial properties. Many of these drugs interact with DNA in a sequence-specific manner. These effects are most pronounced on cells that grow and divide at an accelerated rate, however, the drugs also affect normal cells. Rational drug design is aimed at producing novel compounds, which affect the target with minimal impact on other cells. In order to accomplish this goal, however, an understanding of the mode of interaction between these drugs and DNA is required. These studies aim to explore the interaction between existing medicinal agents and DNA so as to aid in the design and evaluation of a new generation of superior analogs with greater specificity and fewer side effects.

1.1 DNA Structure

Nucleic acids are polymers comprised of nucleotides linked together by phosphodiester bonds, in which the internucleotide linkage involves the 3′ and 5′ hydroxyls of the sugar, so as to form a 3′-5′-phosphodiester linkage. Information relating to the secondary structure of DNA first came from X-ray diffraction studies (Astbury, 1947; Pauling & Corey, 1953; Wilkins *et al.,* 1953; Franklin & Gosling, 1953) and was complemented by early modelling attempts. (Watson & Crick, 1953) In these models, two complementary strands associate together to form a right-handed double-helical structure. The planar bases occur in the center of the helix and stack on top of each other. There are two grooves present on the surface of the B-form of DNA, (more below) known as the major and minor grooves. As their names suggest, the major groove is wide, whilst the minor groove is narrow. These grooves create unique environments for the binding and recognition of ligands. (Long, 1996) The floors of the major and minor grooves vary depending on the base-pair sequences, which provide hydrogen-bond donor and acceptor sites, thus giving rise to the sequence selectivity of binding for different ligands as shown in figure 1.1.

The sugar-phosphate backbone lies on the outside of the helix and is polyanionic (one negative charge per phosphate at neutral pH) making it hydrophilic in nature. The stacked bases form a hydrophobic core, however, the amino and keto groups present on the bases point into the grooves allowing for non-covalent interactions with the solvent and/or other molecules.

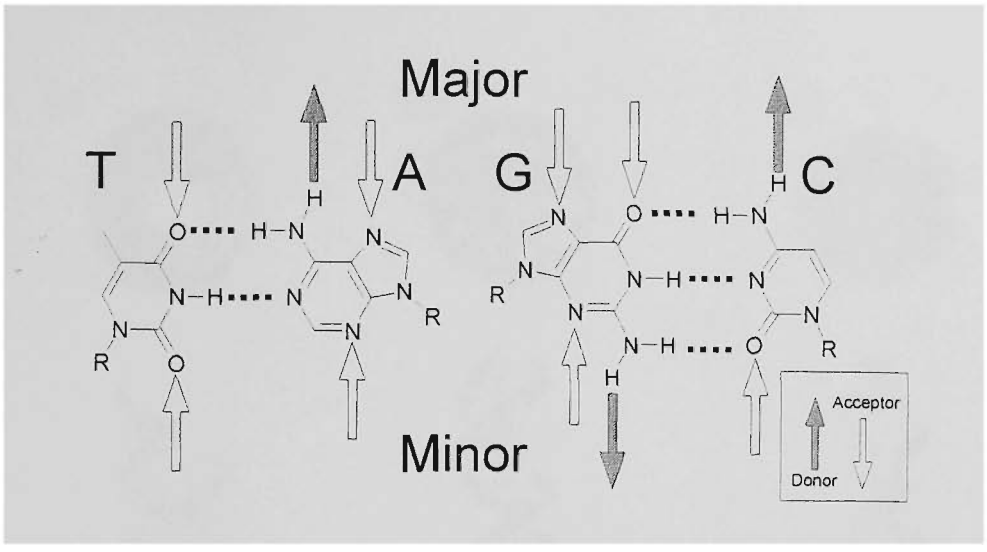


Figure 1.1: Watson-Crick base pairs indicating major and minor grooves and hydrogen bond donor and acceptor sites.

1.2 Polymorphism: B-form versus A and Z-forms

DNA has been shown to adopt a number of conformations depending on both its sequence and environmental conditions as shown in figure 1.2. The B-form is the right-handed helical conformation adopted by DNA under physiological conditions. The A-form is also right-handed, but differs in the pitch, the number of bases per turn and the widths and depths of the minor and major grooves. A-form differs from B-form in the sugar conformation (C3'-*endo* versus C2'-*endo*). DNA is known to adopt the A-form owing to protein associations or in different solvents (*i.e.* in low humidity). Z-DNA is an even more unusual structure formed by alternating purine-pyrimidine sequences, which takes on a left-handed helical in solutions with high salt content. The presence of these different conformations of DNA in solution has been confirmed by various NMR techniques. (Neidle *et al.,* 1987; Cohen, 1987; Delepierre *et al.,* 1986)

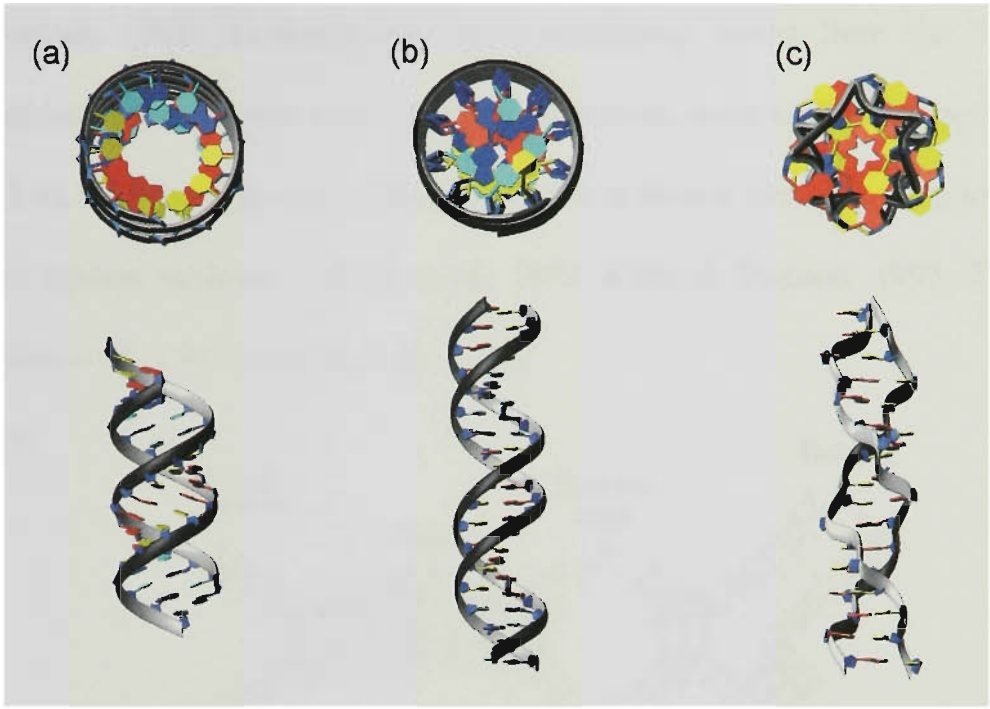


Figure 1.2: Polymorphism of DNA: (a) A-form DNA; (b) B-form DNA and (c) Z-form DNA. (Williams, 2000)

1.3 Novel Structures

DNA also adopts a number of conformations that do not exclusively involve Watson-Crick base pairing. These include triple-stranded structures, tetraplexes, some DNA-drug complexes and hairpin loops (figure 1.3).

Antisense oligonucleotides (*i.e.* oligonucleotides that stop mRNA from being translated into protein) and modified oligonucleotides have been quite widely used for the regulation of gene expression, in particular, for genes associated with cancer or in the replication of viruses and parasites (Agrawal, 1992) However, these oligonucleotides are susceptible to degradation by nucleases. As such, there is an ongoing interest into the modification of these DNA fragments to protect against degradation. Modifications to the phosphate backbone (e.g. conversion to phosphothioates) have lead to increased nuclease resistance, although these modified oligonucleotides are still degraded over

time. (Iverson, 1991) As degradation most commonly occurs from the 3′-terminus, modifications at this site are most effective. Structures, such as DNA hairpins (figures 1.3 and 1.4), have been shown to be useful in the antisense strategy owing to their high resistance against nucleases. (Hirao *et al.*, 1993; Khan & Coulson, 1993; Tang *et al.*, 1993; Jolles *et al.*, 1997; Chraibi *et al.*, 1999)

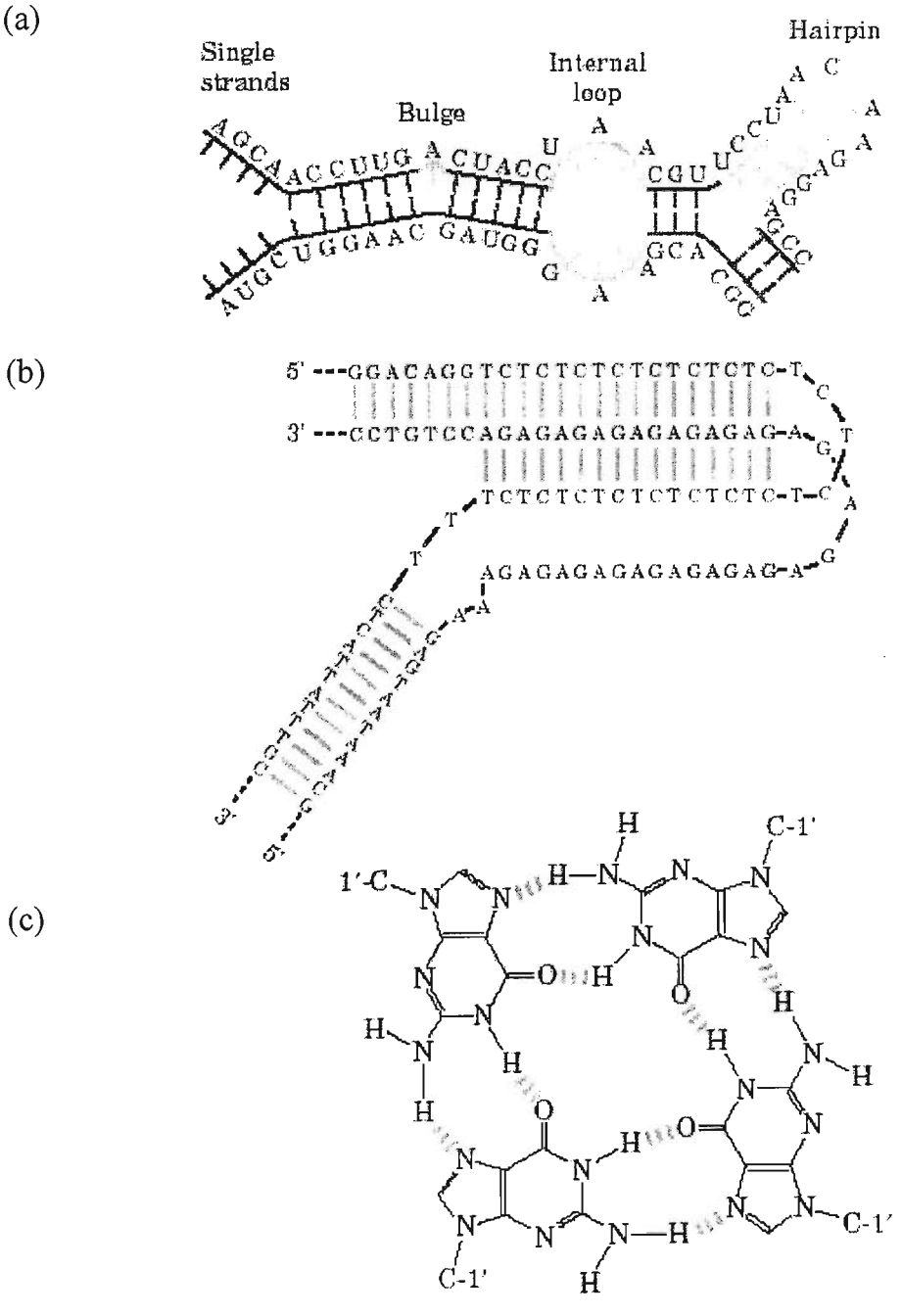


Figure 1.3: Unusual DNA conformations: (a) DNA structure showing bulge, loop and hairpin features; (b) triple-stranded DNA (triplex); and (c) guanine quartet of a G tetraplex. (Nelson & Cox, 2000)

Hairpin loops form when nucleic acid strands fold back on themselves to form base pairs. They consist of a single stranded loop region, which is closed by a base paired stem region. Generally, the greater the number of base pairs in the stem region, the greater the stability of the hairpin structure. (Yoshizawa *et al.*, 1997) Hairpins are a dominant secondary structural feature in RNA. They have high thermodynamic stability and define nucleation sites for folding, determine tertiary interactions in RNA enzymes, protect mRNAs from degradation and are recognised by RNA binding proteins. (Varani, 1995) They are far less common in DNA and consequently they have been less frequently reported. (Vallone *et al.,* 1999) There have, however, been a number of recent reports exploring the RNA and DNA sequences that form hairpin structures, which are extraordinarily stable. (Hirao *et al.,* 1989, 1992, 1993; Tang *et al.,* 1993; Khan & Coulson, 1993; Crooke, 1992; Varani, 1995; Yoshizawa *et al.*, 1994, 1997; Chraibi *et al.,* 2000) As well as their possible application to the antisense strategy, these novel structures allow the creation of a nicked site through the co-axial stacking of two hairpins (more later).

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 also further stabilise the loop section. The loop of the GAA mini-hairpin is composed of a single nucleotide, *i.e.*, the central A. 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.

Hirao and co-workers (Hirao *et al.*, 1992) compared a variety of sequence variants of the 5′-d(GCGAAGC)-3′ mini hairpin as well as the corresponding RNA fragments. 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, *i.e.*, B-form for DNA and A-form for RNA. The structure of the 5′-d(GCGAAGC)-3′ oligonucleotide was determined by NMR spectroscopy. (Hirao *et al.*, 1994) The hairpin structure was found to be folded back between A_4 and A_5, and stabilised by two Watson-Crick base pairs and an additional non-Watson-Crick G-A base pair in the GAA loop. Extensive base stacking interactions, especially between $G_1C_2G_3A_4$ and $A_5G_6C_7$ were observed to further stabilise the hairpin structure.

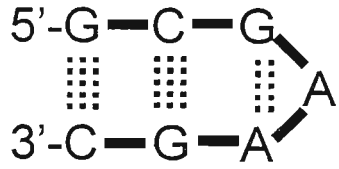


Figure 1.4: Schematic of the DNA mini-hairpin. (Hirao *et al.*, 1994)

In a study by Yoshizawa *et al.*, (Yoshizawa *et al.*, 1997) all 64 variants of the 5′-d(GCNNNGC)-3′ sequence (where N=A, G, C, T) were synthesised and characterised in terms of their ability to form stable hairpin structures. Mobility studies on polyacrylamide gels were conducted to determine which DNA fragments formed a more compact structure. They also examined the resistance of each fragment to exonuclease activity of T4 DNA polymerase. NMR studies, including NOESY, ^{31}P-decoupled DQF-COSY and HOHAHA (see section 2.6) experiments, allowed the elucidation of much structural information and highlighted the importance of the base-stacking interactions.

Overall, these workers found that only hairpins with GNA loops formed extraordinarily stable trinucleotide-loop hairpins. The stability is proposed to arise as a result of the additional G-A pair formed between G_3 and A_5 and the extensive base stacking interactions between $G_1C_2G_3A_4$ and $A_5G_6C_7$ as mentioned previously. The mini-hairpin structures described in this paper have increased thermostability and show high resistance against nucleases. This in turn makes the structures amenable to the stabilisation of oligonucleotides through tagging the mini-hairpin sequence to the 3′-termini. This method has been applied to the protection of antisense DNAs (Khan & Coulson, 1993; Poddevin *et al.*, 1994; Tang *et al.*, 1993) and primers. (Caetono-Anolles & Gresshoff, 1994)

1.4 Bulged DNA Sequences

Bulged structures in DNA are of general biological significance and are potential targets for therapeutic drugs. They are simple and common, non-helical features of RNA and DNA and may vary from single nucleotide bulges to multiple base bulges. Bulges are known to be important in RNA for protein binding and tertiary folding. (Peattie *et al.*, 1981; Romaniuk *et al.* 1987; Flor *et al.*, 1989; Cate *et al.*, 1996, 1997) In DNA, bulges play a role in frame shift mutagenesis in sequences with repeating base pairs. (Okada *et al.*, 1972) There have been relatively few reports involving the characterisation of the interactions of bulged oligonucleotides with ligands.

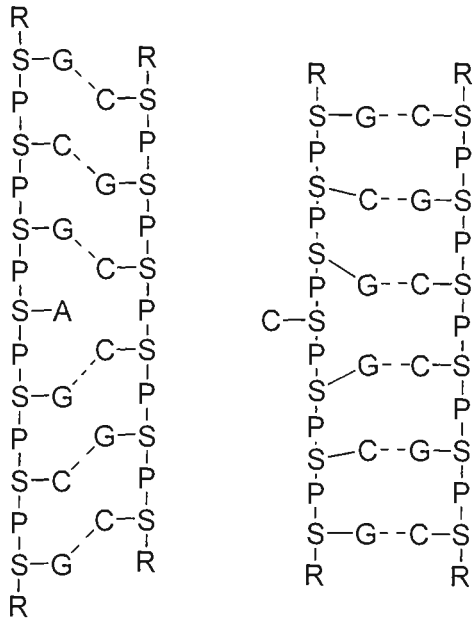


Figure 1.5: Schematic of bulged bases in DNA, where S = deoxyribose and P = phosphate groups in backbone: (a) bulged adenine in staked-in conformation; and (b) bulged cytosine in looped-out conformation.

Nucleic acid duplex regions often contain bulge loops in which there are one or more consecutive bases on one strand with no paired bases 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 types of conformations: a 'stacked-in' structure in which the unpaired base is stacked between neighbouring bases; and a 'looped-out' structure in which the unpaired base faces away from the helical axis (figure 1.5). The extrahelical base and the resultant bulge structure result in the partial misalignment of the primer and template strands during DNA replication so that addition or deletion of bases may take place in what is commonly known as frame shift mutagenesis. (Streisinger *et al.*, 1966) The importance of bulge sites as recognition elements has been established by a number of studies, including the binding of the intercalator, ethidium to bulged DNA sequences. Selective binding of

ethidium bromide to oligonucleotides containing cytosine bulges was found to be an order of magnitude greater in strength than binding to normal duplexes. (Nelson & Tinoco, 1985; White & Draper, 1987) The subsequent stabilisation and extended lifetime of the bulged structure would further increase the probability of frame shift mutagenesis occurring. Further evidence of the heightened affinity of drugs for bulge sites in DNA came from the results of competition dialysis experiments, in which 9-aminoacridine preferentially bound molecules that contained an extrahelical base. (Woodson & Crothers, 1988c) The resultant complex was determined to have additional stabilisation relative to the non-defective DNA complex through examination of the optical meting curves. (Woodson & Crothers, 1988c)

During the 1980's, an intensive research effort focussed on the determination of the conformation of these single nucleotide bulges within short DNA sequences. Patel and co-workers (Patel *et al.*, 1982; Hare *et al.*, 1986) demonstrated that an extra adenine stacks into an otherwise self-complementary oligonucleotide in solution, leaving the flanking base pairs intact. In contrast, studies undertaken by others (Morden *et al.*, 1983) found that the flanking bases to the unpaired cytosine were stacked on top of each other and that the base must be in fact extrahelical. X-ray analysis (Miller *et al.*, 1988; Joshua-Tor *et al.*, 1988, 1992) of oligonucleotides containing a single, unmatched adenine have revealed that the adenine bulge takes up the looped out conformation, whilst NMR studies of similar sequences indicate that the adenine is intrahelical. (Patel *et al.*, 1982; Hare *et al.*, 1986; Niconowicz *et al.*, 1989, 1990) These contrasting results were proposed to arise from differences in hydration in the solid and solution states and the effect of hydrophobic contributions. Conformational search and energy minimisation studies gave results in agreement with the NMR data suggesting that

adenine remains in the 'stacked in' conformation. (Zacharias & Sklenar, 1997) A series of papers by Woodson and Crothers (Woodson & Crothers, 1987; Woodson & Crothers, 1988a; Woodson & Crothers, 1988b; Woodson & Crothers, 1988c) examined the conformation and overall structure of oligonucleotides containing guanine bulges. Their results showed that the guanine stacked in to the helix and added further weight to the hypothesis that additional unpaired purines remain stacked in to the helix, while pyrimidines are in exchange between stacked and unstacked conformations. (Woodson & Crothers, 1988a; Patel *et al.*, 1982; Hare *et al.*, 1986) Bulge migration in homopolymeric sequences was investigated, in which the migration of the guanine bulge over several sites had a lower destabilising effect than when fixed at a particular site. Energy minimisation work revealed the importance of contributions from base stacking interactions toward stabilisation of the 'stacked in' conformation at the bulge site. The helix unwinding owing to the presence of unpaired bases was determined to be ca. 21° from gel electrophoresis studies. (Rice & Crothers, 1989)

A temperature-dependence has been noted for the transition between the 'looped out' and 'stacked in' conformations for the pyrimidine bases at bulge sites. (Kalnik *et al.*, 1989a; Woodson & Crothers, 1988c) At low temperatures, cytosine was observed to take the 'looped out' conformation, whilst at elevated temperature, it takes the 'stacked in' conformation, immediately prior to the helix-coil melting transition. (Kalnik *et al.*, 1989a) The conflicting conformations observed for adenine (Patel *et al.*, 1982) and cytosine (Morden *et al.*, 1983) were postulated to result from the base stacking contributions of either the extra base (*i.e.* purine or pyrimidine) or the flanking bases, to the stability of the helix at the bulge site. Patel and co-workers (Kalnik *et al.*, 1989a; Kalnik *et al.*, 1989b; Kalnik *et al.*, 1990) examined the effect that changing the flanking

bases had on the conformation of the extrahelical base at the interrupted site. They found that adenine stacked into the helix independent of the flanking bases, whilst thymine exhibited the same equilibrium behaviour as cytosine between 'looped out' at low temperature and 'stacked in' at higher temperature when it was flanked by guanine. When the thymine was flanked by cytosines, however, the 'looped out' conformation was observed at all temperatures examined. (Kalnik *et al.*, 1990) The influence of neighbouring base pairs on the stability of single base bulges was examined by temperature gradient gel electrophoresis. (Ke & Wartell, 1995) The results showed that purine bases were less destabilising than pyrimidines, and that when the unpaired base was identical to one of its adjacent bases, the resultant structure was more stable than if flanked by differing bases.

The cleavage of duplex DNA by the neocarzinostatin chromophore (NCS-Chrom), at the 3′-side of cytosine and thymine bulges, has been extensively investigated. (Kappen & Goldberg, 1993a; Kappen & Goldberg, 1993b; Stassinopolous & Goldberg, 1995) These studies have examined oligonucleotide sequences containing different types and numbers of bases within the bulge. In this work, the covalent monoadduct produced from the reaction of NCS-Chrom with a bulged DNA substrate has been characterised. (Kappen & Goldberg, 1997) More recently, the roles of bulged DNA in this interaction have been investigated and shown to enhance the binding of the wedge-shaped drug through allowing greater access to the DNA substrate. (Xi *et al.*, 1999)

Saito and co-workers (Nakatani *et al.*, 1999) recently showed that an altromycin analogue selectively alkylates guanines opposite a bulge. These types of drugs intercalate at 5′-GG-3′ steps prior to forming a covalent linkage with the 5′-guanine. A

^{31}P-labelled oligonucleotide containing four possible binding sites (two GGG triplets and two GG doublets) was used. The complementary sequence was modified to include a single thymine bulge opposite three of these binding sites. The results of these studies, and of the molecular modelling of the oligonucleotide, d(CATG$_4$G$_5$TAC/GTAC$_{12}$T$_{13}$C$_{14}$ATG), suggests that the selectivity observed for alkylation at guanines opposite bulge sites arises from stabilisation of the structure through stacking interactions between the ligand and the G$_5$ and T$_{13}$ bases.

The synthesis of drugs designed to bind specifically to duplex DNA containing a single guanine bulge was carried out by Nakatani *et al.* (Nakatani *et al.*, 2000) These naphthyridine derivatives acted as recognition molecules that differentiated between the different types of bulged bases.

Caceres-Cortes and Wang (Caceres-Cortes & Wang, 1996) used nogalamycin (figure 1.7) as a probe for the binding of intercalators to bulged sites in DNA. In this work, duplex DNA containing a single thymine base bulge was utilised, and the resulting complex was investigated using NMR spectroscopy. This study found that upon binding of the ligand, the bulged thymine and opposing base in the duplex are induced by the nogalamycin to stack on the aglycone chromophore, resulting in a frame-shift. Hence, the weaker base pair interaction is perturbed and the overall structure contains 'dangling' 3′-nucleotides.

1.5 Ligand-DNA Interactions

DNA has been viewed as a potential target for drugs designed to alter a variety of biological functions. There is considerable interest in studying the interaction of

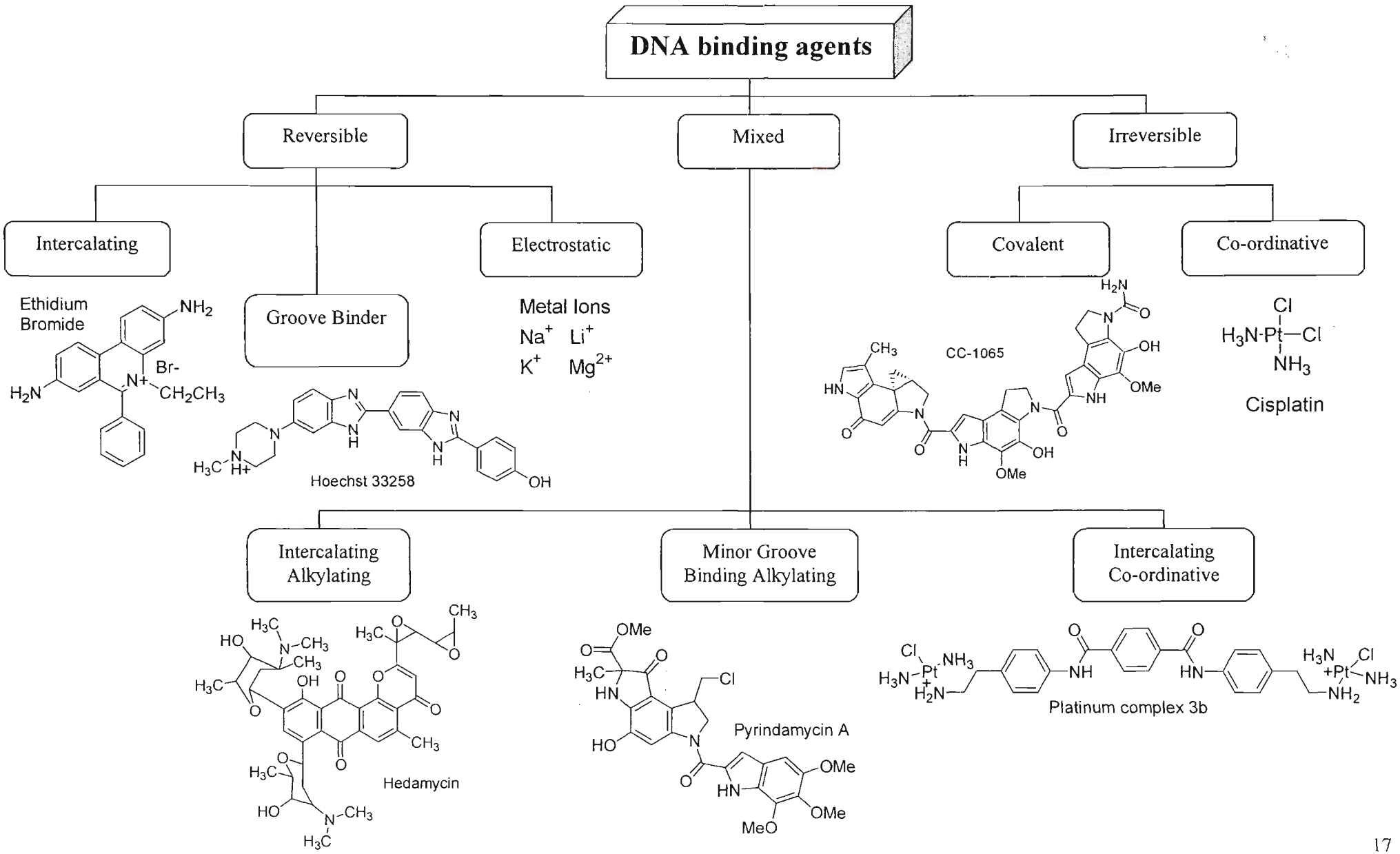
different ligands with DNA, either to understand the role of xenobiotic compounds in carcinogenesis, or to evaluate potential DNA-interactive antitumour drugs. The difficulty in using DNA as a target for drug interactions is that in doing so, there is a risk of disturbing normal cellular function. For this reason, DNA has only been used as a target for serious, life-threatening diseases such as cancer.

The process of designing antitumour drugs or ligands is aided by detailed understanding of their mode of interaction with DNA. Small molecules or ligands can bind to the double helix in a variety of ways as shown in scheme 1.1. The binding may be reversible (or non-covalent), irreversible (or covalent) or be a two-stage process involving an initial non-covalent interaction followed by irreversible binding of the ligand to DNA. Moreover, reversible binding may involve one of several different mechanisms, *i.e.* intercalation or minor groove binding, and irreversible interaction may involve either covalent binding such as alkylation, or co-ordinative metal ion binding.

One of the important features of DNA is the electrostatic interactions that occur between the negatively charged deoxyribose-phosphate backbone and various cations. DNA is stabilised by cations such as Na^+ and Mg^{2+} through the screening of the repulsive interactions of the phosphate anions. (Long, 1996) Interactions may also occur between positively-charged proteins and DNA via groove binding interactions. In this case, the nucleic acid and ligand are bound via weak intermolecular forces, such as hydrogen bonding and Van der Waals forces. These bonds between the ligand and the oligonucleotide usually involve base pairs on the floor of the major and minor grooves. Generally, macromolecules such as proteins interact with the major groove, whilst small molecules interact with the minor groove. Minor groove-binding molecules interact

preferentially with A-T base pairs owing to steric factors. Additionally, the greater negative electrostatic potential of AT-rich regions, when compared to GC-rich regions, may account for the AT-selectivity of specific ligands, which bear a positive charge. (Mountzouris & Hurley, 1996) Examples of minor-groove binding ligands include netropsin, (Fish *et al.*, 1988) distamycin (Boehncke *et al.*, 1991) and Hoechst 33258, (Teng *et al.*, 1988) which are shown in figure 1.6.

Scheme 1.1: Summary of the major modes of ligand-DNA and metal-DNA interactions.



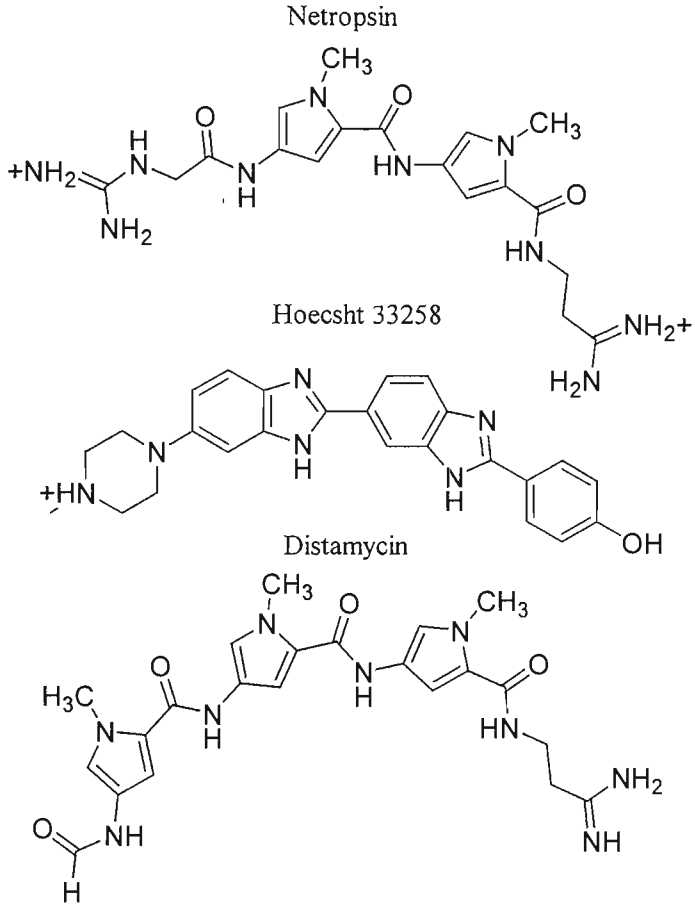


Figure 1.6: Structure of the minor groove binding ligands netropsin, Hoechst 33258 and distamycin.

Intercalation, another reversible non-covalent interaction, occurs when a planar, aromatic molecule slides between the stacked base-pairs of the hydrophobic interior of the helical DNA. (Long, 1996) The classical intercalating model was proposed in the 1960's, (Lerman, 1961) following observations that some DNA associating drugs were seen to increase the viscosity of DNA. (Blackburn & Gait, 1996) Intercalation generally causes structural distortion of the DNA such as partial unwinding and lengthening (up to 3.4 Å), which may lead to changes in the hydrodynamic behaviour of the polymer and thermal stabilisation of the duplex structure. Examples of intercalating agents include daunomycin (Taatjes *et al.*,

1996) and nogalamycin, (Bhuyan & Dietz, 1965; Bhuyan & Reusser, 1970) which are shown in figure 1.7.

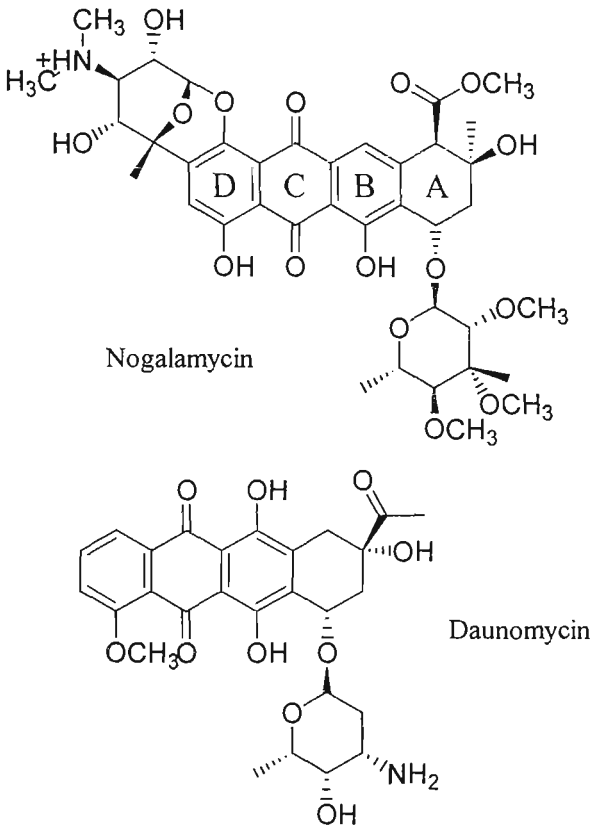


Figure 1.7: Structures of the intercalating agents nogalamycin (Ng) and daunomycin.

Nucleic acids can also be covalently modified through alkylation or metallation reactions. Agents that modify DNA covalently are often mutagenic or carcinogenic and/or may possess antitumour properties. Some carcinogenic compounds are alkylating agents which result in direct modification of the DNA, whilst other carcinogenic compounds are metabolised into alkylating agents in cells, e.g., polyaromatic hydrocarbons (PAHs) are converted to epoxides *in vivo* by the effects of cytochrome P450s. (Stemmler *et al.*, 1994)

Examples of ligands that covalently interact with DNA include nitrogen mustards (Ojwang *et al.*, 1989) and platinum complexes such as cisplatin and its analogues (Murray *et al.*, 1992) as shown in figure 1.8. The co-ordinate binding of metal ions may also result in the modification of DNA.

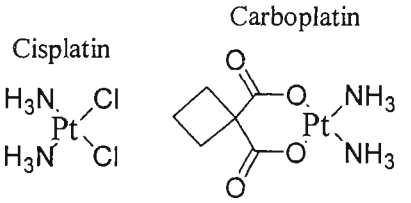


Figure 1.8: Structure of the anti-tumour agent cisplatin and its analogue carboplatin.

1.6 CC-1065 and the Duocarmycins

CC-1065 represents the oldest member of a growing class of potent antitumour antibiotics that also includes the duocarmycins and is one of the most cytotoxic antitumour agents known. (Boger *et al.*, 1990, 1994; Gomi *et al.*, 1992; Boger & Johnson, 1995, 1996) It is produced by a soil culture, *Streptomyces Zelensis.* (Chidester *et al.*, 1981) The interaction of CC-1065 with DNA has been studied extensively. (Hurley *et al.*, 1988, 1990; Boger *et al.*, 1991, 1994) CC-1065 exerts its biological activity through the covalent and irreversible alkylation to DNA in a sequence selective manner. (Boger, 1995; Boger & Sakya, 1992) Its structure, shown in figure 1.9, is characterised by three repeating pyrroloindole subunits, the first of which contains a reactive cyclopropyl function (Hurley *et al.*, 1988) with alkylating potential. (Scahill *et al.*, 1990) (+)-CC-1065 was shown to be efficacious in the treatment of experimental tumours in mice. (Reynolds *et al.*, 1986; Martin *et al.*, 1981) The development of this drug was halted, however, upon the discovery of its delayed toxicity in mice at therapeutic doses. (McGovren *et al.*, 1984) This delayed toxicity was thought to be

a consequence of the large extent of stabilisation by non-covalent binding, which renders the DNA alkylation irreversible. (Boger & Yun, 1993) Synthetic analogues have been prepared which are more effective and do not cause delayed death and therefore show promise as therapeutic agents. (Warpehoski, 1986; Li *et al.*, 1987)

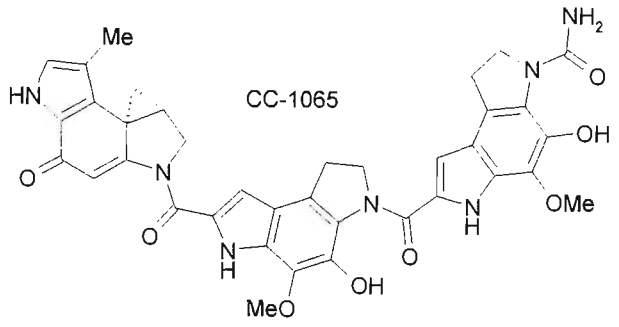


Figure 1.9: Structure of CC-1065.

CC-1065 and the duocarmycins interact with DNA via the minor groove. High selectivity for covalent reaction at adenine N3 within duplex DNA has been observed. (Hurley *et al.*, 1988; Boger, 1995; Boger & Yun, 1994) The stereoelectronically-controlled, adenine N3 addition to the least substituted carbon of the activated cyclopropane occurs within selected AT-rich sites of the minor groove. (Boger *et al.*, 1990; Boger & Yun, 1994; Boger, 1995) The DNA alkylation sites and relative selectivity for CC-1065 and the duocarmycins have been determined through thermally-induced depurination and strand cleavage of labelled DNA, following exposure to the agents as shown in scheme 1.2. (Boger, 1995; Reynolds *et al.*, 1985) The elucidation of alkylation sites is possible as strand cleavage preferentially occurs at the site of covalent alkylation. Since the thermally-induced cleavage of covalently modified DNA only detects adducts which are susceptible to thermal glycosidic bond

cleavage, *i.e.* adenine N3 and/or guanine N3 and N7, the occurrence of additional alkylation reactions would not be detected.

CC-1065 binding to DNA has been shown to occur at a five-base-pair, AT-rich alkylation site. (Boger & Johnson, 1996) Each site of alkylation proved to be an adenine flanked by two 5′-A or T bases. There was also a strong preference for the fourth 5′-base to be an A or T, a weaker preference for the fifth 5′-base to be an A or T and a weak preference for the 3′-base preceding the alkylation site to be a purine. The preference for the 3-base alkylation site followed the order: 5′-AA<u>A</u>-3′ > 5′-TT<u>A</u>-3′ > 5′-TA<u>A</u>-3′ > 5′-AT<u>A</u>-3′. The selectivity of the agent for the AT-rich sites corresponds to the length of the agent and hence the size of the binding site required. (Boger & Johnson, 1996; Boger, 1994; Hurley *et al.*, 1988)

1.7 The Pyrindamycins (Duocarmycins)

Duocarmycins A, C_1 (pyrindamycin B) and C_2 (pyrindamycin A),[1] produced by *Streptomyces sp.* isolated from soil samples, are antitumour antibiotics related to CC-1065. They are active against both Gram-positive and Gram-negative bacteria, and exhibit strong therapeutic effects against drug-sensitive and resistant cells of P388 leukemia in mice. (Ichimura *et al.*, 1988; Ohba *et al.*, 1988; Takahashi *et al.*, 1988; Ishii *et al.*, 1989; Gomi *et al.*, 1992)

[1] The different names for this class of compounds arises as a result of the simultaneous dicovery of them by two different Japanese research teams. In September, 1988, Ichimura *et al.* published their findings on duocarmycin C_1 (Ichimura *et al.*, 1988) and C_2. (Yasuzawa *et al.*, 1988) In October, 1988, Ohba *et al.* (Ohba *et al.*, 1988) published a report on the structure of both pyrindamycins A and B.

The structures of both pyrindamycin A and B, as shown in figure 1.10, were deduced and confirmed by a number of studies which utilised electron ionisation mass spectrometry (EI-MS), (Ichimura *et al.*, 1988) high resolution EI-MS, (Yasuzawa *et al.*, 1988) NMR spectroscopy, (Ichimura *et al.*, 1988; Yasuzawa *et al.*, 1988; Ohba *et al.*, 1988; Takahashi *et al.*, 1988) ultra-violet visible (UV-Vis) and infrared (IR) (Ichimura *et al.*, 1988; Takahashi *et al.*, 1988) and X-ray crystallography. (Ohba *et al.*, 1988) Pyrindamycins A and B possess a pharmacophore similar to that of CC-1065 as shown in figure 1.10.

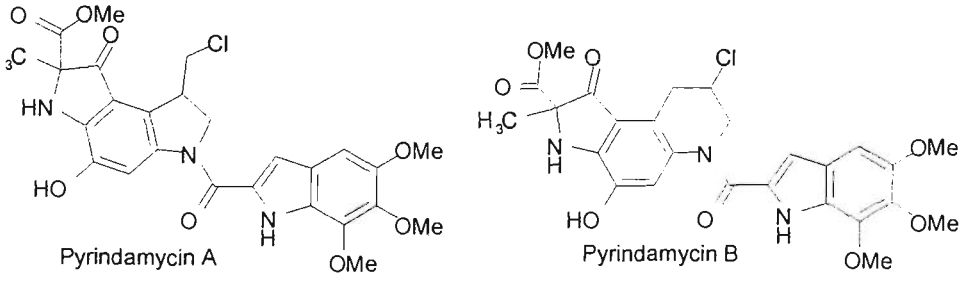


Figure 1.10: Structure of the ligands used in this study pyrindamycins A and B.

In a similar manner to CC-1065, the pyrindamycins (or duocarmycins) exert their biological activity through a sequence selective alkylation of DNA. However, this alkylation is reversible and as such they do not exhibit the same highly toxic effects as does CC-1065. (Gomi *et al.*, 1992; Hurley *et al.*, 1988) The rate, or ease, of reversibility proved to be dependent on the relative reactivity of the agent, the stability of the adduct formed as well as the extent of non-covalent binding interactions. (Boger *et al.*, 1994) The DNA alkylation reaction was also dependent on the stabilising effect of the non-covalent binding interactions. The site of alkylation for the duocarmycin A and the pyrindamycins (A and B) proved to be nearly indistinguishable and proceeds through the 3′-adenine N3 alkylation of

the activated cyclopropane similar to that observed for the (+)-CC-1065 covalent alkylation to DNA. (Boger *et al.*, 1990) It is thought that pyrindamycin A and B are converted to a common agent, duocarmycin A (as shown in figure 1.11), prior to alkylation (although pyrindamycin A may act as an alkylating agent in its own right), (Boger *et al.*, 1990) and hence these ligands and other duocarmycins display identical alkylation properties.

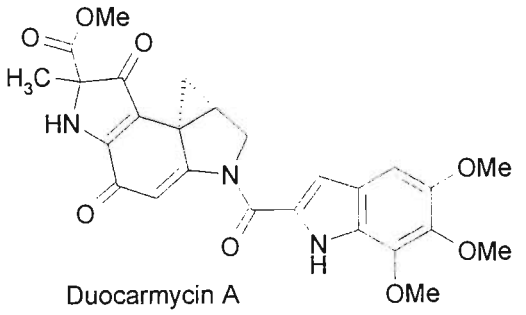


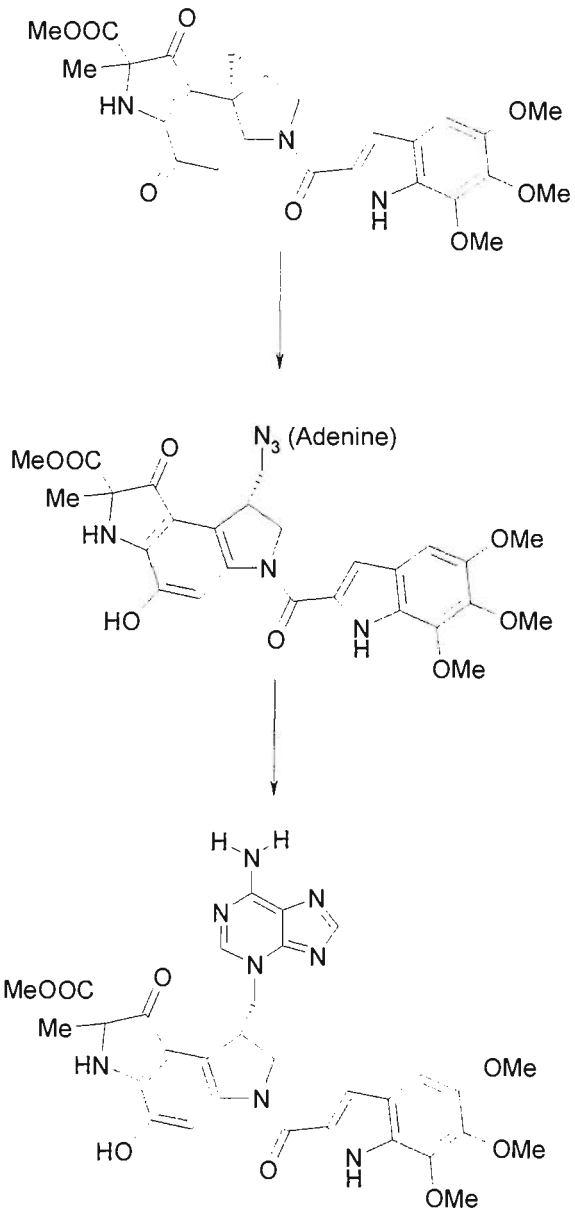
Figure 1.11: Structure of the alkylating agent duocarmycin A.

Figure 1.11 shows the structure of duocarmycin A. The conversion of the pyrindamycins to duocarmycin A would require the cleavage of the halogen *in vitro* in the culture media or *in vivo* in the cells. (Gomi *et al.*, 1992)

The preference for the 3-base sequence follows the same order as described for CC-1065 (section 1.6). A strong preference for the fourth 5′-base to be an A or T versus G or C was also observed. (Boger *et al.*, 1990; Boger, 1995) Thus, it was shown that the sequence selectivity extends in the 5′-direction from the site of modification. The binding of these antibiotics to AT-rich regions of the DNA minor groove was shown to span 3.5 base-pairs, which explains the absolute requirement for the first three base-pairs and the preference for the fourth base-pair. Scheme 1.2 shows the adenine N3 addition to duocarmycin A.

Alkylation has also been found to occur to the N3 of guanine in sequences, which lack the high affinity binding sites. (Mitchell *et al.*, 1993; Boger *et al.*, 1994)

Scheme 1.2: The reaction between the duocarmycin A and DNA.



Boschenok *et al.* (Boschenok, 1997) studied two oligonucleotides containing preferred sites for alkylation by electrospray mass spectrometry, but did not observe the intact adduct for

either sequence. They did, however, observe the pyrindamycin-adenine adduct, a product of the depurination of the oligonucleotide. They hypothesised that the adducts formed were relatively labile and decomposed during ionisation.

1.8 The Anthracyclines

The anthracyclines are a family of DNA-intercalating agents that have found common use as chemotherapeutic agents. The best-known members of this family are daunomycin and doxorubicin (adriamycin), which are used in the treatment of acute leukaemia and solid tumours respectively. (Crooke & Reich, 1980; Acramone & Penco, 1988) Nogalamycin is an antibiotic with a related structure, but is more structurally complex than the other members of the family. (Arora, 1983) Nogalamycin (Ng) was isolated from *Streptomyces nogalator* and has been shown to be active against Gram-positive bacteria and some varieties of experimental tumours (Bhuyan & Dietz, 1965; Bhuyan, 1967; Bhuyan & Reusser, 1970) and it inhibits DNA-directed RNA synthesis *in vivo*. (Li *et al.*, 1979; Ennis, 1981) Nogalamycin consists of a planar aglycone chromophore with bulky sugar substituents at both ends, which gives rise to a dumbbell-shaped molecule. 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. (Bhuyan & Reusser, 1970) Nogalamycin forms a stable intercalation complex (Searle *et al.*, 1988) with slow association and dissociation rates. (Fox & Waring, 1984; Fox *et al.*, 1985) A number of studies have been carried out to show the relationship between the structure of nogalamycin and its analogues and their biological activity. (Gao *et al.*, 1990; Wiley *et al.*, 1982)

Nogalamycin (figure 1.7) is known as a topoisomerase I (TOP I) poison. 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 induces a 22° bend in DNA. (Robinson *et al.,* 1994) It can trap topoisomerase I-cleavable complexes at specific sites on DNA. (Sim *et al.,* 1997) Similar behaviour has also been noted for some minor groove binding ligands (Hoechst 33342 and the terbenzimidazoles). The stimulation of site-specific topoisomerase I-mediated DNA cleavage has been linked to the groove binding properties of these ligands. (Xu *et al.,* 1998) It has been shown that these compounds preferentially bind and stabilise bent DNA domains suggesting a possible correlation between drug-induced bending and TOP I poisoning. (Pilch *et al.,* 1997) The replacement of the DNA sequence at the binding site, with the DNA bending sequence (A_5) also resulted in the stimulation of TOP I mediated DNA cleavage. (Wu & Crothers, 1984) Hence, it was suggested that nogalamycin binding to DNA stimulated TOP I mediated DNA cleavage by inducing a bend in the DNA structure. (Sim *et al.,* 2000)

The binding of nogalamycin to a number of short oligonucleotides, has been studied extensively by X-ray crystallography (Williams *et al.,* 1990; Liaw *et al.,* 1989; Gao *et al.,* 1990; Smith *et al.,* 1995, 1996) and NMR spectroscopy. (Searle *et al.,* 1988; Zhang & Patel, 1990; Searle & Bicknell, 1992; Searle & Lane, 1992; Williams & Searle, 1998, 1999) These studies showed that Ng molecules are intercalated between 5′-TpG (or 5′-CpG) steps in a B-DNA double helix with the nogalose and aminoglucose occupying the minor and

major grooves respectively. More recently, the binding of a variety of drugs to unusual DNA structures has been reported. (Caceres-Cortes & Wang, 1996)

1.9 Outline of this Thesis

In this thesis, the analytical techniques of mass spectrometry and nuclear magnetic resonance (NMR) spectroscopy have been applied to the study of pyrindamycin-oligonucleotide complexes and complexes of nogalamycin with unusual DNA structures. These techniques and their application to the study of modified and unmodified oligonucleotides are reviewed in chapter two. In chapter three, studies of the interaction of the minor groove binding alkylators, the pyrindamycins, with a variety of oligonucleotides by electrospray ionisation tandem mass spectrometry are described with particular emphasis on the sequence selectivity of these agents. The engineering of novel DNA structures that contain nogalamycin intercalation sites is described in chapters four and five. NMR spectroscopy is utilised for the complete structural characterisation of the free and complexed forms of these unusual structures. These studies show that where one technique may be applied to one ligand-DNA interaction another technique may be more applicable to the study of another. Whilst NMR spectroscopy yields detailed structural information, mass spectrometry is a powerful tool for the rapid elucidation of information pertaining to sites of binding, stoichiometry and stability of the interaction.

Bibliography

Acramone, F. and Penco, S. (1988) in *Anthracyclines and anthracenedione-based anticancer agents* (J.W. Lown, Ed) Elsevier, NY

Adams, R.L.P; Knowler, J.T and Leader, D.P. (1992) *The Biochemistry of the Nucleic Acids*, Chapman & Hall, London

Agrawal, S. (1992) *Trends in Biotechnology* **10**,152-158

Arora, S.K. (1983) *J. Am. Chem. Soc.* **105**, 1328-1332

Astbury, W.T. (1947) *Symp. Soc. Exp. Biol.* **1**, 66

Beck, J.L.; Colgrave, M.L.; Ralph, S.R. and Sheil, M.M. (2001) *Mass Spec. Rev.* **20(2)**, 61-87

Bhuyan, B.K. and Dietz, A. (1965) *Antimicrob. Agents Chemother.* 836-844

Bhuyan, B.K. (1967) in *Antibiotics* (D. Gottlieb & P.D. Shaw, Eds), Springer, NY, Vol.1, p. 173-180

Bhuyan, B.K. and Reusser, F. (1970) *Cancer Res.* **30**, 984-989

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

Boehncke, K.; Nonella, M. and Schulten, K. (1991) *Biochemistry* **30**, 5465-5475

Boger, D.L. (1995) *Acc. Chem. Res.* **28**, 20-29

Boger, D.L.; Invergo, B.J.; Coleman, R.S.; Zarrinmayeh, H.; Kitos, P.A.; Collins-Thompson, S.; Leong, T. and McLaughlin, L.W. (1990) *Chem. & Biol. Interact.* **73**, 29-52

Boger, D.L.; Ishizaki, T.; Zarrinmayeh, H.; Kitos, P.A. and Suntornwat, O. (1990) *J. Org. Chem.* **55**, 4499-4502

Boger, D.L. and Johnson, D.S. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 3642-3649

Boger, D.L. and Johnson, D.S. (1996) *Angew. Chem. Int. Ed.* **35**, 1438-1474

Boger, D.L.; Johnson, D.S. and Yun, W. (1994) *J. Am. Chem. Soc.* **116**, 1635-1656

Boger, D.L.; Johnson, D.S.; Yun, W. and Tarby, C.M. (1994) *Bioorg. & Med. Chem.* **2**, 115-135

Boger, D.L.; Munk, S.A. and Zarrinmayeh, H. (1991) *J. Am. Chem. Soc.* **113**, 3980-3983

Boger, D.L. and Sakya, S.M. (1992) *J.Org. Chem.* **57**, 1277-1284

Boger, D.L. and Yun, W. (1993) *J. Am. Chem. Soc.* **115**, 9872-9873

Boger, D.L. and Yun, W. (1994) *J. Am. Chem. Soc.* **116**, 5523-5524

Boschenok, J. (1997) *Mass Spectrometry of DNA Constituents* PhD Thesis, University of Wollongong, Wollongong

Brenner, S.; Jacob, F. and Meselson, M. (1961) *Nature* **190**, 576

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

Caetono-Anolles, G. and Gresshoff, P.M. (1994) *Bio/Technology* **12**, 619-623

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

Cate, J.H.; Hanna, R.L. and Doudna, J.A. (1997) *Nature Struct. Biol.* **4**, 553-558

Chargaff, E. (1955) in *The Nucleic Acids* (E. Chargaff and J.N. Davidson, Eds) Academic Press, New York, Vol. 1, p. 307

Chidester, C.G.; Krueger, W.C.; Mizsak, S.A.; Duchamp, D.J. and Martin, D.G. (1981) *J. Am. Chem. Soc.* **103**, 7629-7635

Chraibi, Z.; Refregiers, M.; Jolles, B. and Laigle, A. (1999) *J. Biomol. Struct. Dyn.* **17**, 539-544

Chraibi, Z.; Refregiers, M.; Jolles, B. and Laigle, A. (2000) *J. Raman Spectrioscop.* **31**, 481-484

Cohen, J.S. (1987) *Trends Biochem. Sci.* **12**, 133

Crooke, S.T. (1992) *Annu. Rev. Pharmacol. Toxicol.* **32**, 329-376

Crooke, S.T. and Reich, S.D. (1980) *Anthracyclines: current status and new developments,* Academic Press, NY

Delepierre, M; Langlois D'Estaintot, B.L; Igolen, J and Roques, B.P. (1986) *Eur. J. Biochem.* **161**, 571-577

Derome, Andrew E., (1995) Modern NMR Techniques for Chemistry Research, vol 6, Tetrahedron Organic Chemistry Series. (J.E. Baldwin & P.D. Magnus, Eds), Pergamon Press, Exeter, UK

Ennis, H.L. (1981) *Antimicrobial Agents & Chemotherapy* **19**, 657-665

Fish, E.L.; Lane, M.J. and Vournakis, J.N. (1988) *Biochemistry* **27**, 6026-6032

Flor, P.; Flanegan, J.B. and Cech, T.R. (1989) *EMBO J.* **8**, 3391-3399

Fox, K.R.; Brassett, C. and Waring, M.J. (1985) *Biochim. Biophys. Acta* **840**, 383-392

Fox, K.R. and Waring, M.J. (1984) *Biochim. Biophys. Acta* **802**, 162-168

Franklin, R.E and Gosling, R.G. (1953) *Nature* **171**, 740

Franklin, R.E and Gosling, R.G. (1953) *Nature* **172**, 156

Gao, Y.G.; Liaw, Y.C.; Robinson, H. and Wang, A.H.J. (1990) *Biochemistry* **29**, 10307-10316

Gomi, K.; Kobayashi, E.; Miyoshi, K.; Ashizawa, T.; Okamoto, A.; Ogawa, T.; Katsumata, S.; Mihara, A.; Okabe, M. and Hirata, T. (1992) *Japan J. Cancer Res.* **83**, 113-120

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

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

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

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

Hirao, I.; Yoshizawa, S. and Miura, K. (1993) *FEBS Letters* **321**, 169-172

Hurley, L.H.; Lee, C.-S.; McGovren, J.P.; Warpehoski, M.A.; Mitchell, M.A.; Kelly, R.C. and Aristoff, P.A. (1988) *Biochemistry* **27**, 3886-3892

Hurley, L.H.; Warpehoski, M.A.; Lee, C.-S.; McGovren, J.P.; Scahill, T.A.; Kelly, R.C.; Mitchell, M.A.; Wicnienski, N.A.; Gebbard, I.; Johnson, P.D. and Bradford, V.S. (1990) *J. Am. Chem. Soc.* **112**, 4633-4649

Ichimura, M.; Muroi, K.; Asano, K.; Kawamoto, I.; Tomita, F.; Morimoto, M. and Nakano, H. (1988) *J. Antibiot.* **41**, 1285-1288

Ishii, S.; Nagasawa, M.; Kariya, Y.; Yamamoto, M.; Inouye, S. and Kondo, S. (1989) *J. Anitbiot.* **42**, 1713-1717

Iverson, P. (1991) *Anti-Cancer Drug Design* **6**, 531-538

Jolles, B.; Refregiers, M. and Laigle, A. (1997) *Nucleic Acids Res.* **25**, 4608-4613

Joshua-Tor, L.; Rabinovich, D.; Hope, H.; Frolow, F.; Appela, E. and Sussman, J.L. (1988) *Nature* **334**, 82-84

Joshua-Tor, L.; Frolow, F.; Appela, E.; Hope, H.; Rabinovich, D. and Sussman, J.L. (1992) *J. Mol. Biol.* **225**, 397-431

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

Kalnik, M.W.; Norman, D.G.; Swann, P.F. and Patel, D.J. (1989) *J. Biol. Chem.* **264**, 3702-3712

Kalnik, M.W.; Norman, D.G.; Li, B.F.; Swann, P.F. and Patel, D.J. (1990) *J. Biol. Chem.* **265**, 636-647

Kappen, L.S. and Goldberg, I.H. (1997) *Biochemistry* **36**, 14861-14867

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

Lerman, L.S. (1961) *J. Mol. Biol.* **3**, 18-30

Li, L.H.; Wallace, T.L.; DeKoning, T.F.; Warpehoski, M.A.; Kelly, R.C.; Prairie, M.D. and Krueger, W.C. (1987) *Investigational New Drugs* **5**, 329-337

Li, L.H.; Kuentzel, S.L.; Murch, L.L.; Pschigoda, L.M. and Krueger, W.C. (1979) *Cancer Res.* **39**, 4816-4822

Liaw, Y.C.; Gao, Y.G.; Robinson, H.; van der Marel, G.A.; van Boom, J.H. and Wang, A.H.J. (1989) *Biochemistry* **28**, 9913-9918

Long, E.C. (1996) *Bioorganic Chemistry: Nucleic Acids* Oxford University Press, New York

Martin, D.G.; Biles, C.; Girpheide, S.A.; Hanka, L.J.; Krueger, W.C.; McGovren, J.P.; Mizsak, S.A.; Neil, G.L.; Stewart, J.C. and Visser, J. (1981) *J. Antibiot.* **34**, 1119

McGovren, J.P.; Clarke, G.L.; Pratt, E.A. and DeKoning, T.F. (1984) *J. Antibiot.* **37**, 63-70

Miescher, F. (1879) *Die histochemischen und physiologischen Arbeiten,* Leipzig

Miller, M.; Harritson, R.W.; Woldaver, A.; Appela, E. and Sussman, J.L. (1988) *Nature* **334**, 85-86

Mitchell, M.A.; Weiland, K.L.; Aristoff, P.A.; Johnson, P.D. and Dooley, T.P. (1993) *Chem. Res. Toxicol.* **6**, 421-424

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

Mountzouris, J.A. and Hurley, L.H. (1996) *Bioorganic Chemistry: Nucleic Acids* Oxford University Press, New York

Murray, V.; Motyka, H.; England, P.R.; Wickham, G.; Lee, H.H.; Denny, W.A. and McFadyen, W.D. (1992) *Biochemistry* **31**, 11812-11817

Nakatani, K.; Okamoto, A. and Saito, I. (1999) *Angew. Chem. Int. Ed.* **38**, 3378-3381

Nakatani, K.; Sando, S. and Saito, I. (2000) *J. Am. Chem. Soc.* **122**, 2172-2177

Neidle, S; Pearl, L.H and Skelly, J.V. (1987) *Biochem. J.* **243**, 1-13

Nelson, D. L. and Cox, M. M. (2000) *Lehninger Principles of Biochemistry (3rd ed.)* p. Worth Publishers, NY, p. 339-343

Nelson, J.W. and Tinoco Jr., I. (1985) *Biochemistry* **24**, 64616-6421

Nikonowicz, E.; Roongta, V.; Jones, C.R. and Gorenstein, D.G. (1989) *Biochemistry* **28**, 8714-8725

Nikonowicz, E.; Meadows, R.P. and Gorenstein, D.G. (1990) *Biochemistry* **29**, 4193-4204

Ohba, K.; Watabe, H.; Sasaki, T.; Takeuchi, Y.; Kodama, Y.; Nakazawa, T.; Yamamoto, H.; Shomura, T.; Sezaki, M. and Kondo, S. (1988) *J. Antibiot.* **41**, 1515-1519

Ojwang, J.O.; Grueneberg, D.A. and Loechler, E.L. (1989) *Cancer Res.* **49**, 6529-6537

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

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

Pauling, L. and Corey, R.B. (1953) *Proc. Natl. Acad. Sci. USA*, **39**, 84

Peattie, D.A.; Douthwaite, S.; Garrett, R.A. and Noller, H.F. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 7331-7335

Pilch, D.S.; Xu, Z.; Sun, Q.; LaVoie, E.J.; Liu, L.F. and Breslauer, K.J. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 13565-13570

Reynolds, V.L.; McGovren, J.P. and Hurley, L.H. (1986) *J. Antibiot.* **39**, 319-334

Reynolds, V.L., Molineux, I.J.; Kaplan, D.J.; Swenson, D.H. and Hurley, L.H. (1985) *Biochemistry* **24**, 6228-6237

Rice, J.A. and Crothers, D.M. (1989) *Biochemistry* **28**, 4512-4516

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

Romaniuk, P.J.; Lowary, P.; Wu, H.-N.; Stormo, G. and Uhlenbeck, O.C. (1987) *Biochemistry* **26**, 1563-1568

Scahill, T.A.; Jensen, R.M.; Swenson, D.H.; Hatzenbuhler, N.T.; Petzold, G.; Wierenga, W. and Brahme, N.D. (1990) *Biochemistry* **29**, 2852-2860

Searle, M.S. and Bicknell, W. (1992) *Eur. J. Biochem.* **205**, 45-58

Searle, M.S.; Hall, J.G.; Denny, W.A.; Wakelin, L.P.G. (1988) *Biochemistry* **27**, 4340-4349

Searle, M.S. and Lane, A.N. (1992) *FEBS Letters* **297**, 292-296

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

Sim, S.P.; Pilch, D.S. and Liu, L.F. (2000) *Biochemistry* **39**, 9928-9934

Smith, C.K.; Brannigan, J.A. and Moore, M.H. (1996) *J. Mol. Biol.* **263**, 237-258

Smith, C.K.; Davis, G.J.; Dudson, E.J. and Moore, M.H. (1995) *Biochemistry* **34**, 415-425

Stemmler, E.A.; Buchanan, M.V.; Hurst, G.B. and Hettich, R.L. (1994) *Anal. Chem.* **66**, 1274-1285

Streisinger, G.; Okada, Y.; Emrich, J.; Newton, J.; Tsugita, A.; Terzhagi, E. and Innouye, M. (1966) *Cold Spring Harbor Symp. Quantum Biol.* **31**, 77-84

Taatjes, D.J.; Gaudiano, G.; Resing, K. and Koch, T.H. (1996) *J. Med. Chem.* **39**, 4135-4138

Takahashi, I.; Takahashi, K.-I.; Ichimura, M.; Morimoto, M.; Asano, K.; Kawamoto, I.; Tomita, F. and Nakano, H. (1988) *J. Antibiot.* **41**, 1915-1917

Tang, J.Y.; Temsamani, J. and Agrawal, S. (1993) *Nucleic Acids Res.* **21**, 2729-2735

Teng, M.; Usman, N.; Frederick, C.A. and Wang, A.-H. (1988) *J. Am. Chem. Soc.* 2671-2690

Vallone, P.M.; Paner, T.M.; Hilario, J.; Lane, M.J.; Faldasz, B.D. and Benight, A.S. (1999) *Biopolymers* **50**, 425-442

Varani, G. (1995) *Annu. Rev. Biophys. Biomol. Struct.* **24**, 379-404

Warpehoski, M.A. (1986) *Tetrahedron Lett.* **35**, 4103-4106

Watson, J.D and Crick, F.H.C. (1953) *Nature* **171**, 737

White, S.A. and Draper, D.E. (1987) *Nucleic Acids Res.* **15**, 4049-4064

Wiley, P.F.; Elrod, D.W.; Houser, D.J. and Richard, F.A. (1982) *J. Med. Chem.* **25**, 560-567

Wilkins, M.F.H.; Stokes, A.R and Wilson, H.R. (1953) *Nature* **171**, 738

Williams, H. E. L. (2000) *PhD thesis*, School of Chemistry, University of Nottingham, UK.

Williams, H.E.L. and Searle, M.S. (1998) *J. Chem. Soc, Perkin Trans. I*, **1**, 3-5

Williams, H.E.L. and Searle, M.S. (1999) *J. Mol. Biol.* **290**, 699-716

Williams, L.D.; Egli, M.; Gao, Q.; Bash, P.; van der Marel, G.A.; van Boom, J.H.; Rich, A. and Frederick, C.A. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 2225-2229

Woodson, S.A. and Crothers, D.M. (1987) *Biochemistry* **26**, 904-912

Woodson, S.A. and Crothers, D.M. (1988a) *Biochemistry* **27**, 436-445

Woodson, S.A. and Crothers, D.M. (1988b) *Biochemistry* **27**, 3130-3141

Woodson, S.A. and Crothers, D.M. (1988c) *Biochemistry* **27**, 8904-8914

Wu, H.M. and Crothers, D.M. (1984) *Nature* **308**, 509-513

Xi, Z.; Mao, Q.K. and Goldberg, I.H. (1999) *Biochemistry* **38**, 4342-4354

Xu, Z.; Li, T.K.; Kim, J.S.; LaVoie, E.J.; Breslauer, K.J.; Liu, L.F. and Pilch, D.S. *Biochemistry* (1998) **37**, 3558-3566

Yasuzawa, T.; Iida, T.; Muroi, K.; Ichimura, M.; Takahashi, K. and Sano, H. (1988) *Chem & Pharm. Bull.* **36**, 3728-3731

Yoshizawa, S.; Kawai, G.; Watanabe, K.; Miura, K. and Hirao, I. (1997) *Biochemistry* **36**, 4761-4767

Yoshizawa, S.; Ueda, T.; Ishido, Y.; Miura, K.; Watanabe, K. and Hirao, I. (1994) *Nucleic Acids Res.* **22**, 2217-2221

Zacharias, M. and Sklenar, H. (1997) *Biophy. J.* **73**, 2990-3003

Zhang, X. and Patel, D.J. (1990) *Biochemistry* **29**, 9451-9466

CHAPTER TWO: *Instrumental Methods for Investigating Ligand-DNA Complexes*

2.0 Introduction

The study of the structure, binding specificity and dynamics of ligand-DNA complexes is of great importance in the understanding of the nature of the interactions between the ligand (or drug) and nucleic acids. Two of the most extensively used methods for the structural characterisation of DNA are nuclear magnetic resonance (NMR) spectroscopy and X-ray crystallography. These techniques can yield detailed structural information for nucleic acids and DNA-ligand complexes. They have some fundamental limitations, however, in that they are time consuming and require either large amounts of sample (NMR) (Searle, 1992; Krugh, 1994) or that the sample be crystalline (X-ray crystallography). (Kennard & Salisbury, 1993)

Conventionally, the sequencing of DNA has been achieved by molecular biology techniques, *i.e.* the Sanger method, (Sanger *et al.,* 1977) or by chemical means, *i.e.* the Maxam-Gilbert method, (Maxam & Gilbert, 1977, 1980) technologies. (Limbach, 1996) Both methods result in the production of a series of oligonucleotides, which differ in length and have a chemical or radio-labelled end. The Sanger method uses dideoxynucleotides as chain-terminating inhibitors of DNA polymerases, whilst the Maxam-Gilbert method relies upon base-specific reactions. Automation and technological advances, such as the development of the polymerase chain reaction (PCR), have improved the sensitivity and efficiency of the Sanger-based technique for determining the primary sequence of DNA. Molecular biology techniques play an important role in the determination of the sequence specificity of ligand-DNA interactions. (Murray, 1999) The combination of gel based

sequencing with radioactive labelling, (D'Andrea & Haseltine, 1978; Haseltine *et al.*, 1980) linear amplification, (Ponti *et al.*, 1991) PCR (Pfeifer *et al.*, 1989; Mueller & Wold, 1989; Grimaldi *et al.*, 1994) and/or footprinting (Portugal, 1989) have enabled the sequence specificity of a range of different DNA-interactive agents to be elucidated, however, these techniques may be quite time consuming. Mass spectrometry, on the other hand, may be used to readily gain sequence information concerning modified DNA together with limited structural information concerning ligand-DNA complexes.

In recent years, the development of two new ionisation methods for mass spectrometry, namely electrospray ionisation (ESI) and matrix-assisted laser desorption ionisation (MALDI), has enabled for the first time, the analysis of large segments of DNA by mass spectrometry. Electrospray ionisation mass spectrometry does not provide direct structural information, unlike NMR and X-ray crystallography, however, it does provide important information concerning the stoichiometry of ligand-DNA complexes and sites of binding and it does so in far less time from much less material. (Loo, 1997)

There have been relatively few studies of ligand-DNA interactions by mass spectrometry, so that the general utility of the technique has yet to be fully exploited (more below). Hence, this thesis is concerned with two aspects of the analysis of ligand-DNA interactions. First, exploring the utility of ESI-MS for the study of relatively labile ligand-DNA complexes. The second aspect of the study uses NMR spectroscopy to derive detailed structural information on novel DNA structures, and their drug complexes.

Investigations of this nature require a multi-disciplinary approach including the evaluation of the preferred binding sites, thermodynamic studies to examine complex stability and high-resolution structural studies for probing the molecular basis for sequence recognition. The current studies have utilised electrospray ionisation mass spectrometry (ESI-MS) for the determination of the sequence selectivity, in combination with a wide range of biophysical methods (UV-Vis and thermal denaturation) and the high-resolution structural analysis by NMR spectroscopy.

2.1 Mass Spectrometry Instrumentation

Mass spectrometry is a technique used to determine the molecular mass of chemical and biological molecules, by measurement of the mass-to-charge ratio (*m/z*) of gaseous ions derived from the molecule of interest. A mass spectrometer is usually comprised of an ionisation source, a mass analyser and a detector, along with a computer system for data storage and processing.

Many different types of samples may be analysed using mass spectrometry. These include low molecular weight organic and inorganic compounds (Ahrer & Buchberger, 1999), sugars (Reinhold & Carr, 1983), peptides (Biemann & Martin, 1987), proteins (Biemann & Martin, 1987; Beavis & Chait, 1990) and oligonucleotides (Sindona *et al.*, 1982; Neri *et al.*, 1983). Recently, mass spectrometry has found extensive use for the characterisation of both naturally-occurring DNA (Krahmer *et al.*, 2000) and synthetic oligonucleotides, (Reddy *et al.*, 1994) as well as the products resulting from covalent binding of carcinogens and anti-tumour drugs to DNA. (Chiarelli & Lay Jr., 1992)

2.1.1 Ionisation Methods

The method of ionisation for mass spectrometry depends upon the volatility and thermal stability of the molecule to be analysed. Each ionisation technique is usually suitable for different types of compounds and produces gas phase ions via different chemical and physical processes.

Electron ionisation, or impact, (EI) and chemical ionisation (CI) are the well-established techniques for MS analysis of volatile compounds. Over the last 20 years much of the development in ionisation techniques has been aimed at ionising thermally labile, involatile compounds. These so called 'soft' ionisation techniques include fast atom bombardment (FAB), (Barber *et al.*, 1981) plasma desorption (PD), (Cotter, 1982) secondary ion mass spectrometry (SIMS), (Sindona, 1989) matrix-assisted laser desorption/ionisation (MALDI), (Karas *et al.*, 1991) field desorption (FD), (Sindona, 1989) electrospray ionisation (ESI) (Smith *et al.*, 1990) and thermospray (TS). (Schmelzeisen-Redeker *et al.*, 1989)

These earlier ionisation techniques (FAB, PD, FD, EI and CI) have been largely superseded by electrospray ionisation mass spectrometry (ESI-MS) and matrix-assisted laser desorption/ionisation mass spectrometry (MALDI-MS). Together these techniques, have revolutionised the analysis of biomolecules by allowing the routine characterisation of proteins up to 150kDa (Siegel *et al.*, 1997) and nucleic acids up to 500 base pairs (in the case of MALDI-MS) (Tang *et al.*, 1994).

2.1.2 Matrix-Assisted Laser Desorption/Ionisation (MALDI)

Developed in 1988 by Karas and Hillenkamp, matrix-assisted laser desorption ionisation (MALDI) was based on older laser desorption techniques and studies of the desorption (ionisation) of small organic molecules, mainly amino acids and dipeptides. (Karas *et al.*, 1987, 1991) In the MALDI technique, the matrix material, generally a low molecular weight organic acid present in large excess, minimises radiation damage to the sample and acts as a 'solvent' for the analyte molecules. (Chapman, 1996) The sample is irradiated with a high intensity pulsed laser beam which is tuned to the wavelength at which the matrix has a high extinction coefficient. The matrix also acts as an energy transfer medium. (Karas *et al.*, 1987) Mass analysis is most often accomplished by time-of-flight (TOF) mass spectrometry (more below). The principles of the MALDI-TOF are shown in figure 2.1.

MALDI has overcome the limitations of 'older' laser desorption/ionisation techniques, namely low-mass range and sensitivity, and enables routine analysis of proteins in the mass range 10 to 100kDa with only picomole amounts of sample required. (Karas *et al.*, 1991) MALDI has allowed the structural characterisation of modified nucleic acids and has been used, for example, to enable a better understanding of biological abnormalities such as those that occur with certain forms of cancer. (Hettich & Buchanan, 1991) A recent major improvement in MALDI-MS has been the implementation of delayed extraction (DE) or time-lag focussing which has dramatically improved the resolution of the technique and overcome one of the major difficulties for nucleic acid analysis, namely base fragmentation at high laser power, by enabling analysis at lower laser intensities. (Vestal *et al.*, 1995)

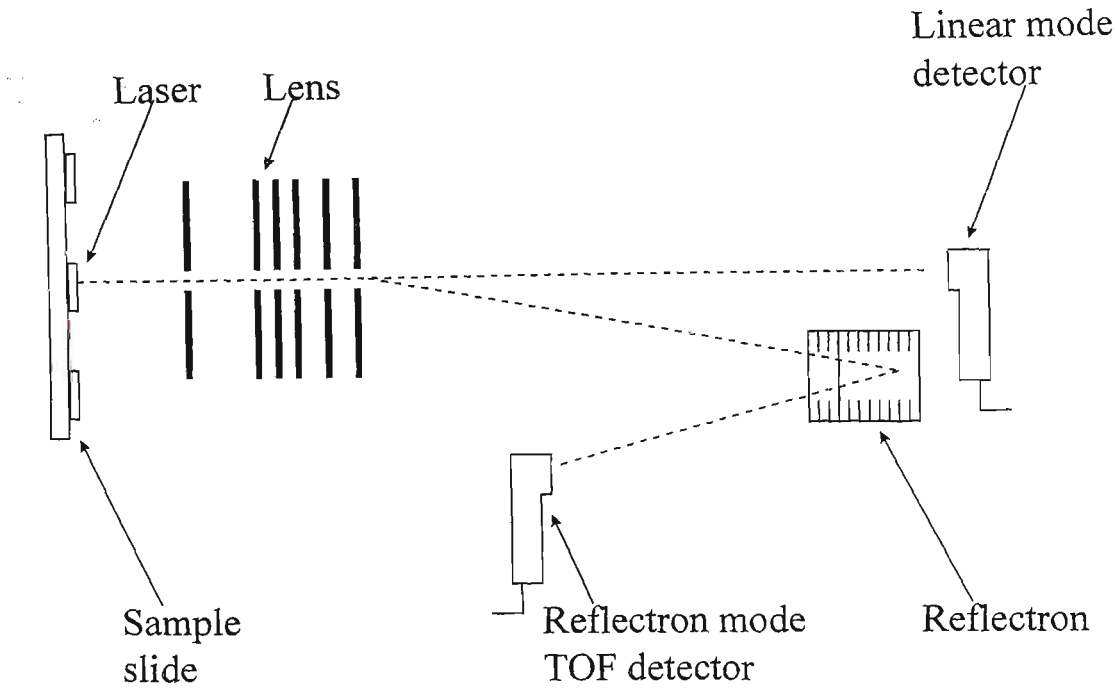


Figure 2.1: Schematic of the most common configuration used with MALDI ion sources. The reflectron improves the energy resolution for analysis of ions below *m/z* 3000.

2.1.3 Electrospray Ionisation (ESI)

Electrospray ionisation (ESI) was developed in 1968 by Dole and coworkers. (Dole *et al.*, 1968) This pioneering work showed that macroions of definite mass-charge states could be produced in the gas phase by electrospray ionisation. The ensuing decade saw little ESI development, however, because of the difficulty of analysing the masses of these large ions. Fenn and coworkers continued this work and in 1984 published a paper showing the application of ESI to large organic molecules that were too complex, too fragile or too involatile for the conventional ionisation methods. (Yamashita & Fenn, 1984) Electrospray, however, has only been in widespread use since 1988 when it was shown to be capable of generating intact ions of large biomolecules. (Mann & Fenn, 1992)

A significant advantage of ESI-MS is that a distribution of multiply charged ions is normally produced when large molecules are analysed. Hence, the *m/z* range of the mass analyser need not be large because, as a result of extensive multiple charging, ions generally fall within the range of *m/z* 1000-4000 (Mann, 1990; Jardine, 1990). Multiple charging also increases the accuracy of molecular weight determinations (to 0.01% or better) since the mass of the sample is determined from more than one peak. (Deroussent *et al.*, 1995) Electrospray also had advantages of high sensitivity (low picomole range) and generally, little fragmentation is observed since the ionisation does not impart a large amount of energy to the ions. Owing to the gentle nature of the electrospray ionisation process, non-covalent interactions occurring in solution may also be detected. (Smith & Light-Wahl, 1993) Examples of non-covalent biomolecular complexes studied to date include enzyme-substrate, (Ganem & Henion, 1991) protein-ligand, (Schwartz *et al.*, 1995; Ganem & Henion, 1991; Jaquinod *et al.*, 1993; Huang *et al.*, 1993; Przybylski *et al.*, 1995; Schwartz *et al.*, 1994) DNA-duplex (Light-Wahl *et al.*, 1993) and DNA-ligand complexes. (Gale *et al.*, 1994; Gao *et al.*, 1996; Gale *et al.*, 1995; Gabelica *et al.*, 1999, 2000; Kapur *et al.*, 1999)

During the electrospray ionisation process, the sample solution is pumped through a metal capillary held at a potential of several kilovolts relative to a counter electrode as shown in in figure 2.2. This results in charging of the liquid as it is sprayed from the capillary resulting in a fine mist of charged droplets. Evaporation is assisted by a flow of warm bath gas (usually nitrogen). Evaporation of the solvent increases the charge density on the droplet surface until electrostatic repulsion is greater than surface tension at which point the droplets explode (Coulombic explosion) as is shown in figure 2.3. If the sample solution is

sufficiently dilute, a sequence of such explosions results in droplets with such small size that each would contain one analyte molecule. (Mann & Fenn, 1992)

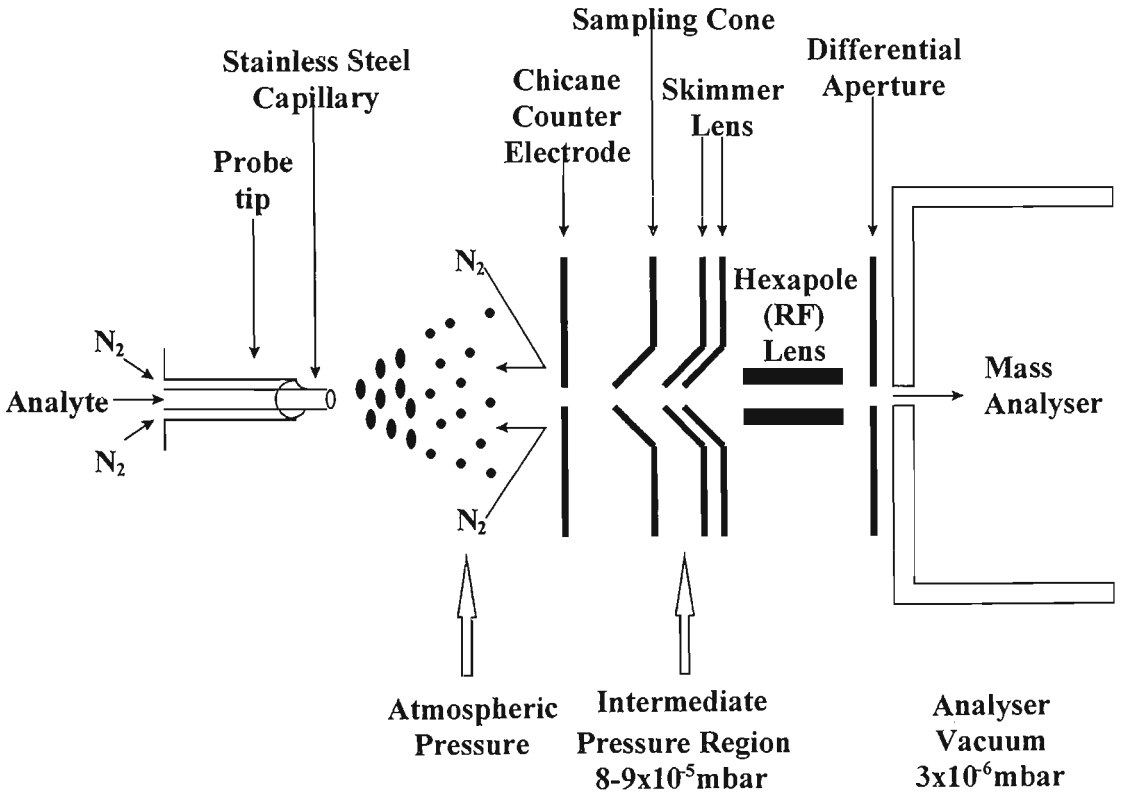


Figure 2.2: Schematic showing a configuration of the electrospray ionisation source in the triple quadrupole instrument used in this study.

The resulting ions from the electrospray ionisation process pass through a skimmer or orifice[2] where the analyte ions are separated from the drying gas and solvent vapour and enter the mass analyser.

[2] More recent electrospray instruments employ a variety of different methods to separate the ionisation source (at atmospheric pressure) from the low vacuum of the analyser. These include orthogonal ionisation, heated nebulisers and the Z-spray configuration used in the Micromass Q-TOF 2 that was employed for some of the later experiments in these studies.

An alternative mechanism for the ionisation process has been proposed by Iribarne and Thompson. (Iribarne & Thomson, 1976) The mechanism is similar to that described above to the point of producing a fine spray of charged droplets. The workers proposed that the droplets evaporate in the bath gas, and as they decrease in size, the charge density increases. The decrease in radius of curvature results in electric fields sufficiently strong to result in the desorption of the analyte ions before the droplets explode. Definitive experimental evidence to support either mechanism has not been forthcoming, however, in general the charged residue mechanism is the more frequently-cited mechanism.

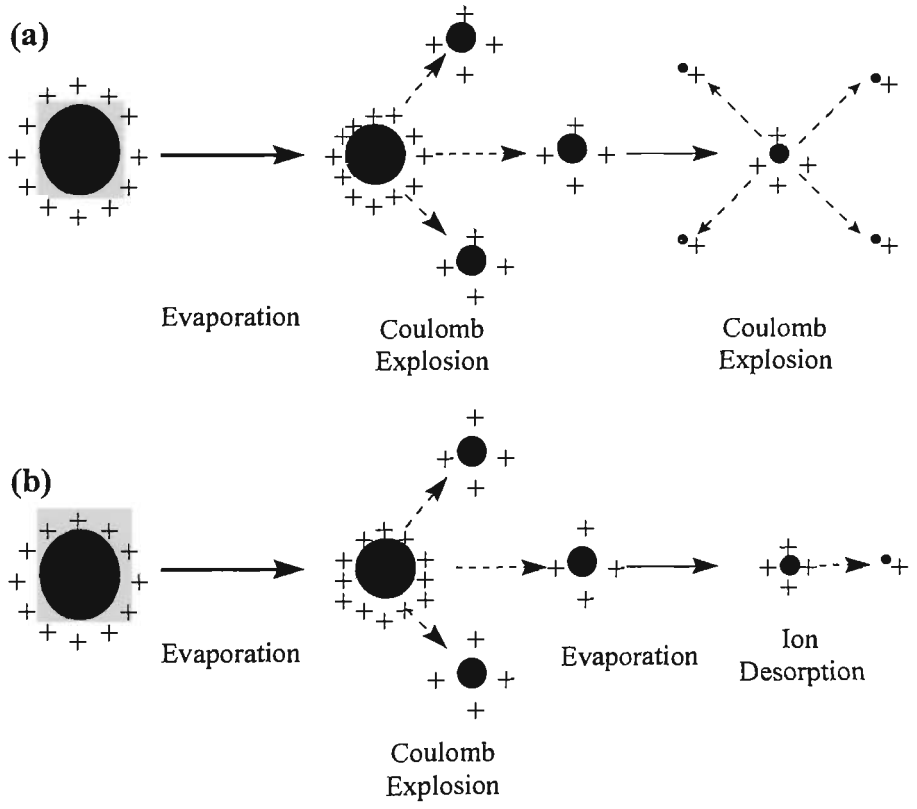


Figure 2.3: Mechanisms of electrospray ionisation: (a) Charged residue model; (b) Ion evaporation model.

As noted above, electrospray ionisation may result in a distribution of molecular ion species with different number of charges being observed in the mass spectrum. The molecular

weight can be determined very accurately from these by solving two simultaneous equations. These are given below for the analysis of negative ions formed by deprotonation.

$$m_2 = \frac{M - nH}{n} \quad \ldots(1)$$

$$m_1 = \frac{M - (n+1)H}{(n+1)} \quad \ldots(2)$$

In equations 1 and 2, m_1 and m_2 are the *m/z* values for two peaks differing by n and (n+1) protons respectively, M is the molecular weight of the species, n is the number of charges and H is the mass of a proton. In most cases, commercial electrospray mass spectrometers have software that facilitates molecular mass calculations based on this principle.

2.1.4 Mass Analysers

The function of the mass analyser is to separate ions according to their mass-to-charge ratio. Electrospray ionisation mass spectrometry has been coupled with quadrupole, time-of-flight, (Verentchikov *et al.*, 1994) quadrupole ion trap, (McLuckey *et al.*, 1988) magnetic sector (Meng *et al.*, 1990) and Fourier Transform (FT) (Buchanan & Hettich, 1993) mass analysers. More recently, hybrid analysers such as quadrupole-time-of-flight (Q-TOF) (Morris *et al.*, 1996) have been effectively coupled with ESI sources. The mass analyser used for the majority of experiments in this study was a quadrupole mass analyser.

2.1.5 Tandem Mass Spectrometry (MS/MS)

Tandem MS or mass spectrometry/mass spectrometry (MS/MS) is very useful for obtaining structural information on a variety of compounds and also the identification of compounds in complex mixtures. Tandem mass spectrometry involves two stages of mass analysis with the stages separated in space (triple quadrupoles and hybrid instruments) or in time (quadrupole ion traps or FT-MS instruments). Figure 2.4 shows a schematic of a tandem mass spectrometry involving spatial separation. The first stage is the mass selection of the precursor ion from a mixture of ions formed in the ion source. Any other ions present, *i.e.* non-selected *m/z* values are filtered out. The precursor ion then fragments to a series of product ions and neutral fragments (in the case of singly charged precursors), *i.e.*

$$M_p^+ \rightarrow M_d^+ + M_n \qquad \ldots(3)$$

where M_p^+ is the precursor ion mass, M_d^+ is the product ion mass and M_n is the neutral fragment mass.

Multiply charged precursors may also fragment to yield charged product ions, *i.e.*

$$M_p^{n+} \rightarrow M_{d1}^{(n-x)+} + M_{d2}^{x+} \qquad \ldots(4)$$

Fragmentation of the precursor ion is achieved by collision-induced dissociation (CID) in a reaction region, or collision cell, containing an inert gas such as argon. In the second stage of mass analysis, the mass spectrum of the product ions, which result from the fragmentation of the precursor ion is recorded.

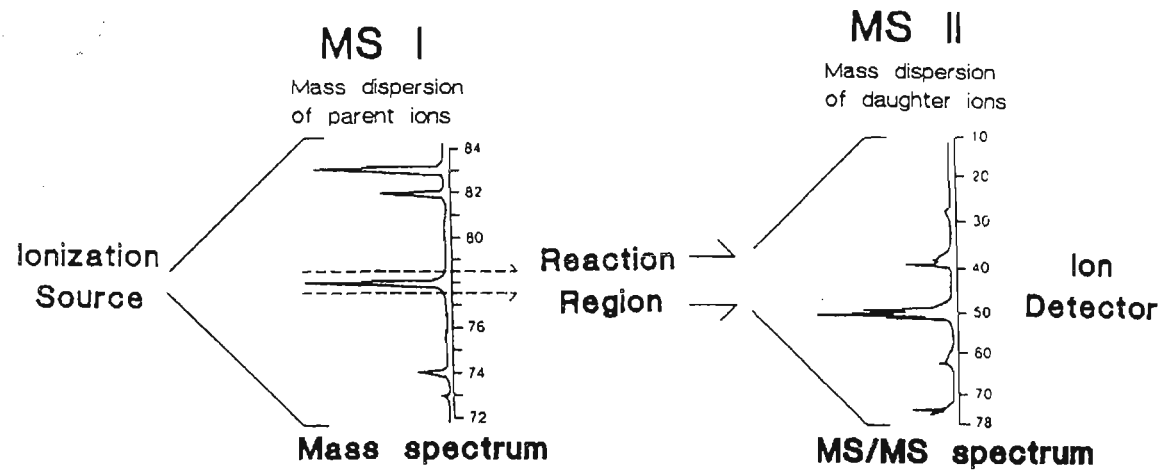


Figure 2.4: Tandem Mass Spectrometry. (Busch *et al.*, 1988)

2.1.6 ESI-MS of Oligonucleotides

Mass spectrometric techniques have been used extensively for the characterisation of nucleic acid constituents, such as nucleosides and nucleotides. (Crain, 1990) Electron ionisation, (EI) and chemical ionisation (CI) mass spectrometry have been utilised for the study of small species, such as nucleosides and bases. (Crain, 1990; Wilson & McCloskey, 1975) Field desorption also provided information on the structure of nucleosides and nucleotides, but is technically very difficult to implement. (Norwood & Vouros, 1994) Plasma desorption (PD) (McNeal & MacFarlane, 1981) and fast-atom-bombardment (FAB) (Crain, 1990; Grotjahn & Taylor, 1985) were used to observe larger species. PD allowed the observation of an oligonucleotide up to 12 bases in length, (McNeal & MacFarlane, 1981) whilst FAB was used to analyse up to 13-base oligonucleotides. (Grotjahn *et al.*, 1988) These were the largest species analysed by mass spectrometry prior to the advent of electrospray ionisation (ESI) and matrix-assisted laser desorption ionisation (MALDI).

The introduction of electrospray ionisation enabled mass spectrometry to be used for the routine analysis of oligonucleotides, which had previously proven very difficult to analyse by those older MS techniques. Covey *et al.* (Covey *et al.*, 1988) were amongst the first to examine the application of ESI-MS to oligonucleotides up to 14 bases in length. The very small amounts of sample required (picomole) gave rise to the rapid ascendancy of ESI for oligonucleotide analysis. The application of MALDI and ESI to nucleic acid analysis has been the subject of comprehensive reviews. (McCloskey & Crain, 1992, 1998; Limbach *et al.*, 1995; Nordhoff *et al.*, 1996; Miketova & Schram, 1997) High-resolution tandem mass spectrometry has been utilised extensivley for the sequencing of oligonucleotides. (Little *et al.*, 1994; Owens *et al.*, 1998; Lichtenwalter *et al.*, 2000) Electrospray mass spectrometry has also found applications for the analysis of modified nucleotides (Reinhold & Carr, 1983; Reddy *et al.*, 1994; Chapman, 1996; Tito, 1998; Reddy & Iden, 1993; Potier *et al.*, 1994; Kowalak *et al.*, 1993; Ni *et al.*, 1998; Wu & Aboleneen, 2000, 2001), intact plasmid DNA (Cheng *et al.*, 1996; Lichtenwalter *et al.*, 2000) and non-covalent ligand-DNA complexes (Gale *et al.*, 1994; Loo, 1997; Smith *et al.*, 1997; Beck *et al.*, 2001).

Routine analysis of oligonucleotides by ESI-MS, however, is more difficult than for proteins owing to the highly labile nature of the oligonucleotides, which makes them prone to fragmentation. The high affinity of the polyanionic backbone for alkali metals (especially sodium) also causes problems. Since the presence of sodium both reduces ionisation efficiency and limits mass accuracy owing to the presence of multiple species with different numbers of sodium ions. Methods commonly employed for desalting oligonucleotides include ethanol precipitation (Potier *et al.*, 1994) and the use of a variety of base additives. (Connolly *et al.*, 1984) In the work described in this thesis, the problems

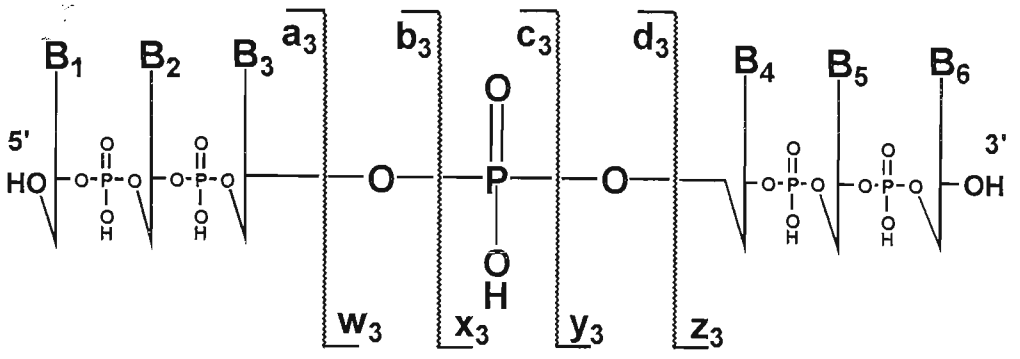
with salt adduction were overcome by the use of ammonium acetate in the HPLC purification of the oligonucleotides, which results in ammonium ions replacing any adventitious sodium ions.

ESI-MS has also been applied to study non-covalent biomolecular complexes. (Light-Wahl *et al.*, 1993) McCloskey and co-workers observed non-covalent interactions including duplexes of an 8-mer, 20-mer and a tetraplex, which was observed in the presence of Na^+ ions. (Limbach *et al.*, 1995) These studies have been extended by Smith *et al.* to encompass specific metal ion binding to oligonucleotides (Wu *et al.*, 1995). The early work by Gale and co-workers on the binding and stoichiometry of the distamycin A-DNA interaction explored the optimisation of the sample conditions required for the detection of these non-covalent complexes (Gale *et al.*, 1994). These studies were extended to include the minor groove binding ligands pentamidine and Hoecsht 33258 (Gale & Smith, 1995) and a number of other studies examining the non-covalent binding of ligands to DNA have been carried out (Triolo *et al.*, 1997; Kapur *et al.*, 1999; Wan *et al.*, 2000a, 2000b; Gabelica *et al.*, 1999, 2000; Gao *et al.*, 1996). The binding of the intercalators actinomycin D (Hsieh *et al.*, 1994; Wan *et al.*, 2000a, 2000b), daunomycin (Triolo *et al.*, 1997; Kapur *et al.*, 1999) and nogalamycin (Kapur *et al.*, 1999) have been investigated in terms of their binding affinity through competition experiments. The relative binding strengths of a variety of minor groove binding ligands has been explored through the examination of the dissociation profiles (Wan *et al.*, 2000b). Gabelica and co-workers found good correlation between the gas-phase dissociation and solution data resulting from the thermal denaturation of the ligand-DNA duplexes of Hoechst 33258 and netropsin (Gabelica *et al.*, 2000).

2.1.7 ESI-MS/MS of Oligonucleotides

Tandem mass spectrometry combined with ESI, has potential for the characterisation of oligonucleotides (Doctycz *et al.*, 1994) and modified oligonucleotides (Barry *et al.*, 1995) as well as covalent ligand-oligonucleotide complexes. (Wickham *et al.*, 1995a, 1995b; Iannitti-Tito *et al.*, 2000)

The fundamentals of oligonucleotide decompositions were first studied by McLuckey and co-workers (McLuckey *et al.*, 1995; McLuckey & Habibi-Goudarzi, 1993, 1994; McLuckey *et al.*, 1992) on a quadrupole ion trap. These workers proposed a nomenclature scheme for the characterisation of oligonucleotide fragmentation similar to the scheme which has been widely used for peptides. (McLuckey & Habibi-Goudarzi, 1993) Scheme 2.1 shows the nomenclature for a 6-mer with the variation introduced by Vouros and co-workers, (Barry *et al.*, 1995) whereby the oligonucleotide appears as a neutral species. The vertical lines show the possible sites of cleavage along the phosphodiester backbone. The lower case letters represent the nomenclature given to these cleavages with the subscript denoting the position along the backbone at which the cleavage occurs. B_1 to B_6 represent the bases. McLuckey *et al.* demonstrated that two series of ions, the 5′-(a-B_n) series and the complementary 3′-(w_n)-sequence ions, are most commonly observed in MS/MS spectra of oligonucleotides.



Scheme 2.1: Oligonucleotide fragmentation by tandem mass spectrometry.

McLuckey proposed that the ions are formed following fragmentation of multiply charged oligonucleotide anions proceeds via the loss of a base followed by the subsequent cleavage of the 3′-C-O of the sugar from which the base is lost. (McLuckey & Habibi-Goudarzi, 1993) When the charge is retained on the 5′-terminus, an (a-B) ion results, with the complementary w ion being formed if the charge is retained on the 3′-terminus. They also observed that the order for base loss tended to follow the order A>T>>C>G, and that the loss of the 3′-base was unfavourable. The loss of a phosphate group from phosphorylated ions can compete with the loss of a base. Neutral base loss was also found to compete with charged base loss. (McLuckey & Habibi-Goudarzi, 1993) The higher the charge state of the precursor ion, the greater is the preference for charged base loss. Increasing solvation stabilises both the products resulting from neutral base loss as well as charged base loss with the product of highest charge being stabilised to the greatest extent. Hence increasing solvation tends to favour neutral base loss. (McLuckey *et al.*, 1995)

The same tendency for loss of adenine was not found under more energetic collision conditions by Boschenok and Sheil (Boschenok & Sheil, 1996) and Barry *et al.* (Barry *et al.*, 1995) on triple quadrupole instruments. Based on ESI-MS/MS, Ni *et al.* have developed an automated sequencing algorithm, which allows the sequence elucidation of unknown oligonucleotides up to 15-mers. (Ni *et al.*, 1996)

In addition to the (a-B) and w sequence ions, many other product ions are observed in ESI-MS/MS spectra that result from cleavages along the phosphodiester backbone. In many cases, it is difficult to identify these ions conclusively owing to the sugar, C_5H_4O (80.02 Da) and HPO_3 (79.97 Da) and sugar + H_2O, $C_5H_6O_2$, (98.04 Da) and H_3PO_4 (97.98 Da) having the same nominal mass. This results in difficulties in the analysis of these spectra as peaks may occur at the same or very similar *m/z* value requiring high resolution to allow their identification. A typical mass spectrum may contain hundreds of peaks. Hence ESI-MS/MS has not been used routinely for the sequencing of oligonucleotides as it has been with peptides.

There have been relatively few MS/MS studies of covalently modified oligonucleotides. (Barry *et al.*, 1995) The relative stabilities and sequence selectivity of adducts formed from the reaction of aflatoxin B_1 with DNA was explored. (Marzilli *et al.*, 1998) The covalent interactions of hedamycin, cisplatin and *n*-phenanthridinium bromides with oligonucleotides were explored by MS/MS. (Iannitti-Tito *et al.*, 2000; Iannitti *et al.*, 1997) The alkylation of guanine labilises the 3′-glycosidic bond, resulting in cleavage on the 3′-side adjacent to the alkylation site. The sequence ions resulting from this fragmentation

pathway allow the identification of the binding site. The ESI-MS/MS analysis of the HPLC purified adducts formed from the reaction of the carcinogen benzo[*a*]pyrene with an 11-mer showed guanine N2 alkylation. (Ni *et al.*, 1998) The fragmentation pattern was not as predominant as was observed for hedamycin and aflatoxin B_1, suggesting that guanine N2 alkyaltion does not have the same destabilising effect on the DNA backbone as alkylation to the N7 of guanine.

Reports of MS/MS studies of non-covalent complexes are even more scarce. Gabelica and co-workers utilised a quadrupole-time-of-flight (Q-TOF) mass spectrometer to examine the interaction of Hoechst 33258 and netropsin with oligonucleotides. (Gabelica *et al.*, 1999) Their results indicated that netropsin showed selective binding to the GC-rich strand of DNA, whilst Hoechst 33258 showed no preference. The stabilities of the DNA complexes of the antibiotics distamycin (Gale & Smith, 1995) and paromomycin (Griffey *et al.*,1999) have also been examined by tandem mass spectrometry. The 2:1 complex of distamycin with duplex DNA was found to have greater stability than the 1:1 complex. The binding of paromomycin was found to protect oligonucleotides from the common fragmentation pathway involving loss of adenine and cleavage of the 3′-C-O glycosidic bond. Recently, Gross and co-workers looked at a range of drugs, including minor groove binding agents (distamycin, Hoechst 33258, Hoechst 33342, beneril, FeTMpy-P4 and MnTMpy-P4) and an intercalator (actinomycin D). (Wan *et al.*, 2000b) The energy of dissociation, of each of the complexes, was determined and found to correlate with the number of hydrogen bonds involved in stabilising the complex.

2.2 Other Techniques for the Structural Analysis of Ligand-DNA Complexes

Many studies have been conducted on ligand-DNA complexes by both the high-resolution techniques of nuclear magnetic resonance (NMR) spectroscopy and X-ray crystallography. Although X-ray crystallography may have the advantage of allowing direct structural determination, there are numerous significant differences between such determinations in the solid and solution states. In the case of nucleic acids, their conformation is often different in solution owing to hydration and/or crystal packing. (Luxon & Gorenstein, 1995) Also, studies to examine the kinetics of the interaction between the ligand and the DNA or to probe conformational changes, *i.e.* by altering the temperature or salt composition are far more easily undertaken in solution. For the same reason, NMR is preferable to X-ray crystallography for investigating the stoichiometry of complex formation (although mass spectrometry may also be used in this regard).

2.3 Ultraviolet (UV) Spectroscopy and the Thermal Denaturation of DNA

Ultraviolet (UV) spectroscopy is a useful tool for the characterisation of nucleic acids. The UV absorbance at 260 nm is generally used for estimating the concentration of nucleic acids using the Beer-Lambert Law and calculated extinction co-efficients. Absorbance versus temperature (melting curves) can be used to determine the melting temperature (T_m) of an oligonucleotide duplex giving an indication of its stability and also for determining the GC content in normal double-stranded DNA. (Mathews & van Holde, 1990) When an oligonucleotide is heated in solution, the absorbance increases. For double-stranded DNA, there is a sigmoidal increase as the duplex dissociates into single-stranded DNA owing to a reduction in stacking interactions between the bases. The melting curve is characterised by the melting temperature, T_m, which is defined as the temperature at the midpoint of the

absorbance transition. The T_m is concentration independent, but may be influenced by the salt concentration.

UV spectroscopy can be used to study the melting behaviour of DNA, which gives an indication of the stability of the duplex or ligand–DNA complex. Hence, it can be used to compare the relative binding affinities of different DNA ligands. The transition from the double stranded DNA helix to single strands, or DNA melting, occurs when sufficient energy has been supplied to overcome the interactions between the strands. Owing to base pair stacking, the energy required is greater than just the energy to disrupt the hydrogen bonding interactions. DNA melting is a highly cooperative event. Therefore, when the DNA begins to melt, the helix is destabilised and rapidly un-zips into its component single strands. So when the DNA melting is followed by rising temperature, the transition from double stranded helix to single strands occurs over a very narrow temperature range.

DNA melting can be examined by UV spectroscopy because when the two strands form a double helix, the aromatic nucleobases display hypochromicity. That is, the stacking interactions alter the electron density in the rings and they display a reduction in UV absorbance at 260 nm. When the helix separates, the UV absorbance is restored, giving rise to the familiar DNA melting curve. Therefore, the temperature of denaturation, T_m, or melting temperature, is found by following the change in absorption at 260 nm with increased temperature. The T_m is the point of inflexion of the curve.

Changes in the melting temperature of the DNA sample (ΔT_m) indicate changes in the stability of the helix. An increase in T_m shows that the helix is being stabilised, and this usually indicates that the ligand is binding to the double helix. It is also possible that a ligand could destabilise the helix by binding preferentially to single stranded DNA.

2.4 Analysis of Ligand-DNA Complexes by NMR Spectroscopy

Regardless of the mode of binding, one of the most important properties of a ligand-DNA system from the viewpoint of NMR studies is binding specificity. The most desirable ligands are those that are highly specific, so that when reacted with DNA sequences, binding at a single site may be obesrved. This dramatically improves the resulting spectra in comparison to those obtained when several bound complexes co-exist in an equilibrium mixture. For example, actinomycin D has been extensively studied in terms of its binding specificity. (Reid *et al.,* 1983; Brown *et al.,* 1984; Zhou *et al.,* 1989) Mass spectrometry may be utilised in the early stages of analysis to screen for binding specificity prior to larger scale, more detailed, structural studies by NMR spectroscopy.

Selection of the DNA sequence is to a large extent determined by the binding properties of the drug. For a drug displaying a high level of binding specificity, a single site may be incorporated into a duplex, but depending on the exact binding site, it may or may not be possible to use a self-complementary sequence, for example, $d(ATGCAT)_2$ has one site, whilst $d(CGTACG)_2$ has two sites for actinomycin D. Examples of drugs whose complexes with nucleic acids have been examined include daunomycin, (Quigley *et al.,* 1980; Odefey

et al., 1992) echinomycin (Gilbert & Feigon, 1992; Park & Choi, 1995) and mithramycin. (Banville *et al.*, 1990; Sastry & Patel, 1993; Keniry *et al.*, 1991, 1993)

2.5 NMR Spectroscopy

Nuclear magnetic resonance spectroscopy has become an important tool for structural investigations of ligand-DNA interactions as a result of the many recent developments in technology. These include the production of reliable, super-conducting magnets coupled to the application of the pulse technique and the associated Fourier transform (FT). Increased sensitivity has lead to smaller sample requirements (microgram range), whilst the greater control and manipulation of a sample's magnetisation arising from the pulse technique has resulted in NMR providing more detailed structural information than any other single technique. (Derome, 1995) NMR spectroscopy has one great advantage over X-ray crystallography in its requirement for the sample to be in solution rather than a crystalline state. This means the environment around the DNA is closer to *in vivo* conditions and also allows the structure and dynamics of interacting molecules to be observed.

2.5.1 Theory of NMR

NMR is concerned with the interaction between an oscillating magnetic field (the *rf* field), B_1, and the net magnetisation of a sample, which arises as a result of the application of a static magnetic field, B_0. In a magnetic field, nuclei are observed to occupy different energy levels according to their orientation to that field. The number of different orientations that can be adopted within a magnetic field is 2I+1, where I represents the spin quantum number for the nucleus. Most NMR-active nuclei within biological systems possess a spin number of $^1/_2$. The spin-$^1/_2$ nuclei can adopt one of two orientations within a magnetic field

referred to as $+^1/_2$ and $-^1/_2$. The nuclei that adopt an orientation with the magnetic field are of lower energy than those aligned against. This energy difference is the basis of the NMR signal. The NMR absorption is a consequence of transition between these two energy states and is stimulated by the applied *rf* field. A number of different nuclei that possess spin-$^1/_2$ exist in nucleic acids, namely ^1H, ^{13}C, ^{31}P and ^{15}N which can all be detected using NMR.

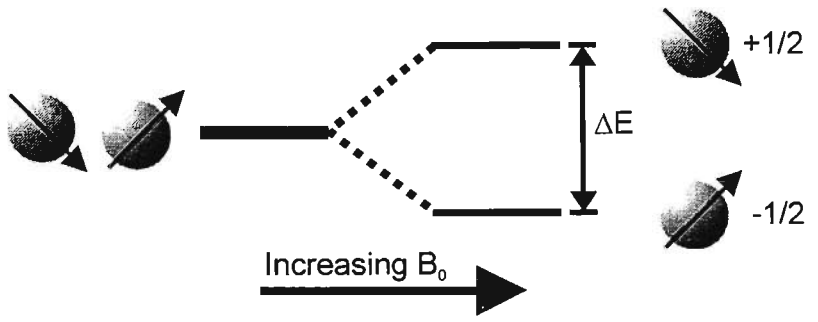


Figure 2.5: The energy levels of the two spin states in an applied magnetic field.

The difference in energy can be quantified using equation 5. ΔE is the energy separating the high and low energy spin states, B_0 is the magnetic field strength (typically in Tesla), *h* is Plank's constant, and γ is the magneto-gyro ratio, a constant specific to the isotope being observed.

$$\Delta E = \frac{\gamma h B_0}{2\pi} \qquad \ldots(5)$$

The energy required to excite the transition between the two energy levels occurs in the radio frequency (*rf*) region of the electromagnetic spectrum. Magnet strengths are often described in terms of the frequency at which protons resonate, for example a 2.35 or 11.74 Tesla magnet causes protons to resonate at about 100 MHz or 500 MHz respectively. The

distribution between the energy levels can be described by the Boltzmann equation. This shows the excess of nuclei that reside in the lower energy state at thermal equilibrium (Equation 6). The equation reveals that there is only an excess of one ^{1}H nuclei in the low energy state for every hundred thousand at 298 K and a field strength of 2.35 Tesla (100 MHz). This small excess even in very strong magnetic fields accounts for the relative insensitivity of NMR as a spectroscopic technique.

$$\frac{N_\beta}{N_\alpha} = e^{(-\Delta E/kT)} \qquad \ldots(6)$$

Equation 6 shows the populations of the high energy state N_β over low energy level N_α for a specific energy gap ΔE, where k is Boltzmann's constant (1.3805×10^{-23} J/Kelvin) and T is the temperature in Kelvin.

The insensitivity is further compounded by the natural abundance of some NMR active nuclei studied. To overcome these problems, sensitive electronics, signal averaging techniques and relatively high concentrations of sample (compared with other techniques) are required to obtain high quality spectra.

An alternating voltage is applied across the ends of the coil in the NMR probe, which induces an alternating magnetic field (B_1) throughout the sample. The amplitude of this field is much smaller than that of the static magnetic field, B_0, and its frequency covers that of the NMR resonances to be observed. At equilibrium, the spins are all aligned with the applied magnetic field, which defines the *z*-axis, with no component along the *x*- or *y*-axes. However, when a pulse of *rf* radiation is applied to the spins, an *x*-*y* component is induced, and the *z*-component is altered from its equilibrium value. The spins return to their

equilibrium values following a first-order exponential decay. The plot of intensity versus time is referred to as the Free Induction Decay (FID) in an NMR experiment. The actual FID obtained comprises the sum of many overlapping sine waves and the individual frequencies are obtained by a Fourier transform (FT). The result is a plot of amplitude versus frequency as shown in figure 2.6.

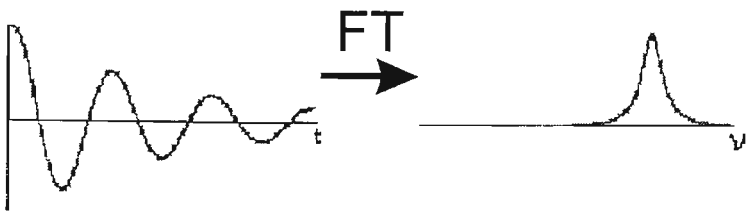


Figure 2.6: FID to frequency plot (FT).

Signal averaging involves the repetition of experiments in order to gain a better signal-to-noise ratio, however, this may be time consuming owing to the relatively long relaxation times required by some nuclei. T_1 is known as the *longtitudinal* or spin-lattice relaxation time. It describes the behaviour of the magnetisation in the z-component, and is the time taken for the z-component of the magnetisation to return to its equilibrium value. There is a second type of relaxation known as *transverse* or spin-spin relaxation. It occurs in the x-y plane, is described by the time constant, T_2 and is owing to the inhomogeneity of the magnetic field. The relaxation time, T_2, is inversely proportional to the resonance peak width.

The NMR spectrum represents the frequencies at which the transitions are induced between the two allowed states of the spin $+^1/_2$ and $-^1/_2$. (Wemmer, 2000) Spins of a given type, *i.e.*

^{1}H, resonate around a given frequency, *i.e.* 500 MHz in a field of 11.74 T. However, nuclei of the same type, but in different chemical environments, have slightly different frequencies. These differences are known as chemical shifts. The chemical shift (δ) defines the location of the NMR line and is calibrated against a reference compound, tetramethylsilane (TMS) or 3-trimethylsilylpropionate (TSP) for the ^{1}H nuclei. In frequency units, the chemical shift is proportional to the applied static magnetic field, and as such are commonly quoted in parts per million (ppm). The relative line intensity in a 1D-NMR experiment is representative of the number of nuclei associated with each resonance line. (Wuthrich, 1986)

2.5.2 2D NMR Techniques

Two-dimensional NMR techniques provide increased resolution and information over 1D-NMR spectra. One major advantage of 2D-NMR is the ability to investigate the tertiary structure of protein and DNA molecules up to ca. 10000 Da in molecular mass. To obtain a 2D spectrum (amplitude versus two frequencies), two Fourier transforms must be carried out, one for rows and one for columns. Some spins may exchange magnetisation during the mixing period, which gives rise to cross-peaks in the 2D spectrum. These cross-peaks are of great interest and use in the assignment of the resonances, which may be either close through bonds or through space.

2.5.3 DQF-COSY and TOCSY

Double quantum filtered correlation spectroscopy (DQF-COSY) and total correlation spectroscopy (TOCSY) are the two most common ^{1}H-NMR experiments used to determine

through bond information. They involve scalar spin-spin coupling and identify spin systems. (Piantini, 1982; Rance *et al.*, 1983; Bax & Davis, 1985)

DQF-COSY (Piantini, 1982; Rance *et al.*, 1983) observes connectivities between protons that are up to three bonds away. This type of experiment has advantages over normal COSY (Bax & Freeman, 1981) experiments in that there is a significant reduction in singlet resonances (eg, methyl or methoxy) and improved resolution owing to the diagonal and cross peaks being in phase, compared to 90° out of phase for COSY experiments.

A DQF-COSY experiment allows the assignment of spin systems within each nucleotide for an oligonucleotide. However, no information may be gained as to the position of the base or sugar within the sequence as each deoxyribose is an isolated spin system.

A disadvantage of DQF-COSY is the overlap of resonances that occurs, especially of sugar resonances, as well as the peaks being anti-phase, which leads to cancellation, which limits the usefulness of this technique. This may be overcome by the use of TOCSY (Bax & Davis, 1985) experiments, which have the advantage of resolving longer-range connectivities.

2.5.4 NOE: the Nuclear Overhauser Effect

Experiments based on the nuclear Overhauser effect (NOE) examine the interaction involved in the direct magnetic coupling between nuclei (dipolar coupling). (Derome, 1987) The information gained from this technique pertains to internuclear distances and

molecular motion. The NOE is a change in intensity of an NMR resonance when the transitions of another are perturbed.

2.5.5 NOESY: Nuclear Overhauser Effect Spectroscopy

The NOESY (Jeener *et al.*, 1979; Kumar *et al.*, 1980) experiment determines correlation between nuclei that are close in space. Cross-peaks are observed in NOESY spectra as a result of mutual (cross) relaxation between nuclei that are close in space. The magnitude of the effect is dependent on the distance between spins and how fast the molecule, or the part of the molecule that contains the spins, is moving.

2.5.6 Water Suppression and Exchangeable Resonances

To avoid the large proton signal that occurs in water (H_2O), samples may be dissolved in D_2O. However, the more acidic nuclei, such as those attached to O or N, have very short lifetimes and rapidly exchange with deuterium. To characterise such resonances, water must be used as the solvent and thus, special methods for solvent suppression must be employed. There are two common approaches. The first involves saturating the solvent with a weak *rf* pulse before the normal pulse sequence, *i.e.* presaturation. This is a simple and easy method to use. The second method involves designing a new pulse sequence, which excites the resonances of interest, but not those of the solvent.

Many of the exchangeable protons are involved in H-bonding and thus can be used as an indicator of secondary and tertiary structure of the nucleic acid. The difference in chemical shift of the imino and amino protons means that they can be independently assigned. The hydrogen-bonded iminos resonate between 12-15 ppm, whilst the amino protons appear in

the 6.5-8.5 ppm range. In a NOESY spectrum of DNA, it is possible to observe cross-peaks from each imino to to the imino groups of its immediate neighbours and also to the amino and H2 protons within the base pair.

2.5.7 Chemical Shift Perturbations Arising from Ligand Binding

By examining the alterations in chemical shift of a ligand resonance resulting from non-covalent binding, an indication of whether a particular substituent is involved in binding can be obtained.

Intercalating agents produce the most dramatic changes in DNA conformation since the process of intercalation results in the unwinding of the helix to accommodate the intercalating agent. (Searle, 1988) This effect may be observed on neighbouring protons as well as protons further away from the intercalation site.

2.5.8 Typical Chemical Shifts of Short Olignucleotides

The assignment of oligonucleotides and complexes was achieved using established techniques based on the combined analysis of DQF-COSY, TOCSY and NOESY spectra. (Wijmenga *et al.*, 1993; Chazin *et al.*, 1986)

The typical ^{1}H-NMR spectrum of a short oligonucleotide can be divided into five distinct regions. Aromatic resonances that arise from the four bases of DNA (*i.e.* H6, H8 and AH_2) occur between 6.5 and 8.5 ppm. Cytosine H5 and H1' sugar resonances occur between 5.0 and 6.5 ppm. H3' sugar resonances are found between 4.5 and 5.5 ppm, while H4', H5' and H5" resonances are usually overlapping between 3.5 and 4.5 ppm. H2' and H2" resonances

occur further upfield between ca. 1.8 and 3.0 ppm. Thymine methyl resonances are generally the most upfield resonances, being observed between ~1.0 and 1.8 ppm.

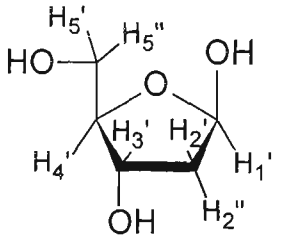


Figure 2.7: Structure of deoxyribose including standard nomenclature scheme.

Exchangeable imino resonances of thymine and guanine and amino resonances of guanine, adenine and cytosine can be observed in H_2O solvent. Imino resonances occur between 12 and 15 ppm, while amino resonances are found between 6.5 and 8.5 ppm.

2.5.9 Comparison of Techniques: Mass Spectrometry vs NMR spectroscopy

Both mass spectrometry and nuclear magnetic resonance spectroscopy can provide detailed sequence and structural information on biomolecules and their complexes with ligands. The information obtained differs between the two techniques, however, the combination of modern mass spectrometric and multi-dimensional NMR spectroscopic techniques allows the structural investigation of intact biomolecules such as proteins, oligonucleotides and complex carbohydrates.

Mass spectrometry gives accurate mass measurement and can yield detailed information on primary structure (sequence), whilst NMR spectroscopy allows the elucidation of the

secondary and tertiary structure of peptides, proteins and other biomolecules. In addition, MS can also provide insight into higher order structure and inter- and intramolecular interactions, whilst NMR can pinpoint the position and orientation of a bound ligand and can provide information about the ligand-receptor contacts. Mass spectrometry holds the potential of providing structural information on proteins with larger masses (>100 kDa) than can be studied with nuclear magnetic resonance (NMR) spectroscopy (40 kDa). Mass spectrometry also requires considerably smaller sample sizes (picomolar) than NMR (millimolar).

Bibliography

Ahrer, W. and Buchbreger, W. (1999) *J. Chrom. A* **854**, 275-287

Banville, D.L.; Keniry, M.A. and Shafer, R.H. (1990) *Biochemistry* **29**, 9294-9304

Barber, M.; Bordoli, R.S.; Sedgwick, R.D. and Tyler, A.N. (1981) *J. Chem. Soc. Chem. Comm.* 325-327

Barry, J.P.; Vouros, P.; Van Schepdael, A. and Law, S.-J. (1995) *J. Mass Spectrom.* **30**, 993-1006

Bax, A. and Davis, D.G. (1985) *J. Mag. Res.* **65**, 355-360

Bax, A. and Freeman, R. (1981) *J. Mag. Res.* **44**, 542-561

Beck, J.L., Colgrave, M.; Ralph, S.R. and Sheil, M.M. (2001) *Mass Spec. Rev.* **20(2)**, 61-87.

Beavis, R.C. and Chait, B.T. (1990) *Proc. Nat. Acad. Sci. USA* **87**, 6873-6877

Biemann, K. and Martin, S.A. (1987) *Mass Spectrom. Rev.* **6**, 1-76

Boschenok, J. and Sheil, M.M. (1996) *Rapid Commun. Mass Spectrom.* **10**, 144-149

Brown, S.C.; Mullis, K.; Levenson, K.C. and Shafer, R.H. (1984) *Biochemistry* **23**, 403-408

Buchanan, M.V. and Hettich, R.L. (1993) *Anal. Chem.* **65**, 245A-259A

Busch, K.L.; Glish, G.L. and McLuckey, S.A. (1988) *Mass Spectrometry/Mass Spectrometry: Techniques and applications of tandem mass spectrometry* VCH, NY

Chapman, J.R. "*Mass Spectrometry: Ionisation methods and instrumentation*" (1996) in *Methods in Molecular Biology* (J.R. Chapman, Ed.) Humana Press, NJ, vol. 61

Chazin, W.J.; Wuthrich, K.; Hyberts, S.; Rance, M.; Denny, W.A. and Leupin, W. (1986) *J. Mol. Biol.* **190**, 439-453

Cheng, X.; Camps, D.G.; Wu, Q.; Bakhtiar, R.; Springer, D.L.; Morris, B.J.; Bruce, J.E.; Anderson, G.A. and Smith, R.D. (1996) *Nucleic Acids Res.* **24**, 2183-2189

Chiarelli, M.P. and Lay Jr., J.O. (1992) *Mass Spec. Rev.* **11**, 447-493

Connolly, B.A.; Potter, B.V.; Eckstein, F.; Pingoud, A. and Grotjahn, L. (1984) *Biochemistry* **23**, 3443-3453

Cotter, R.J. (1988) *Anal. Chem.* **60**, 781A-793A

Covey, T.R.; Bonner, R.F. and Shushan, B.I. (1988) *Rapid Commun. Mass Spectrom.* **2**, 249-253

Crain, P.F. (1990) *Mass Spectrom. Rev.* **9**, 505-554

D'Andrea, A. D. and Haseltine, W. A. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 3608-3612

Derome, A.E. (1987) *Modern NMR Techniques for Chemistry Research,* Pergamon Press, Oxford, p97

Deroussent, A.; LeCaer, J.P.; Rossier, J. and Gouyette, A. (1995) *Rapid Commun. Mass Spectrom.* **9**, 1-4

Doctycz, M.J.; Habibi-Goudarzi, S. and McLuckey, S.A. (1994) *Anal. Chem.* **66**, 3416-3422

Dole, M.; Mack, L.L.; Hanes, R.L.; Mobley, R.C.; Ferguson, L.D. and Alice, M.B. (1968) *J. Chem. Phys.* **49**, 2240

Gabelica, V.; de Pauw, E. and Rosu, F. (1999) *J. Mass Spectrom.* **34**, 1328-1337

Gabelica, V.; Rosu, F.; Houssier, C. and de Pauw, E. (2000) *Rapid Commun. Mass Spectrom.* **14**, 464-467

Gale, D.C. and Smith, R.D. (1995) *J. Am. Soc. Mass Spectrom.* **6**, 1154-1164

Gale, D.C.; Goodlett, D.R.; Light-Wahl, K.J. and Smith, R.D. (1994) *J. Am. Chem. Soc.* **116**, 6027-6028

Gao, Q.; Cheng, X.; Smith, R.D.; Yang, C.F. and Goldberg, I.H. (1996) *J. Mass Spectrom.* **31**, 31-36

Ganem, B.; Li, Y.-T. and Henion, J.D. (1991) *J. Am. Chem. Soc.* **113**, 6294-6296

Ganem, B.; Li, Y.-T. and Henion, J.D. (1991) *J. Am. Chem. Soc.* **113**, 7818-7819

Gilbert, D.E. and Feigon, J. (1992) *Nucleic Acid Res.* **20**, 2411-2420

Griffey, R.H.; Greig, M.J.; An, H.; Sasmor, H. and Manalili, S. (1999) *J. Am. Chem. Soc.* **121**, 474-475

Grimaldi, K. A.; McAdam, S. R.; Souhami, R. L. and Hartley, J. A. (1994) *Nucleic Acids Res.* **22**, 2311-2317

Grotjahn, L.; Blocker, H. and Frank, R. (1985) *Biomed. Mass Spectrom.* **12**, 514-524

Grotjahn, L. and Taylor, L.C.E. (1985) *Org. Mass Spectrom.* **20**, 146-152

Haseltine, W. A.; Lindan, C. P.; D'Andrea, A. D. and Johnsrud, L. (1980) *Methods Ezymol.* **65**, 235-248

Hettich, R.L. and Buchanan, M.V. (1991) *Int. J. Mass Spectrom. Ion Proc.* **111**, 365-395

Hsieh, Y.L.; Li, Y-T.; Henion, J.D. and Ganem, B. (1994) *Biol. Mass Spectrom.* **23**, 272-276

Huang, E.C.; Pramanik, B.N.; Tsarbopoulos, A.; Reichert, P.; Ganguly, A.K.; Trotta, P.; Nagabhushan, T.L. and Covey, T.R. (1993) *J. Am. Soc. Mass Spectrom.* **4**, 624-630

Iannitti, P.; Sheil, M.M. and Wickham, G. (1997) *J. Am. Chem. Soc.* **119**, 1490-1491

Iannitti-Tito, P.; Weimann, A.; Wickham, G. and Sheil, M.M. (2000) *Analyst* **125**, 627-633

Iribarne, J.V. and Thomson, B.A. (1976) *J. Chem. Phys.* **64**, 2287-2294

Jaquinod, M.; Leize, E.; Potier, N.; Albrecht, A.-M.; Shanzer, A. and Van Dorselaar, A. (1993) *Tetrahedron Lett.* **34**, 2771-2774

Jardine, I. (1990) *Nature* **345**, 747-748

Jeener, J.; Meier, B.H.; Bachmann, P. and Ernst, R.R. (1979) *J. Chem. Phys.* 71, 4546

Kapur, A.; Beck, J.L. and Sheil, M.M. (1999) *Rapid Commun. Mass Spectrom.* **13**, 2489-2497

Karas, M.; Bachmann, D.; Bahr, U. and Hillenkamp, F. (1987) *Int. J. Mass Spectrom. Ion Proc.* **78**, 53-68

Karas, M.; Bahr, U. and Giebmann, U. (1991) *Mass Spectrom. Rev.* **10**, 335-357

Keniry, M.A.; Banville, D. L. and Shafer, R.H. (1991) *FEBS Letters* **289**, 210-212

Keniry, M.A.; Banville, D.L.; Simmonds, P.M. and Shafer, R. (1993) **231**, 753-767

Kennard, O. and Salisbury, S.A. (1993) *J. Biol. Chem.* **268**, 10701

Kowalak, J.A.; Pomerantz, S.C.; Crain, P.F. and McCloskey, J.A. (1993) *Nucleic Acids Res.* **21**, 4577-4585

Krahmer, M.T.; Walters, J.J.; Fox, K.F.; Fox, A.; Creek, K.E.; Pirisi, L.; Wunschel, D.S.; Smith, R.D.; Tabb, D.L. and Yates, J.R. 3rd. (2000) *Anal. Chem.* **72**, 4033-4040

Krugh, T.R. (1994) *Current Opin. Struct. Biol.* **4**, 351-364

Kumar, A.; Ernst, R.R. and Wuthrich, K. (1980) *Biochem. Biophys. Res. Commun.* **95**, 1-6

Light-Wahl, K.J.; Loo, J.A.; Edmonds, C.G.; Smith, R.D.; Witkowska, H.E.; Shackleton, C.H.L. and Wu, C.-S.C. (1993) *Biol. Mass Spectrom.* **22**, 112-120

Light-Wahl, K.J.; Springer, D.L.; Winger, B.E.; Edmonds, C.G.; Camp II, D.G.; Thrall, B.D. and Smith, R.D. (1993) *J. Am. Chem. Soc.* **115**, 803-804

Lichtenwalter, K.G.; Apffel, A.; Bai, J.; Chakel, J.A.; Dai, Y.; Hahnenberger, K.M.; Li, L. and Hancock, W.S. (2000) *J. Chrom. B* **745**, 231-241

Limbach, P.A. (1996) *Mass Spectrom. Rev.* **15**, 297-336

Limbach, P.A.; Crain, P.F. and McCloskey, J.A. (1995) *Current Opin. Biotech.* **6**, 96-102

Little, D.P.; Chorush, R.A.; Speir, J.P.; Senko, M.W.; Kelleher, N.M. and McLafferty, F.W. (1994) *J. Am. Chem. Soc.* **116**, 4893-4897

Loo, J.A. (1997) *Mass Spectrom. Rev.* **16**, 1-23

Luxon, B.A. and Gorenstein, D. (1995) *Methods in Enzymology* TL James (Ed) v.261, Academic Press, San Diego, CA, p45

Mann, M., Shen, S. and Fenn, J. B. (1992) *Electrospray Mass Spectrometry: Principles and Methods*, in *Mass Spectrometry: Clinical and Biomedical Applications* (D. M. Desiderio, Ed.) Plenum Press, New York. p.1-35.

Mann, M. (1990) *Org. Mass Spectrom.* **25**, 575-587

Mathews, C.K. and van Holde, K.E. (1990) *Biochemistry* Benjamin Cummings Publishing Co., California, p.117

Maxam, A.M. and Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 560-564

Maxam, A.M. and Gilbert, W. (1980) *Methods in Ezymology* **65**, 499-560

McCloskey, J.A. and Crain, P.F. (1992) *Int. J. Mass Spectrom. Ion Processes* **118/119**, 593

McCloskey, J.A. and Crain, P.F. (1998) *Curr. Opin. Biotechnol.* **9**, 25-34

McLuckey, S.A.; Glish, G.L.; Asano, K.G. and van Berkel, G.J. (1988) *Anal. Chem.* **60**, 2312-2314

McLuckey, S.A. and Habibi-Goudarzi, S. (1993) *J. Am. Chem. Soc.* **115**, 12085-12095

McLuckey, S.A. and Habibi-Goudarzi, S. (1994) *J. Am. Soc. Mass Spectrom.* **5**, 740-747

McLuckey, S.A.; Vaidyanathan, G. and Habibi-Goudarzi, S. (1993) *J. Mass Spectrom.* **30**, 1222-1229

McLuckey, S.A.; Van Berkel, G.J. and Glish, G.L. (1992) *J. Am. Soc. Mass Spectrom.* **3**, 60-70

McNeal, C.J. and MacFarlane, R.D. (1981) *J. Am. Chem. Soc.* **103**, 1609-1610

Meng, C.-K.; McEwen, C.N. and Larsen, B.S. (1990) *Rapid Commun. Mass Spectrom.* **4**, 147-150

Miketova, P. and Schram, K.H. (1997) *Molecular Biochem.* **8**, 249-253

Morris, H.R.; Paxton, T.; Dell, A.; Langhorne, J.; Berg, M.; Bordoli, R.S.; Hoyes, J. and Bateman, R.H. (1996) *Rapid Commun. Mass Spectrom.* **10**, 889-896

Mueller, P. R. and Wold, B. (1989) *Science* **246**, 780-786

Murray, V. (1999) *Prog. Nucl. Acid Res. Mol. Biol.* **63**, 367-415

Neri, N.; Sindona, G. and Uccella, N. (1983) *Gazz. Chim. Ital.* **113**, 197

Ni, J.; Liu, T.; Kolbanovskiy, A.; Krzeminski, J.; Amin, S. and Geacintov, N.E. (1998) *Anal. Biochem.* **264**, 222-229

Ni, J.; Pomerantz, S.C.; Rozenski, J.; Zhang, Y. and McCloskey, J.A. (1996) *Anal. Chem.* **68**, 1989-1999

Nordhoff, E.; Kirpekar, F. and Roepstorff, P. (1996) *Mass Spectrom. Rev.* **15**, 67-138

Norwood, C.B. and Vouros, P. (1994) *DNA modifications: Investigations by mass spectrometry* (C.B. Norwood & P. Vouros, Eds.) Plenum Press, NY. Vol. 2, 89-133

Odefey, C.; Westendorf, J.; Dieckmann, T. and Oschkinat, H. (1992) *Chem-Biol. Interact.* **85**, 117-126

Owens, D.R.; Bothner, B.; Phung, Q.; Harris, K. and Siuzdak, G. (1998) *Bioorg. Med. Chem.* **6**, 1547-1554

Park, J.Y. and Choi, B.S. *J. Biochem.* **118**, 989-995

Pfeifer, G. P.; Steigerwald, S. D.; Mueller, P. R.; Wold, B. and Riggs, A. D. (1991) *Science* **246**, 810-813

Piantini, U.; Sorensen, O.W. and Ernst, R.R. (1982) *J. Am. Chem. Soc.* **104**, 6800-6801

Poddevin, B.; Meguenni, S.; Elias, I.; Vasseur, M. and Blumenfeld, M. (1994) *Antisense Res. Dev.* **4**, 147-154

Ponti, M.; Forrow, S. M.; Souhami, R. L.; D'Incalci, M. and Hartley, J. A. (1991) *Nucleic Acids Res.* **19**, 2929-2933

Portugal, J. (1989) *Chem-Biol. Interact.* **71**, 311-324

Potier, N.; van Dorsselaer, A.; Cordier, Y., Roch, O. and Bischoff, R. (1994) *Nucleic Acids Res.* **22**, 3895-3903

Przybylski, M.; Kast, J.; Glocker, M.O.; Durr, E.; Bosshard, H.R.; Nock, S. and Sprinzl, M. (1995) *Toxicol. Lett.* **82/83**, 567-575

Quigley, G.J.; Wang, A.H.; Ughetto, G.; van der Marel, G.; van Boom, G.J. and Rich, A. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 7204

Rance, M.; Sorenson, O.W.; Bodenhausen, G.; Wangner, G.; Ernst, R.R. and Wuthirch, K. (1983) *Biochem. Biophys. Res. Commun.* **117**, 479-485

Reddy, D.M. and Iden, C.R. (1993) *Nucleosides and nucleotides* **12**, 815-826

Reddy, D.M.; Rieger, R.A.; Torres, M.C. and Iden, C.R. (1994) *Anal. Biochem.* **220**, 200-207

Reid, D.G.; Salisbury, S.A. and Williams, D.H. (1983) *Biochemistry* **22**, 1377-1385

Reinhold, V.N. and Carr, S.A. (1983) *Mass Spectrom. Rev.* **2**, 153-221

Sanger, F.; Nicklen, S. and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467

Sastry, M. and Patel, D.J. (1993) *Biochemistry* **32**, 6588-6604

Schmelzeisen-Redeker, G.; Butfering, L. and Rollgen, F.W. (1989) *Int. J. Mass Spectrom. Ion Proc.* **90**, 139-150

Schwartz, B.L.; Gale, D.C.; Smith, R.D.; Chilkoti, A. and Stayton, P.S. (1995) *J. Mass Spectrom.* **30**, 1095-1102

Schwartz, B.L.; Light-Wahl, K.J. and Smith, R.D. (1994) *J. Am. Soc. Mass Spectrom.* **5**, 201-204

Searle, M.S. (1992) *Progress in NMR Spectroscopy* **25**, 403-480

Searle, M.S.; Hall, J.G.; Denny, W.A. and Wakelin, L.P.G. (1988) *Biochemistry* **27**, 4340-4349

Siegel, M.M.; Tabei, K.; Kunz, A.; Hollander, I.J.; Hamann, R.R.; Bell, D.H.; Berkenkamp, S. and Hillenkamp, F. (1997) *Anal. Chem.* **69**, 2716-2726

Sindona, G. (1989) *Chimicaoggi* 71-75

Sindona, G.; Uccelle, N. and Weclawek, K. (1982) *J. Chem. Res.* 184-185

Smith, R.D.; Loo, J.A.; Edmonds, C.G.; Barinaga, C.J. and Udseth, H.R. (1990) *Anal. Chem.* **62**, 882-899

Smith, R.D. and Light-Wahl, K.J. (1993) *Biol. Mass Spectrom.* **22**, 493-501

Smith, R.D.; Bruce, J.E.; Wu, Q.Y. and Lei, Q.P. (1997) *Chem. Soc. Rev.* **26**, 191-202

Tang, K.; Taranenko, N.I.; Allman, S.A.; Chang, L.Y. and Chen, C.H. (1994) *Rapid Commun. Mass Spectrom.* **8**, 727-730

Tito, M.A. (1998) *"Nanoelectrospray and tandem mass spectrometry of biological molecules"* BSc (Hons) Thesis, University of Wollongong.

Triolo, A.; Arcamone, F.M.; Rafaelli, A. and Salvaori, P. (1997) *J. Mass Spec.* **32**, 1186-1194

Varani, G. (1995) *Annu. Rev. Biophys. Biomol. Struct.* **24**, 379-404

Verentchikov, A.N.; Ens, W. and Standing, K.G. (1994) *Anal. Chem.* **66**, 126-133

Vestal, M.; Juhasz, P. and Martin, S. (1995) *Rapid Commun. Mass Spectrom.* **9**, 1044-1050

Wan, K.X.; Shibue, T. and Gross, M.L. (2000a) *J. Am. Chem. Soc.* **122**, 300-307

Wan, K.X.; Gross, M.L. and Shibue, T. (2000b) *J. Am. Soc. Mass Spectrom.* **11**, 450-457

Wemmer, D. in *"Structure and dynamics by NMR"* (2000) Nucleic Acids: Structures, properties and functions. (V.A. Bloomfield, D.M. Crothers and I. Tinoco Jr., Eds) University Science Books, CA, p111-163

Wickham, G.; Iannitti, P.; Boschenok, J. and Sheil, M.M. (1995) *Rapid Commun. Mass Spectrom.* S197-203

Wickham, G.; Iannitti, P.; Boschenok, J. and Sheil, M.M. (1995) *FEBS Lett.* **360**, 231-234

Wijmenga, S.S.; Mooren, M.M. and Hilbers, C.W. in *"NMR of nucleic acids: from spectrum to structure"* (1993) NMR of Macromolecules: A Practical Approach, (G.C.K. Roberts, Ed.), IRL Press, Oxford

Wilson, M.S. and McCloskey, J.A. (1975) *J. Am. Chem. Soc.* **97**, 3436-3444

Wu, H. and Aboleneen, H. (2000) *Anal. Biochem.* **287**, 126-135

Wu, H. and Aboleneen, H. (2001) *Anal. Biochem.* **290**, 347-352

Wu, Q.; Cheng, X.; Hofstadler, S.A. and Smith, R.D. (1995) *J. Mass Spectrom.* **31**, 669

Wuthrich, K. (1986) NMR of proteins and nucleic acids, John Wiley & Sons, NY

Yamashita, M. and Fenn, J.B. (1984) *J. Phys. Chem.* **88**, 4451-4459

Zhou, N.; James, T.L. and Shafer, R.H. (1989) *Biochemistry* **28**, 5231-5239

CHAPTER THREE: *Sequence Selectivity, Stability, the Pyrindamycins and Mass Spectrometry*

3.0 Introduction

As new DNA-interactive compounds become available, through discovery of natural products or synthetic drug development, there is a need to investigate their chemical, biological and toxicological properties. As outlined in chapter two, mass spectrometry readily enables the examination of ligand-DNA interactions, with particular capabilities for the determination of the sequence selectivity and stability of the ligand-DNA adducts. This chapter details the study of pyrindamycin-oligonucleotide adducts by mass spectrometry. The pyrindamycins are minor groove alkylating agents (see section 1.7). They were selected for study because they are still of interest because of their biological activity against bacteria and some tumour cell lines and their lower toxicity compared to the lethal related compound CC-1065. Further, compared to earlier studies of guanine alkylators by ESI-MS and MS/MS, (Iannitti, 1999; Iannitti-Tito *et al.*, 2000) pyrindamycins alkylate adenine, are more labile and it is more difficult to isolate the intact adducts. Hence, the compounds provide a more demanding test case of the potential of mass spectrometry for rapid characterisation of sequence selectivity.

3.1 Oligonucleotides Used in this Study

As noted in chapter one (section 1.7), the pyrindamycins bind in a sequence selective manner to the adenine N3 in a minor groove alkylation. This binding spans 3.5 base-pairs. Hence, there is an absolute requirement for the sequence of three bases, whilst a preference for the identity of the fourth base has also been observed. (Boger *et al.*, 1990) The

oligonucleotides used were designed to incorporate the high affinity binding site 5′-(AAA)-3′. We varied the base composition in the fourth 5′-position to examine the effect on binding and stability. From the literature, (Boger *et al.*, 1990) it was expected that the binding and stability of the adducts formed would follow the trend:

5′-(AAA)-3′ > 5′-(TTA)-3′ > 5′-(TAA)-3′ > 5′-(ATA)-3′

Moreover, alkylation should exhibit a strong, but not requisite specificity, preferring an A or T versus G or C in the fourth 5′-position. For the lower affinity sites, 5′-(TAA)-3′ and 5′-(ATA)-3′, the preferred fourth 5′-base would be expected to be a purine (A or G) versus a pyrimidine (C or T).

The oligonucleotides 5′-CGCAAATTTGCG-3′ (ON1), 5′-CGGAAATTTCCG-3′ (ON2), 5′-CGAAAATTTTCG-3′ (ON3), 5′-GCAAAATTTTGC-3′ (ON4), 5′-CGTAAATTTACG-3′ (ON5), 5′-CTAAAATTTTAG-3′ (ON6), 5′-CGTTTATAAACG-3′ (ON7), 5′-CGCAAAGCTTTG-3′ (ON8), 5′-GCGTTTCGAAAC-3′ (ON9), 5′-CGAAAGCTTTCG-3′ (ON10) and 5′-CGCAAAGCTTTGGGC-3′ (ON11) were purchased from Bresagen, (Adelaide, Australia) where they were synthesised on a 1 μmol scale. All oligonucleotides were purchased in the crude form with the dimethoxytrityl (DMT) protecting group attached to the 5′-hydroxyl end of the nucleoside as shown in figure 3.1.

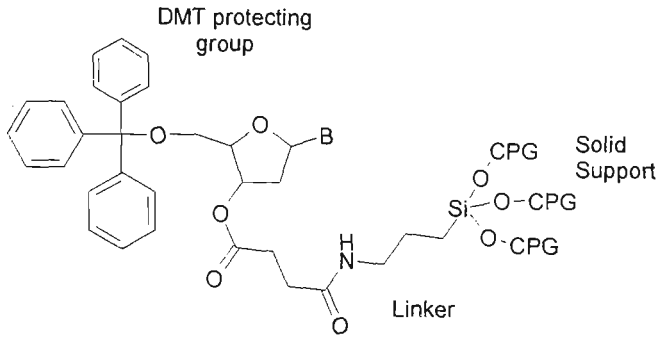


Figure 3.1: DMT-protected nucleoside, attached to controlled glass pore solid support (CPG). B is one of the four bases: cytosine(C), thymine(T), adenine(A) or guanine(G).

The majority of the sequences used were self-complementary which means that they can form specific base pairs with an antiparallel strand of the same sequence resulting in double-stranded, or duplex, oligonucleotides. Several of the oligonucleotides were designed to be non-self complementary to investigate whether alkylation would occur to single-stranded oligonucleotides, e.g., 5′-d(CGCAAAGCTTTG)-3′ (ON8). The oligonucleotide 5′-d(GCGTTTCGAAAC)-3′ (ON9) was designed so that, when combined in solution with ON8, a double-stranded complex would be formed. The oligonucleotide 5′-d(CGAAAGCTTTCG)-3′ (ON10) was designed to incorporate the same binding site as ON8, whilst being self-complementary, so that a double-stranded complex could also be formed in solution. Cytosine and guanine bases were included at the termini in order to increase the stability of the double-stranded conformation since CG-containing oligonucleotides have elevated melting temperatures owing to the increased hydrogen bonding interactions relative to AT-containing oligonucleotides. (Metzler, 1977). The oligonucleotide 5′-d(CGCAAAGCTTTGGC)-3′ (ON11) was designed to exclude the formation of an offset duplex as will be discussed in more detail in section 3.6.5.

3.1.1 Oligonucleotide Synthesis

The oligonucleotides were synthesised by Bresagen using the automated β-cyanoethyl phospharamidite solid phase method. (Sinha *et al.*, 1984; Caruthers *et al.*, 1987) In this method, the bases are protected by the groups shown in figure 3.2, *i.e.* benzoyl groups on the adenine and cytosine bases and isobutyryl groups on the guanine bases. The cyanoethyl group protects the reactive phosphate oxygen during the coupling step. The oligonucleotide is attached to a controlled glass support (CPG) via an N-succinimide linkage.

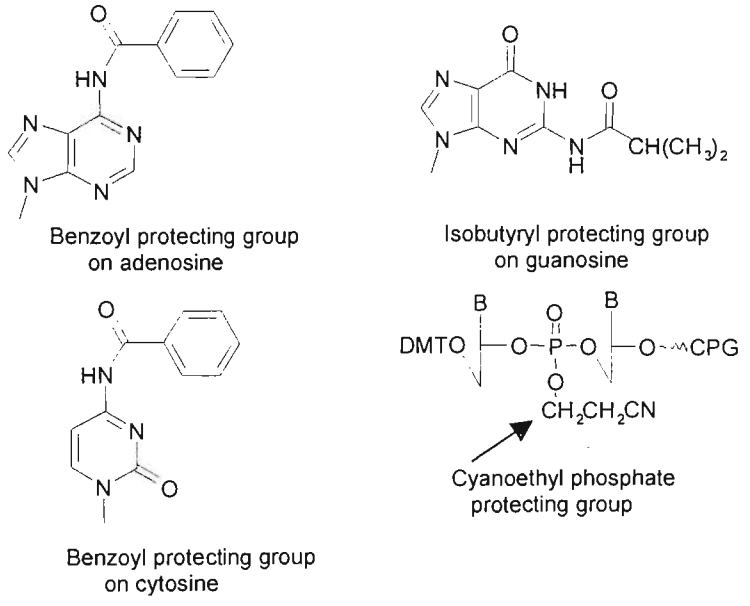


Figure 3.2: Protecting groups used for the bases in oligonucleotide synthesis.

The major steps in the phosphoramidite method are shown in figure 3.3. (Sinha *et al.*, 1984; Caruthers *et al.*, 1987) The first stage in the synthesis involves the removal of the DMT-protecting group from the 5′-end of the nucleotide that is attached to the CPG support. The free hydroxyl group may then participate in a coupling step with the 3′-nucleoside, which occurs as a result of nucleophilic attack. The next stage of synthesis involves the capping of any unreacted nucleotides which are attached to the solid support, *i.e.* the failure sequences.

This prevents these failure sequences from further reaction with nucleotides in subsequent coupling steps. The last stage in the synthesis cycle involves the oxidation of the internucleotide phosphite linkage formed in the coupling step to a phosphate triester. The above cycle is then repeated for every addition of a base to the oligonucleotide chain. The 5'-DMT protecting group is generally left on for the purpose of purification as discussed in section 3.1.3.

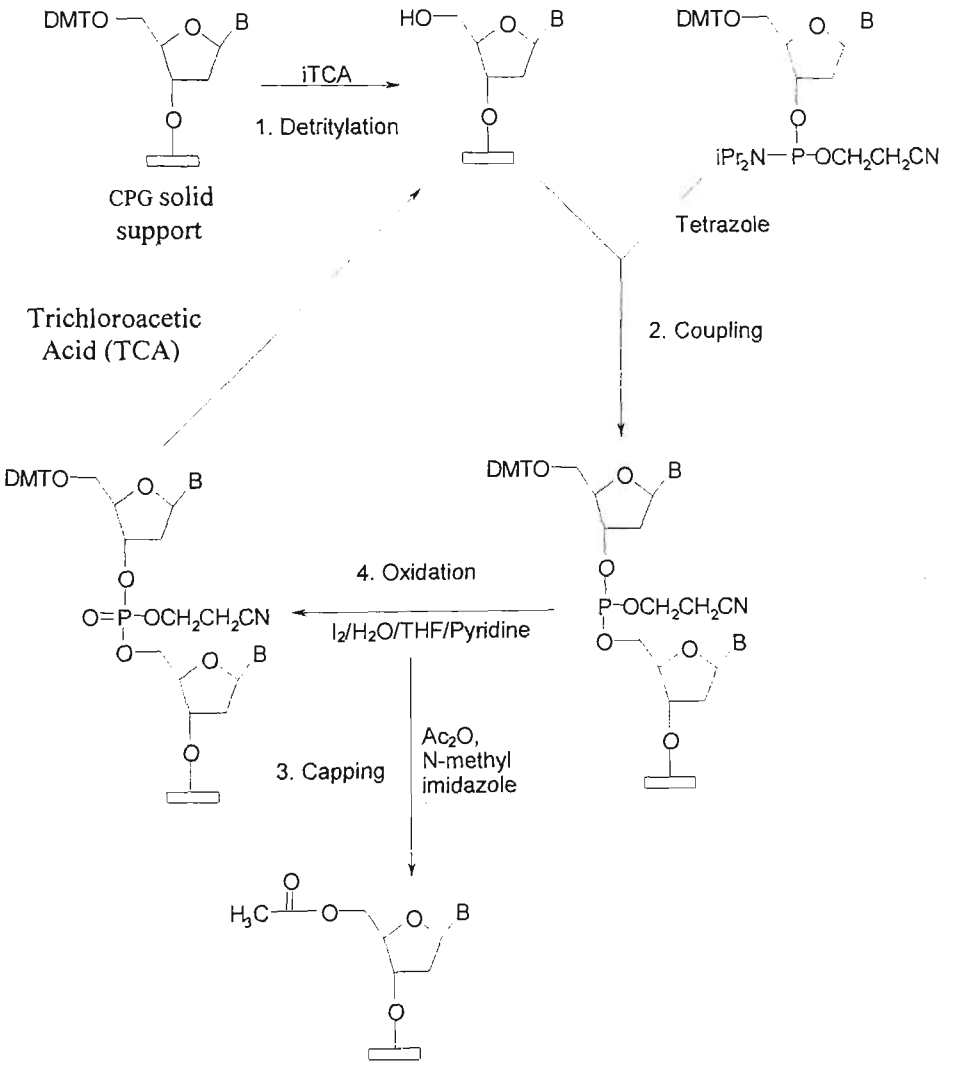


Figure 3.3: Phosphoramidite Synthesis Cycle.

3.1.2 Cleavage and Deprotection

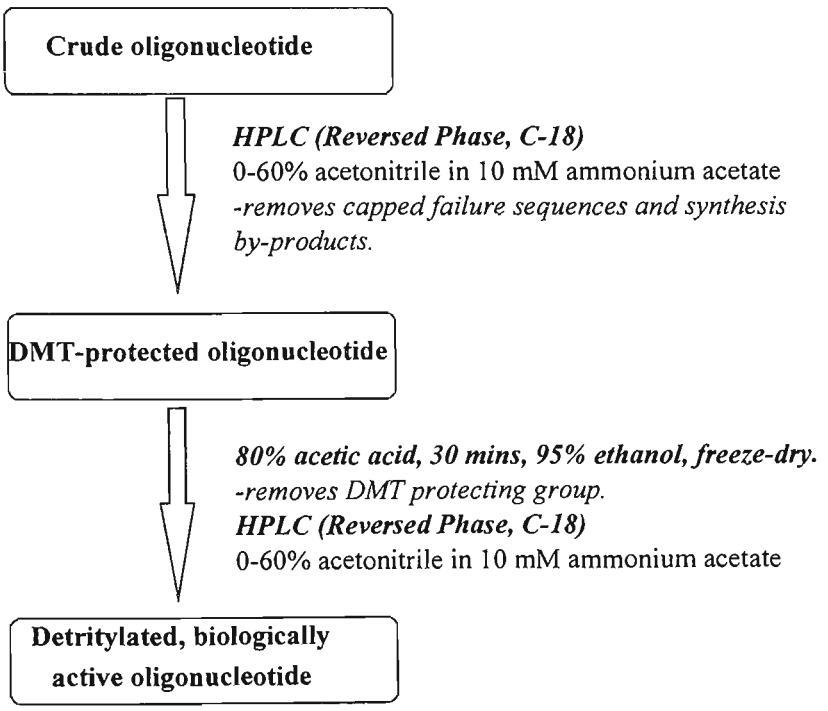
As noted above, the oligonucleotides were received from Bresagen still attached to the CPG support and fully protected. Cleavage of the oligonucleotide from the solid support was achieved by treatment with concentrated ammonia for one hour, which also removes the cyanoethyl protecting group. The process was repeated twice to ensure complete cleavage. The oligonucleotide was then heated in ammonia solution at 55°C for 15 hours to remove the exocyclic amine protecting groups. After complete deprotection, the solution was cooled and stored at –4 °C. Nitrogen gas was bubbled through the solutions to remove excess ammonia, which were then freeze-dried in a Savant Speedivac™ vacuum centrifuge.

3.1.3 HPLC Purification

The oligonucleotides were purified on a Beckman high performance liquid chromatography (HPLC) system (Sydney, Australia), which consisted of 7725I sample injection valve, 127s programmable solvent module and a UV visible 166 detector module set to measure absorbance at 254 nm. Purification was achieved by reversed phase HPLC using a C18 octadecasilyl column (8x100 mm Waters Deltapak Radial Pak cartridge).

The dry product, after removal of ammonia, was redissolved in a minimum of 10 mM ammonium acetate, pH 7.0. Purification of the oligonucleotide required two stages of HPLC as illustrated in scheme 3.1. The DMT protecting group was retained on the oligonucleotide for the first stage of purification by HPLC. This enhances separation of the oligonucleotide, which is hydrophobic owing to the presence of the trityl group, from the capped failure sequences and other synthesis impurities, which are not tritylated.

Scheme 3.1: Scheme showing the steps involved in the purification of oligonucleotides.



In the first stage of HPLC, a linear gradient of 0 to 60% acetonitrile in 10 mM ammonium acetate over 30 minutes was used. The "trityl-on" fraction was then collected and freeze-dried.

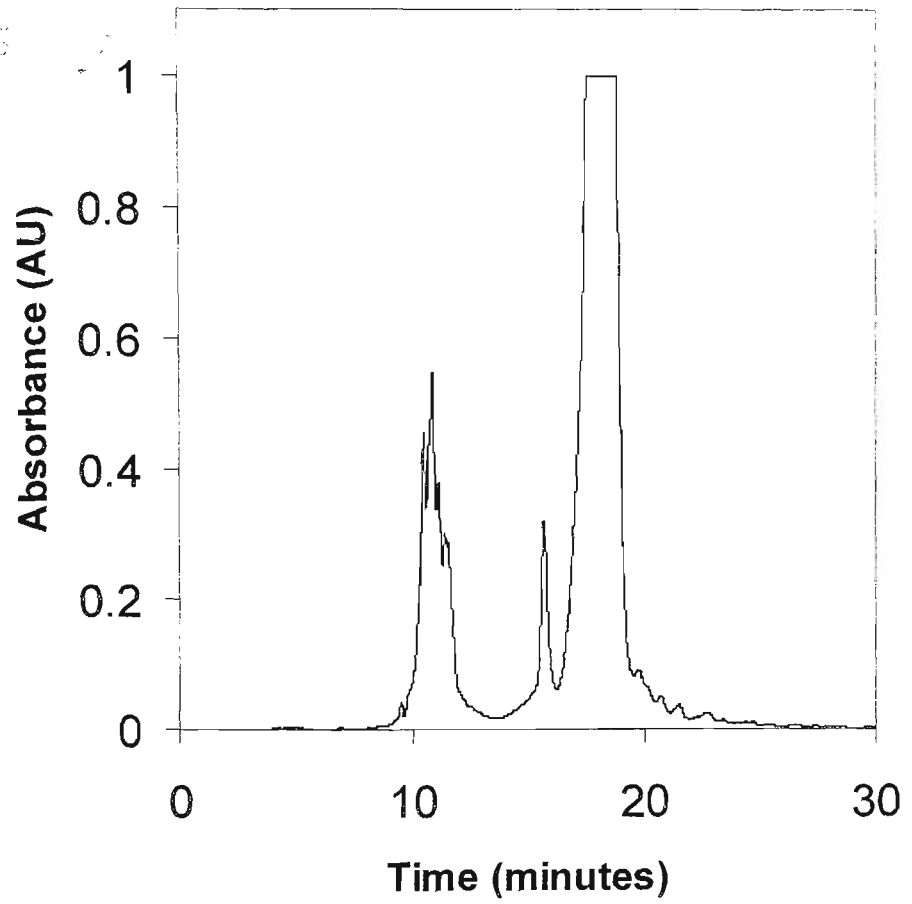


Figure 3.4: HPLC profile of the crude, protected oligonucleotide 5′-DMT-d(CGCAAATTTGCG)-3′.

Figure 3.4 shows the HPLC profile of the protected oligonucleotide 5′-DMT-CGCAAATTTGCG-3′. The cluster of peaks occurring at retention times between 10 and 12 minutes arise from the capped failure sequences and synthesis by-products. There may also have been peaks in this region owing to any of the desired oligonucleotide that has been detritylated. The peak at approximately 16 minutes is thought to be benzamide, which is produced when the benzoyl protecting group is cleaved during the deprotection step. The major peak, which elutes at a time of 18 minutes, is from the desired 5′-DMT-

CGCAAATTTGCG-3′. The difference in elution times between the protected and deprotected oligonucleotides is a consequence of the hydrophobic nature of the DMT group and its increased affinity for the non-polar reverse phase column.

The oligonucleotides were detritylated by treating the dried fraction with 80% acetic acid at room temperature for 30 minutes. After this time, an equal volume of ethanol was added to the solution and the sample was again freeze-dried. The detritylated oligonucleotide was purified by RP-HPLC to remove any remaining tritylated oligonucleotide and DMT-cations on the same column and HPLC system that was used in the first stage. A linear gradient of 0 to 60% acetonitrile with 10 mM ammonium acetate buffer at a flow rate of 1 mL/min over 30 minutes was used in this stage of purification. The detritylated oligonucleotide fractions were collected, freeze-dried and stored in the freezer prior to analysis by electrospray ionisation mass spectrometry (ESI-MS).

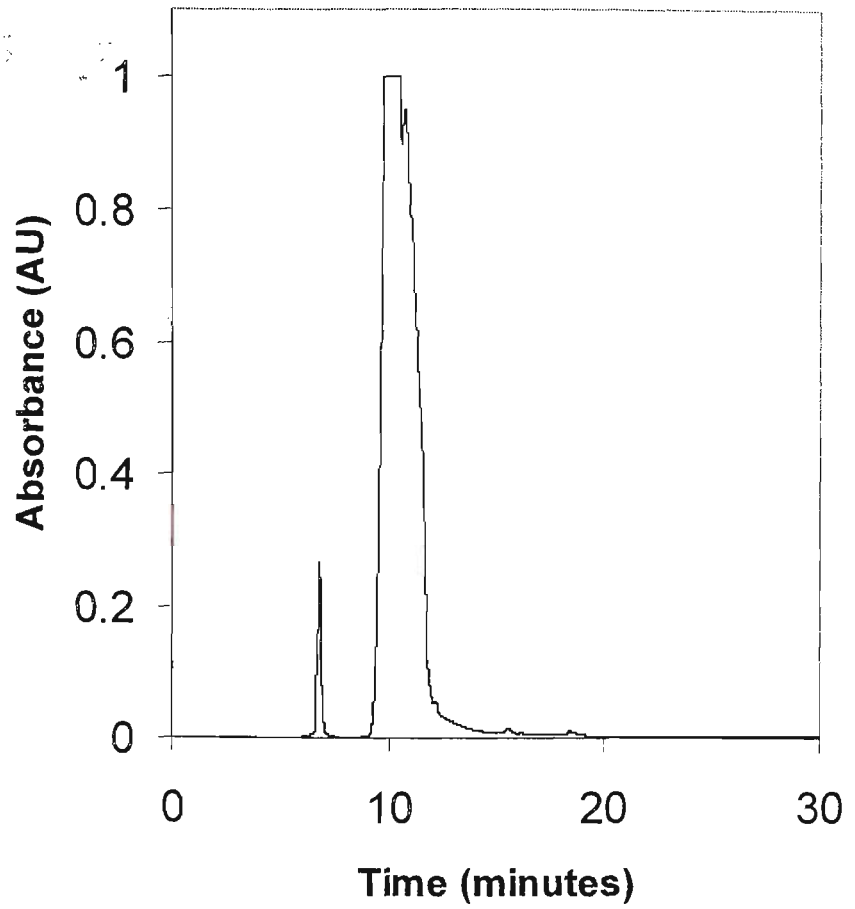


Figure 3.5: HPLC profile of the detritylated oligonucleotide 5′-d(CGCAAATTTGCG)-3′

Figure 3.5 shows the HPLC profile for the detritylated oligonucleotide, 5′-CGCAAATTTGCG-3′. The chromatogram consists of one major peak eluting at a retention time of 12 minutes corresponding to the detritylated oligonucleotide. The absence of a small peak at ~18 minutes owing to any remaining tritylated oligonucleotide, indicates complete detritylation occurred in this instance.

All the remaining oligonucleotides used in the experiments described in this chapter gave very similar HPLC profiles for both "trityl-on" and "trityl-off" purifications, with the exception that the retention times varied slightly with sequence. The purity of each of the

oligonucleotides after the second stage of HPLC was confirmed by examining the resultant product by ESI-MS on the VG Quattro triple quadrupole mass spectrometer.

3.1.4 Concentrations of Oligonucleotide Solutions

The concentration of the solutions of purified oligonucleotides was estimated by measuring the UV absorbance on a model 265 Shimadzu UV Visible spectrophotometer.

The purine and pyrimidine bases strongly absorb light with maxima near 260 nm. The concentrations were determined using Beers Law by measurement of UV absorbance at 260 nm (in a 1 cm quartz cuvette) using estimates of the molar extinction coefficient (ε_{260}) based on the proportion of adenine, guanine, cytosine and thymine with ε_{260} of 15400, 11700, 7300 and 8300 M^{-1} cm^{-1} respectively. (Sambrook *et al.*, 1989) The freeze-dried oligonucleotide samples were redissolved in a known volume of distilled, deionised water and diluted within the absorbance range of 0.2-1.2 so that errors owing to deviations from Beer's Law were minimised. (Sambrook *et al.*, 1989) The absorbance was measured at 260 nm in a 1 cm quartz cuvette. The oligonucleotide stock solutions prepared with a concentration of 100 pmol/μL in 50% aqueous acetonitrile containing 1% ammonia solution (pH~9.0).

3.2 Ligand-Oligonucleotide Binding Reactions

For reactions between pyrindamycin A and the oligonucleotides, the ligand (Pyrindamycin A) was dissolved in a minimum amount of dimethylformamide (DMF) and the resulting solution was added to the 1.0 mM oligonucleotide solution, in 0.1 M Na_2HPO_4/NaH_2PO_4

buffer (pH~8.0), in a 5:1 ratio. The reaction mixtures were kept under nitrogen in the dark at room temperature for 7 days.

The adducts were further purified by HPLC using the same conditions as for oligonucleotide purification except the gradient from 0 to 60% acetonitrile with 10 mM ammonium acetate buffer was over a longer time (40 mins). This additional step of HPLC was utilised to remove the unreacted products and the high concentration of sodium used in the reaction mixture. The HPLC did not give sufficient resolution to yield fractions containing only one adduct and as such the fractions were then pooled and analysed by ESI-MS.

3.3.1 Electrospray Ionisation Mass Spectrometry (ESI-MS) Studies

The electrospray ionisation mass spectra were obtained on a VG Biotech Quattro™ mass spectrometer (VG Biotech Ltd. (now Micromass), Altrincham, UK) equipped with an electrospray ionisation source and a quadrupole/hexapole/quadrupole mass analyser, which has a mass range up to *m/z* 4000. The instrument was upgraded to meet Quattro II™ specifications prior to the commencement of this study. A schematic diagram of the triple quadrupole mass spectrometer is shown in Figure 3.6.

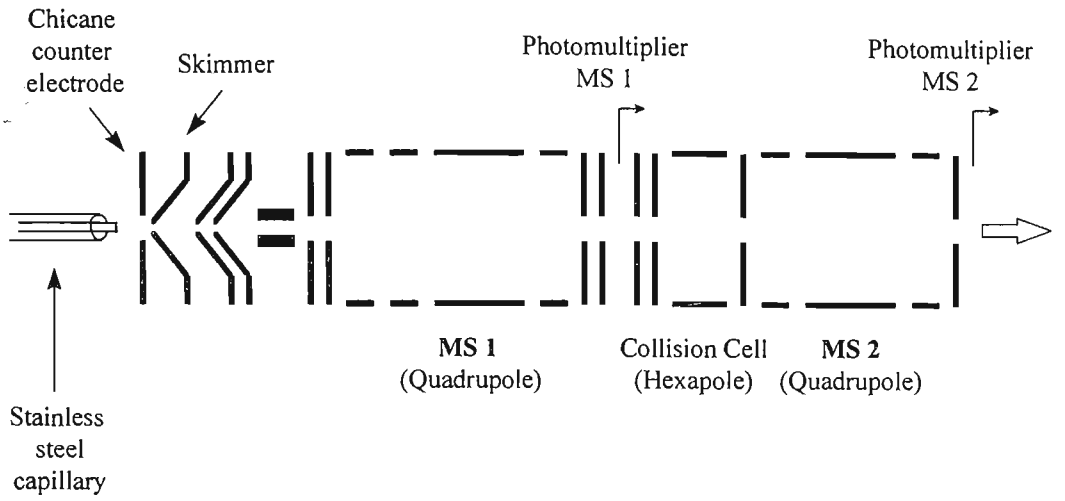


Figure 3.6: Schematic illustration of the triple quadrupole mass spectrometer showing the quadrupole/hexapole/quadrupole mass analyser employed.

ESI mass spectra were acquired for oligonucleotides in 50% aqueous acetonitrile. The stock oligonucleotide solutions (100 pmol/μL) were dissolved in this solvent and subsequently diluted to give a concentration of single-stranded DNA of approximately 5 pmol/μL for conventional electrospray. The solutions were directly infused via a Harvard syringe pump, model 55-1111 (Holliston, MA) at a flow rate of 5 μL/min. Total sample consumption for a typical experiment was 125 pmol. No attempt was made in this study to explore the absolute sensitivity of the technique, which could have been improved via the use of nanoelectrospray or capillary LC-MS on the Q-TOF, as it was deemed unnecessary.

Mass spectra were acquired over a range of 400–1000 *m/z* which were calibrated using a solution of sodium iodide at a concentration of 3 mg/mL. Calibration was undertaken after all other data had been acquired so as to avoid the introduction of sodium ions into the

ionisation source prior to the analysis of the oligonucleotide samples. Datafiles were then recalibrated accordingly.

A DEC Personal Computer workstation equipped with Micromass Mass Spectrometry Data System software MassLynx™ (version 3.1) was used for all data acquisition and processing.

The temperature of the electrospray ionisation source was maintained at 80°C. The operating voltage of the electrospray capillary was set at -3 kV for negative ions and 3 kV for positive ions. ESI mass spectra were typically recorded with a cone voltage of 25V. All ESI mass spectra were obtained by scanning in the multi-channel analysis (MCA) mode with a scan time of 1 second per *m/z* 100. The resulting spectra presented in this thesis are generally the sum of 10-15 scans. The ESI mass spectra presented here have been background subtracted and smoothed using standard algorithms provided in the Mass Lynx ™ software. Other MS parameters are given in table 3.1.

Table 3.1: Typical tuning parameters for ESI-MS on the VG Quattro mass spectrometer.

Source parameter	-ve ion mode	Comments
Capillary	3.0-3.5 kV	
Cone	20-50 V	sample dependent
HV Lens	0.1 kV	
Lens 3	1.0 V	
Source Temperature	80 °C	
LM Resolution	15.0	arbitrary units
HM Resolution	15.0	arbitrary units
Ion Energy	1.0	arbitrary units
Ion Energy Ramp	0.0	
Lens 5	30-50 V	
Lens 6	-2.0-8.0 V	sample dependent
MS1 Photomultiplier	650 V	

3.3.2 Electrospray IonisationTandem Mass Spectrometry (ESI-MS/MS) Studies

Tandem mass spectrometry experiments in negative ion mode were undertaken on the triple quadrupole mass spectrometer using 5 pmol/μL solutions of the products resulting from the reactions between the ligands and the various oligonucleotides. An ESI mass spectrum was obtained first to determine if the ion of interest was present. This ion was then viewed on a tuning monitor and optimised in MS1 before being selected for MS/MS in MS2. A collision energy ramp was utilised with high collision energy at low *m/z* and low collision energy at high *m/z*. This involved scanning *m/z* 250-1250 while the collision energy was ramped

from 50-10 eV. The collision gas (argon) was introduced into the collision cell initially at low pressure ($1x10^{-4}$ mbar) and gradually increased until the point where the intensity of the ion of interest had dropped to approximately 40% transmission and product (or daughter) ions were observed. Typically, a collision gas pressure of ~$1x10^{-3}$ mbar was sufficient to yield good levels of fragmentation in the MS/MS spectra of ions of interest in this study.

MS/MS data were acquired in product (or daughter) ion mode over the range *m/z* 250-1250. A scan time of 10 seconds with an interscan time of 0.10 seconds was used. Typically, 10-20 scans were summed to give a final spectrum.

The mass spectrometer was tuned in negative ion mode with a solution of the oligonucleotide to be analysed at a concentration of 5 pmol/μL. Typical tuning parameters for MS/MS experiments are given in table 3.2.

The mass range was calibrated with a sugar mixture for negative ions and with a solution of sodium iodide for positive ions. The sugar mixture consisted of corn syrup (maltooligosaccharide), raffinose, maltose and maltotetraose at a concentration of 100 pmol/μL in 50% aqueous acetonitrile. The sodium iodide solution had a concentration of 3 mg/mL.

Table 3.2: Typical tuning parameters for ESI-MS/MS on the VG Quattro mass spectrometer.

Source parameter	**-ve ion mode**	**Comments**
MS1:		
LM Resolution	12.0	arbitrary units
HM Resolution	12.0	arbitrary units
MS2:		
LM Resolution	10.0	arbitrary units
HM Resolution	10.0	arbitrary units
Ion Energy	0.5	arbitrary units
Ion Energy Ramp	0.0	
Lens 7	~330 V	
Lens 8	~280 V	
Lens 9	~15 V	
Photomultiplier	650 V	

3.4 Nomenclature

Throughout this thesis, the letter 'M' is be used to denote the neutral form of the oligonucleotide being examined. The designation ON1 through ON11 is given in parentheses after each oligonucleotide sequence to facilitate cross-referencing between data in tables and the discussion in the text. Standard nomenclature *i.e.* adenine (A), guanine (G), cytosine (C) and thymine (T) is used for the bases. All of the oligonucleotides used in this study had hydroxyl groups present at each terminus. The anti-tumour agents, pyrindamycin A and B, are abbreviated to PyA and PyB in their free state, whilst in bound

form, *i.e.* after the loss of HCl, they are abbreviated to Py. For example, the $[M+Py-4H]^{4-}$ species represents a pyrindamycin bound to the intact oligonucleotide with a charge of 4- arising from loss of 4 protons from a neutral oligonucleotide backbone, whilst $[M-A-3H]^{3-}$ represents a triply charged ion of a depurinated oligonucleotide.

For the designation of product (or sequence) ions we have used a variation of the system for nomenclature for oligonucleotide fragmentation introduced by McLuckey and co-workers, (McLuckey & Habibi-Goudarzi, 1993) whereby the oligonucleotide is assumed to be uncharged. (Boschenok & Sheil, 1996; Barry *et al.*, 1995) Depurination of the oligonucleotide occurs following alkylation. Subsequent cleavage of the 3′-C-O bond in the depurinated residue results in the formation of a w-ion when the charge is retained by the 3′ fragment and the corresponding 5′-cleavage product is designated as an $(a\text{-}B_n)$-ion, where B_n denotes the base lost as shown in scheme 2.1 (chapter two).

Table 3.3(a): Major species observed in ESI mass spectra of the reaction mixtures of pyrindamycins and oligonucleotides.

	Calculated *m/z*[a]			
Species	**ON 1, 2, 8, 9, 10[b]**	**ON 3, 4, 5, 7**	**ON 6**	**ON 11**
[M]	3645.4	3644.4	3643.4	4263.8
$[M-5H]^{5-}$	728.1	727.9	727.7	851.8
$[M-4H]^{4-}$	910.3	910.1	909.9	1064.9
$[M-3H]^{3-}$	1214.1	1213.8	1213.5	1420.3
[M+Py]	4152.9	4151.9	4151.0	4771.3
$[M+Py-5H]^{5-}$	829.6	829.4	829.2	953.3
$[M+Py-4H]^{4-}$	1037.2	1037.0	1036.7	1191.8
$[M+Py-3H]^{3-}$	1383.3	1383.0	1382.6	1589.4
[M-A]	3510.3	3509.3	3508.3	4112.7
$[M-A-5H]^{5-}$	701.1	700.9	700.7	-
$[M-A-4H]^{4-}$	876.6	876.3	876.1	-
$[M-A-3H]^{3-}$	1169.1	1168.8	1168.4	-
$[Py+A-H]^{-}$	641.6	641.6	641.6	-
[M-G]	3494.3	-	-	4112.7
$[M-G-5H]^{5-}$	697.9	-	-	821.5
$[M-G-4H]^{4-}$	872.6	-	-	1027.2
$[M-G-3H]^{3-}$	1163.8	-	-	1369.9
$[Py+G-H]^{-}$	657.6	-	-	657.6

a. Calculated m/z values are given since there are small variations in m/z (0.1-0.5) between individual spectra

b. ON 1 = 5′-d(CGCAAATTTGCG)-3′ ON 2 = 5′-d(CGGAAATTTCCG)-3′
ON 3 = 5′-d(CGAAAATTTTCG)-3′ ON 4 = 5′-d(GCAAAATTTTGC)-3′
ON 5 = 5′-d(CGTAAATTTACG)-3′ ON 6 = 5′-d(CTAAAATTTTAG)-3′
ON 7 = 5′-d(CGTTTATAAACG)-3′ ON 8 = 5′-d(CGCAAAGCTTTG)-3′
ON 9 = 5′-d(GCGTTTCGAAAC)-3′ ON 10 = 5′-d(CGAAAGCTTTCG)-3′
ON 11 = 5′-d(CGCAAAGCTTTGGC)-3′

Table 3.3(b): Major species observed in ESI-MS and MS/MS spectra of the reaction mixtures of pyrindamycins and oligonucleotides.

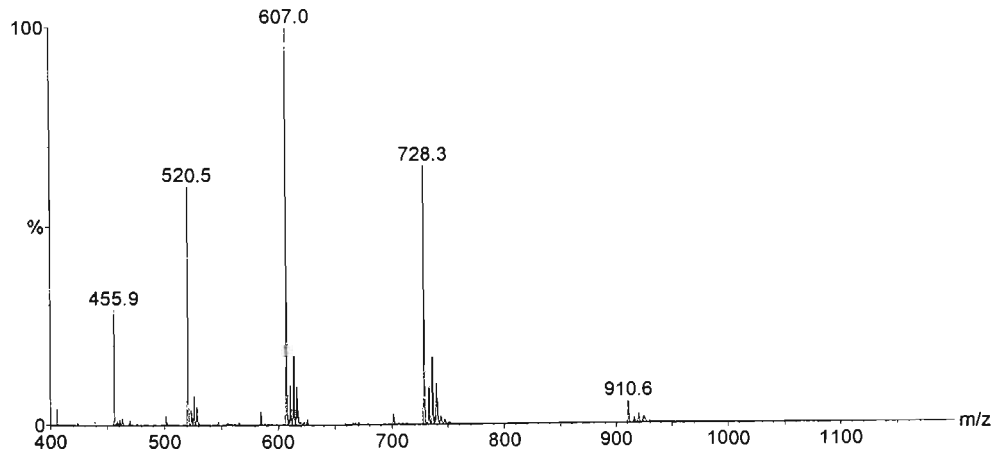
Species	ON 1	ON 2	ON 3, 4	ON 5	ON 6
$[w_1]^-$	-	-	-	-	346.2
$[w_2]^-$	-	-	-	635.4	-
$[w_6\text{-}2H]^{3-}$	625.1	611.7	616.7	619.7	624.7
$[w_6\text{-}H]^{2-}$	938.1	918.1	925.6	930.1	937.6
$[w_7\text{-}2H]^{3-}$	-	-	721.1	-	-
$[w_7\text{-}H]^{2-}$	-	-	1082.2	-	-
$[a_5\text{-}B_5\text{-}4H]^{3-}$	-	-	446.6	-	-
$[a_5\text{-}B_5\text{-}3H]^{2-}$	-	-	670.4	-	-
$[a_6\text{-}B_6\text{-}4H]^{3-}$	543.0	556.4	551.0	-	
$[a_6\text{-}B_6\text{-}3H]^{2-}$	815.0	835.1	827.1	-	
$[a_{10}\text{-}B_{10}\text{-}5H]^{4-}$	-	-	-	717.2	-
$[a_{10}\text{-}B_{10}\text{-}4H]^{3-}$	-	-	-	956.6	-
$[a_{11}\text{-}B_{11}\text{-}5H]^{4-}$	-	-	-	-	789.3
$[a_{11}\text{-}B_{11}\text{-}4H]^{3-}$	-	-	-	-	1052.7

Species	ON 7	ON 8	ON 9	ON 10	ON 11
$[w_1]^-$	-	-	306.2	-	-
$[w_2]^-$	635.4	-	-	-	-
$[w_5\text{-}2H]^{3-}$	-	515.3	-	-	-
$[w_5\text{-}H]^{2-}$	-	773.5	-	-	-
$[w_6\text{-}2H]^{3-}$	625.7	625.1	625.1	-	-
$[w_6\text{-}H]^{2-}$	939.1	938.1	938.1	-	-
$[w_7\text{-}2H]^{3-}$	-	-	-	721.5	721.5
$[w_7\text{-}H]^{2-}$	-	-	-	1082.7	1082.7
$[a_5\text{-}B_5\text{-}4H]^{3-}$	-	-	-	446.6	-
$[a_5\text{-}B_5\text{-}3H]^{2-}$	-	-	-	670.4	-
$[a_6\text{-}B_6\text{-}4H]^{3-}$	-	543.0	543.0	-	-
$[a_6\text{-}B_6\text{-}3H]^{2-}$	-	815.0	815.0	-	-
$[a_7\text{-}B_7\text{-}4H]^{3-}$	-	647.4	-	-	647.4
$[a_7\text{-}B_7\text{-}3H]^{2-}$	-	971.6	-	-	971.6
$[a_{10}\text{-}B_{10}\text{-}5H]^{4-}$	717.2	-	-	-	-
$[a_{10}\text{-}B_{10}\text{-}4H]^{3-}$	956.6	-	-	-	-
$[a_{11}\text{-}B_{11}\text{-}5H]^{4-}$	-	-	799.7	-	-
$[a_{11}\text{-}B_{11}\text{-}4H]^{3-}$	-	-	1066.7	-	-

Table 3.3 shows the calculated *m/z* values for the ions commonly observed in the ESI mass spectra of the reaction mixtures. The calculated *m/z* values for the sequence ions observed in both the ESI-MS and ESI-MS/MS spectra are given in table 3.3(b).

3.5 Electrospray Ionisation Mass Spectrometry (ESI-MS) Results

(a)



(b)

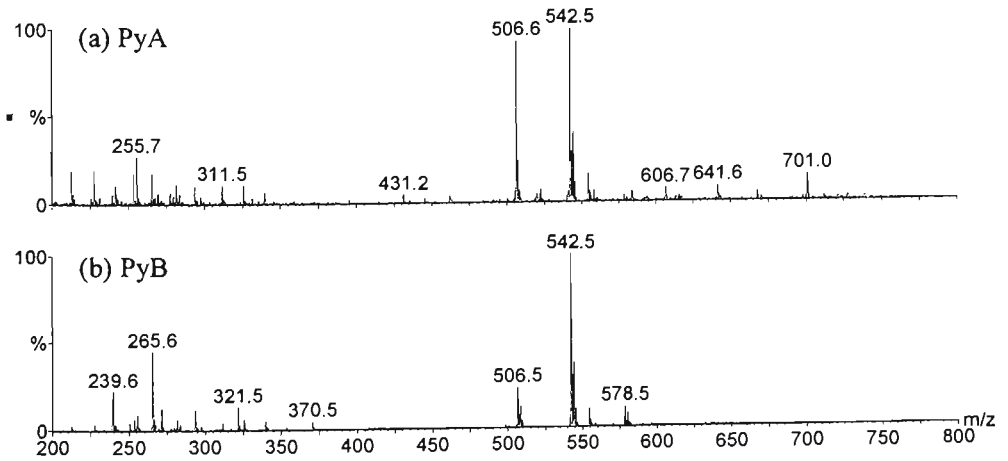
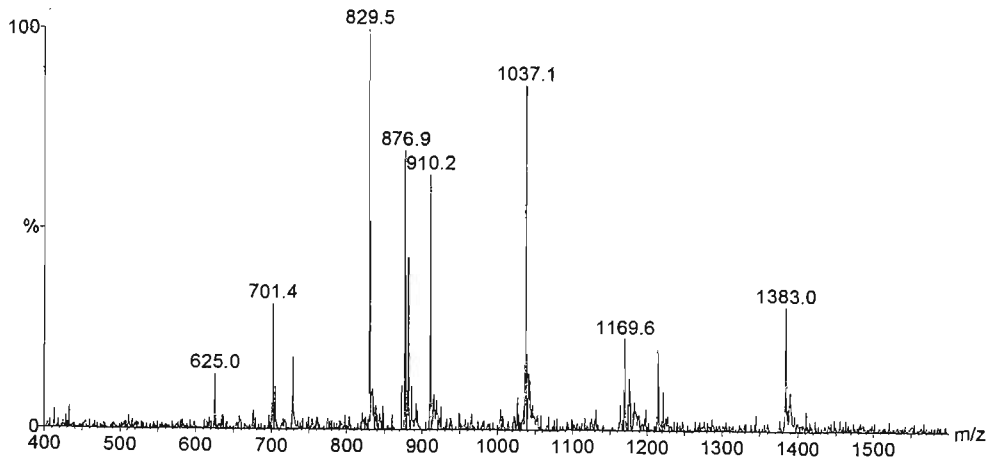


Figure 3.7: ESI mass spectra of: (a) 5′-d(CGCAAATTTGCG)-3′ (b) Pyrindamycin A (PyA) and Pyrindamycin B (PyB)

The ESI mass spectrum of the free oligonucleotide 5′-d(CGCAAATTTGCG)-3′ is shown in figure 3.7(a). The peaks at *m/z* 455.9, 520.6, 607.0, 728.4 and 910.6 represent the $[M-8H]^{8-}$, $[M-7H]^{7-}$, $[M-6H]^{6-}$, $[M-5H]^{5-}$ and $[M-4H]^{4-}$ ions of the free oligonucleotide. The ESI mass spectra of all oligonucleotides studied showed very similar spectra and were examined to ensure the purity of the DNA sequences after HPLC purification. Figure 3.7 also shows the ESI mass spectra recorded for the free ligands, pyrindamycin A and pyrindamycin B (figure 3.7b). In each case, the peaks at *m/z* ~542 represent the $[PyA-H]^-$ or $[PyB-H]^-$ ion of the unmodifed ligand, whilst the peaks observed at *m/z* 506 correspond to the hydrolysed form of the ligand, $[Py-H]^-$. This modification corresponds to a neutral loss of HCl with a *m/z* difference of 36.

(a)



(b)

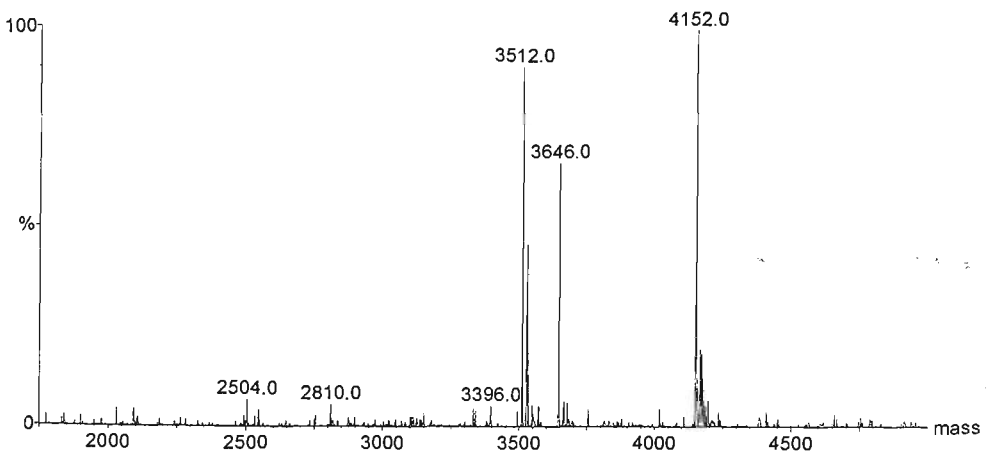
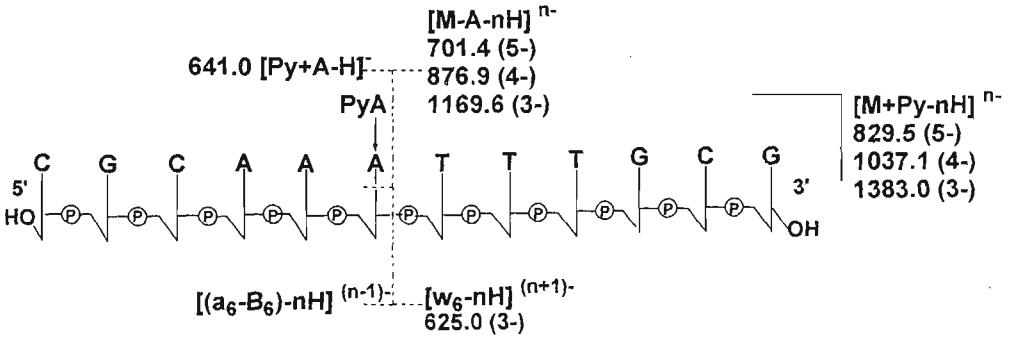


Figure 3.8: Electrospray ionisation mass spectrum of 5′-d(CGCAAATTTGCG)-3′ (ON1) and PyA: (a) *m/z* scale (b) mass scale

Figure 3.8 shows the ESI mass spectrum of the mixture formed on reaction between 5′-d(CGCAAATTTGCG)-3′ (ON1) and PyA on both an m/z scale (figure 3.8a) and converted to a mass scale (figure 3.8b). In the spectrum shown in figure 3.8a, there are peaks from the unreacted oligonucleotide at *m/z* 728.0 $[M-5H]^{5-}$, *m/z* 910.2 $[M-4H]^{4-}$ and *m/z* 1213.8 [M-

$3H]^{3-}$. The peaks at *m/z* 829.5, 1037.1, and 1383.0 represent the $[M+Py-5H]^{5-}$, $[M+Py-4H]^{4-}$ and $[M+Py-3H]^{3-}$ ions of the Py-oligonucleotide adduct. Fragmentation of this adduct results in the depurinated oligonucleotide, for which ions were present at *m/z* 701.4 $[M-A-5H]^{5-}$, *m/z* 876.9 $[M-A-4H]^{4-}$ and *m/z* 1169.6 $[M-A-3H]^{3-}$. The 3′-sequence ion $[w_6-2H]^{3-}$ (*i.e.* $[d(TTTGCG)+H_2O-3H]^{3-}$), *m/z* 625.0, was also observed in the ESI mass spectrum. The corresponding doubly charged ion *m/z* 938.1 was also present, but was of low intensity. Figure 3.8(b) provides a better representation of the relative abundance of the major species, showing the depurinated oligonucleotide at 3512 Da (there is also a small peak at 3534 Da owing to the corresponding sodium adduct of this species), the unreacted oligonucleotide at 3646 Da and the intact adduct at 4152 Da.



Scheme 3.2: Fragmentation of the $[M+Py-4H]^{4-}$ ion of 5′-d(CGCAAATTTGCG)-3′ (ON1) and PyA.

Scheme 3.2 shows the fragementation that was observed for 5′-d(CGCAAATTTGCG)-3′ (ON1). Pyrindamycin A alkylated the oligonucleotide at position 6 (A6) of the sequence, resulting in the labilisation of the 3′-glycosidic bond and cleavage as illustrated.

Table 3.4: Relative intensities of the major species observed in the ESI-MS spectra of pyrindamycin A oligonucleotide reaction mixtures.

	Relative Intensity (%)						
Species	**ON 1[a]**	**ON 2**	**ON 3**	**ON 4**	**ON 5**	**ON 6**	**ON 7**
$[M-5H]^{5-}$	18	-	-	5	47	8	20
$[M-4H]^{4-}$	64	-	-	6	30	5	11
$[M-3H]^{3-}$	20	-	-	-	7	-	3
$[M+Py-5H]^{5-}$	100	16	69	8	28	33	21
$[M+Py-4H]^{4-}$	86	100	100	100	100	100	82
$[M+Py-3H]^{3-}$	31	54	11	30	18	12	18
$[M-A-5H]^{5-}$	32	-	9	7	58	59	100
$[M-A-4H]^{4-}$	70	-	38	11	66	15	62
$[M-A-3H]^{3-}$	23	-	8	6	3	3	5
$[Py+A-H]^{-}$	-	17	15	22	39	47	31
$[w_2]^{-}$	-	-	-	-	16	-	6
$[w_6-2H]^{3-}$	14	2	30	-	-	-	-
$[w_6-H]^{2-}$	3	-	6	-	-	-	-
$[w_7-2H]^{3-}$	-	-	31	-	-	-	-
$[w_7-H]^{2-}$	-	-	5	-	-	-	-
				-	-	-	-
$[a_{10}-B_{10}-4H]^{3-}$	-	-	-	-	5	-	6
$[a_{11}-B_{11}-5H]^{4-}$	-	-	-	-	-	8	-
$[a_{11}-B_{11}-4H]^{3-}$	-	-	-	-	-	19	-

a. ON 1 = 5′-d(CGCAAATTTGCG)-3′ ON 2 = 5′-d(CGGAAATTTCCG)-3′
ON 3 = 5′-d(CGAAAATTTTCG)-3′ ON 4 = 5′-d(GCAAAATTTTGC)-3′
ON 5 = 5′-d(CGTAAATTTACG)-3′ ON 6 = 5′-d(CTAAAATTTTAG)-3′
ON 7 = 5′-d(CGTTTATAAACG)-3′

Table 3.4 shows the relative intensities of the major species and sequence ions observed in the ESI-MS spectra for reaction of pyrindamycin A with seven of the oligonucleotides studied. The ESI mass spectra of the remaining oligonucleotides were similar to that of the reaction mixture with ON1, with the exception that the relative intensities of the different species varied. For example, in the ESI mass spectrum of the mixture formed on reaction of 5′-d(CGGAAATTTCCG)-3′ (ON2) and PyA (not shown), the peaks arising from

depurinated species are absent (apart from $[Py+A-H]^-$) and only the $[w_6-2H]^{3-}$ sequence ion was present in very low abundance (2%).

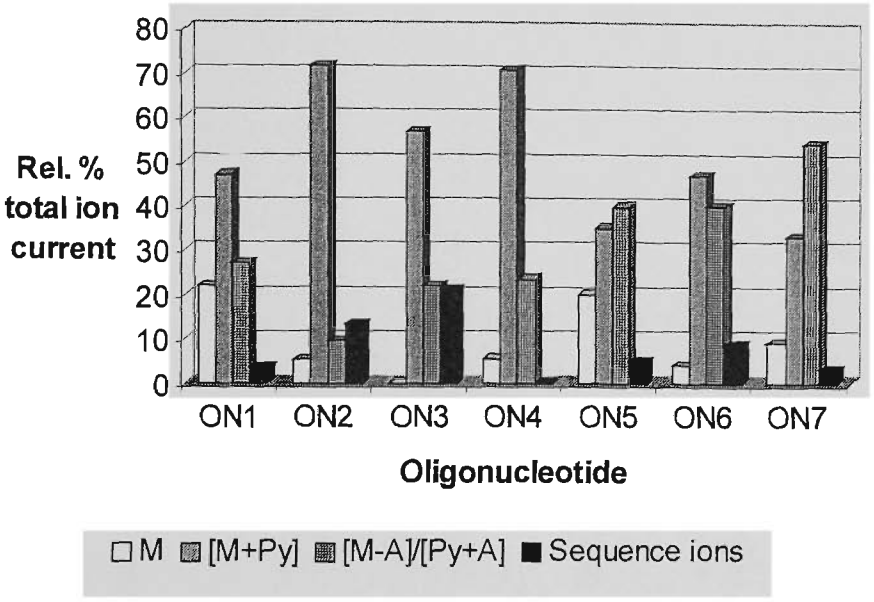
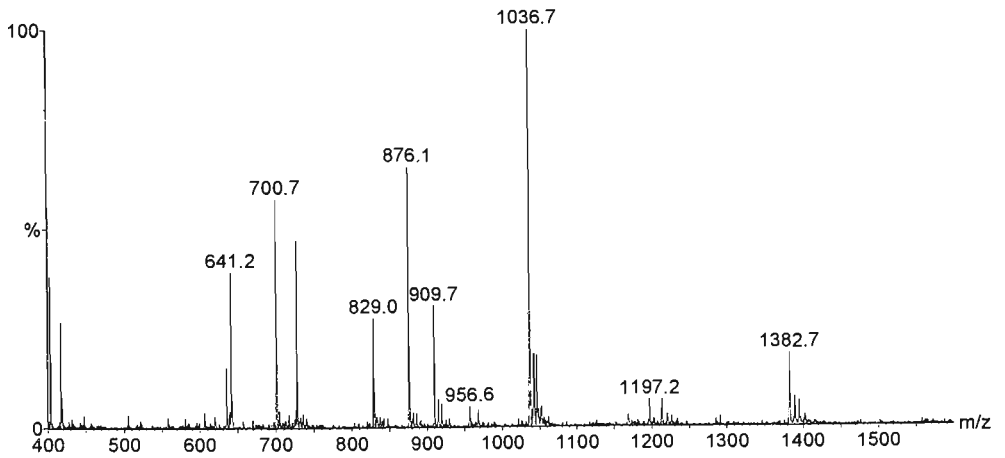


Figure 3.9: Relative intensity (% total ion current) of ions from the intact adduct (M+Py), the depurinated species (M-A)/(Py+A) and sequence ions compared to ions from the free oligonucleotide (M).

The differences between the relative intensity of ions from the intact (M+Py) adducts compared to ions from depurinated species (M-A) and (Py+A) and cleavage products (w and a-B ions) are demonstrated more clearly in figure 3.9. In the figure, the relative intensities of different ions from each species (*i.e.* $[M+Py-5H]^{5-}$ $[M+Py-4H]^{4-}$ and $[M+Py-3H]^{3-}$ for (M+Py) etc) are compared for each of the ESI mass spectra of the pyrindamycin A reaction mixtures with ON1-ON7. Some marked differences between the stabilities of different oligonucleotide adducts are evident from these data. For example, for ON2 (and, albeit to a lesser extent, ON1, ON3 and ON4) the (M+Py) adduct represents the major

species, whereas for ON5, ON6 and ON7, ions from depurinated species were of comparable or greater abundance to those of the intact adducts.

(a)



(b)

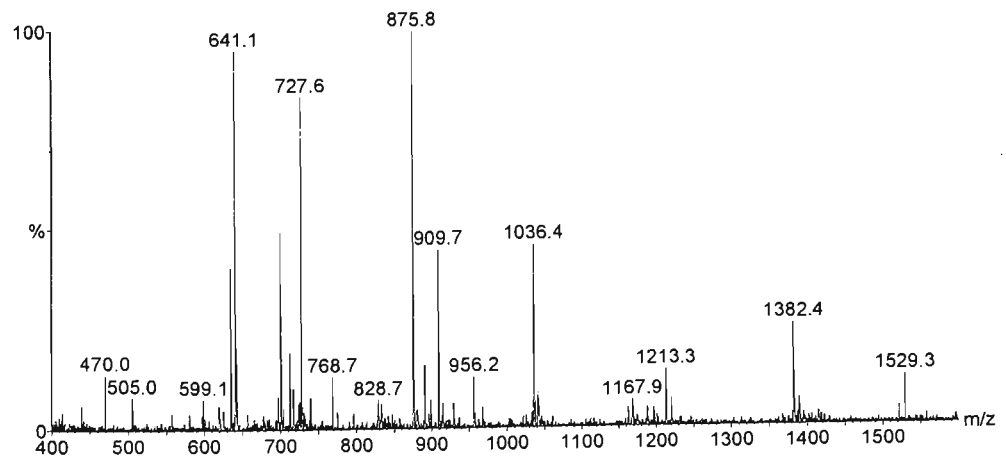


Figure 3.10: ESI mass spectrum of reaction mixture between 5′-d(CGTAAATTTACG)-3′ (ON5) and (a) pyrindamycin A and (b) pyrindamycin B

Figure 3.10 shows the ESI mass spectra of the mixture from the reaction of pyrindamycin A (figure 3.10a) and pyrindamycin B (figure 3.10b) each with 5′-d(CGTAAATTTACG)-3′ (ON5). There were a number of ions of different charge present for each species. For example, for the intact adduct (M+Py) there are peaks at *m/z* 829.0, 1036.7 and 1382.7 from the $[M+Py-5H]^{5-}$, $[M+Py-4H]^{4-}$ and $[M+Py-3H]^{3-}$ ions respectively. The peaks observed at *m/z* 727.7 and 909.7 represent the $[M-5H]^{5-}$ and $[M-4H]^{4-}$ ions of the intact oligonucleotide, whilst those at *m/z* 700.7 and 876.1 are from the $[M-A-5H]^{5-}$ and $[M-A-4H]^{4-}$ depurinated species. It is interesting to note that the $[w_2]^-$ and $[a_{10}-B_{10}-4H]^{3-}$ sequence ions are present at *m/z* 635.3 and 956.6 respectively. This indicates that the binding of the ligand occurred at the adenine at position 10 in the oligonucleotide rather than at the higher affinity site at A6 (more below).

In the spectrum of the reaction mixture of ON5 and pyrindamycin B (figure 3.10b), peaks are also observed owing to the intact adduct, free oligonucleotide, depurinated species and the cleavage products. It is notable, however, that upon reaction with pyrindamycin B, the intensity of the depurinated species (*m/z* 700.5, 875.8 and 1168.3), pyrindamycin-adenine adduct (*m/z* 641.1) and sequence ions (*m/z* 635.0 and 956.2) are the most intense. In contrast, the spectrum from the reaction with pyrindamycin A shows peaks of greater intensity for the intact adduct at *m/z* 829.0, 1036.7 and 1382.7. The ESI mass spectra for all of the oligonucleotides reacted with pyrindamycin B displayed very similar results and MS/MS spectra revealed identical fragmentation pathways. The two agents are thought to alkylate DNA via a common agent, duocarmycin A and the data collected is consistent with this hypothesis.

The ESI mass spectra of the remaining self-complementary oligonucleotides ON2, ON3, ON4, ON6 AND ON10 (not shown) were similar to those already presented here, with the exception that the relative intensities of the different species varied.

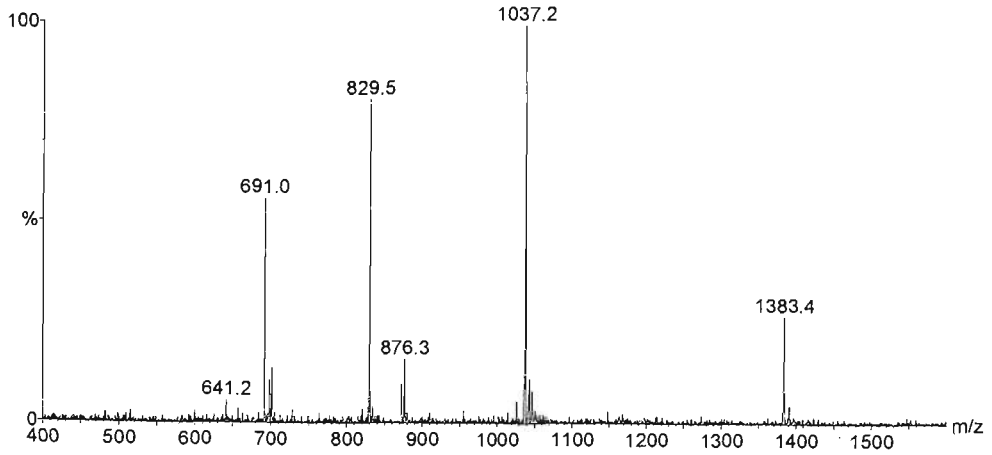


Figure 3.11: ESI mass spectrum of reaction mixture between 5′-d(CGCAAAGCTTTG)-3′ (ON8) and pyrindamycin A

Some differences were observed for the non-self-complementary sequences. For example, figure 3.11 shows the ESI mass spectrum of the mixture formed on reaction between the non-self-complementary sequence, 5′-d(CGCAAAGCTTTG)-3′ (ON8), and PyA. The ions at *m/z* 691.0, 829.5, 1037.2 and 1383.4 represent the $[M+Py-6H]^{6-}$, $[M+Py-5H]^{5-}$, $[M+Py-4H]^{4-}$ and $[M+Py-3H]^{3-}$ of the Py-oligonucleotide adduct. The depurinated oligonucleotide gives rise to the ions $[M-A-4H]^{4-}$ (*m/z* 876.3) and $[M-A-3H]^{3-}$ (*m/z* 1168.5). The adduct formed between pyrindamycin and adenine is observed at *m/z* 641.2 and a low intensity peak is also seen at *m/z* 657.2 which corresponds to the adduct formed after guanine

alkylation. The intact adduct was then analysed by MS/MS and the results of these experiments will be discussed in section 3.6.8.

3.6 Electrospray Ionisation Tandem Mass Spectrometry (MS/MS) Results

Since sequence ions are not observed for all adducts in the ESI mass spectra, tandem mass spectrometry (MS/MS) was necessary for definitive identification of the ligand binding sites. Hence, each of the major species observed in the ESI mass spectra of the reaction mixtures were analyzed further by MS/MS. Table 3.5 shows the major ions observed in the MS/MS spectra of the $[M+Py-4H]^{4-}$ and $[M-A-4H]^{4-}$ ions for each of the oligonucleotide adducts formed with pyrindamycin A and for adducts of ON1, ON2, ON3, ON5, and ON7 with pyrindamycin B. In cases where guanine alkylation was observed, the $[M+Py-4H]^{4-}$ and $[M-G-4H]^{4-}$ ions were also analysed by MS/MS. Additionally, for the 14-mer studied, 5′-d(CGCAAAGCTTTGGC)-3′, the $[M+Py-5H]^{5-}$ ion was also analysed by MS/MS as it was of much greater intensity in the ESI mass spectrum, and therefore was more readily analysed than the $[M+Py-4H]^{4-}$ ion. In each case the MS/MS spectra readily enabled the binding site(s) to be identified as described below.

Table 3.5(a): Precursor and product ions observed in the ESI-MS/MS spectra of pyrindamcin A and B with oligonucleotides ON1-ON6.

	Precursor Ions		Product Ions				Binding Site	
	Species	m/z^b	Species	m/z^b	Species	m/z^b	PyA	PyB
ON1[a]	$[M+Py-4H]^{4-}$	1037.1	$[M-A-3H]^{3-}$ $[M-A-4H]^{4-}$	1169.6 876.9	$[Py+A-H]^-$ $[Py+A]^0$	- c		
	$[M-A-4H]^{4-}$	876.9	$[w_6-H]^{2-}$	938.0	$[a_6-B_6-3H]^{2-}$	814.2	A6	A6
ON2	$[M+Py-4H]^{4-}$	1036.8	$[M-A-3H]^{3-}$ $[M-A-4H]^{4-}$	1168.9 876.6	$[Py+A-H]^-$ $[Py+A]^0$	640.6 c		
	$[M-A-4H]^{4-}$	876.6	$[w_6-H]^{2-}$	917.4	$[a_6-B_6-3H]^{2-}$	833.6	A6	A6
ON3	$[M+Py-4H]^{4-}$	1036.1	$[M-A-3H]^{3-}$ $[M-A-4H]^{4-}$	1167.3 875.6	$[Py+A-H]^-$ $[Py+A]^0$	641.0 c		
	$[M-A-4H]^{4-}$	875.6	$[w_7-H]^{2-}$ $[w_6-H]^{2-}$	1081.5 925.4	$[a_7-B_7-3H]^{2-}$ $[a_6-B_6-3H]^{2-}$	669.2 827.2	A5 A6	A5 A6
ON4	$[M+Py-4H]^{4-}$	1037.0	$[M-A-3H]^{3-}$ $[M-A-4H]^{4-}$	1168.0 875.4	$[Py+A-H]^-$ $[Py+A]^0$	641.2 c		
	$[M-A-4H]^{4-}$	875.4	$[w_7-H]^{2-}$ $[w_6-H]^{2-}$	1081.2 925.9	$[a_7-B_7-3H]^{2-}$ $[a_6-B_6-3H]^{2-}$	- -	A5 A6	d
ON5	$[M+Py-4H]^{4-}$	1036.5	$[M-A-3H]^{3-}$ $[M-A-4H]^{4-}$	1168.1 875.6	$[Py+A-H]^-$ $[Py+A]^0$	640.8 c		
	$[M-A-4H]^{4-}$	875.6	$[w_2]^-$	634.4	$[a_{10}-B_{10}-3H]^{2-}$	956.6	A10	A10
ON6	$[M+Py-4H]^{4-}$	1035.8	$[M-A-3H]^{3-}$ $[M-A-4H]^{4-}$	1167.8 876.2	$[Py+A-H]^-$ $[Py+A]^0$	640.9 c		
	$[M-A-4H]^{4-}$	876.2	$[w_1]^-$	346.4	$[a_{11}-B_{11}-3H]^{2-}$	1053.2	A11	d

a. ON 1 = 5′-d(CGCAAATTTGCG)-3′ ON 2 = 5′-d(CGGAAATTTCCG)-3′
ON 3 = 5′-d(CGAAAATTTTCG)-3′ ON 4 = 5′-d(GCAAAATTTTGC)-3′
ON 5 = 5′-d(CGTAAATTTACG)-3′ ON 6 = 5′-d(CTAAAATTTTAG)-3′

b. Observed m/z

c. This species is neutral and therefore not detected

d. These oligonucleotides were not reacted with pyrindamycin B

Table 3.5(b): Precursor and product ions observed in the ESI-MS/MS spectra of pyrindamcin A and B with oligonucleotides ON7-ON11.

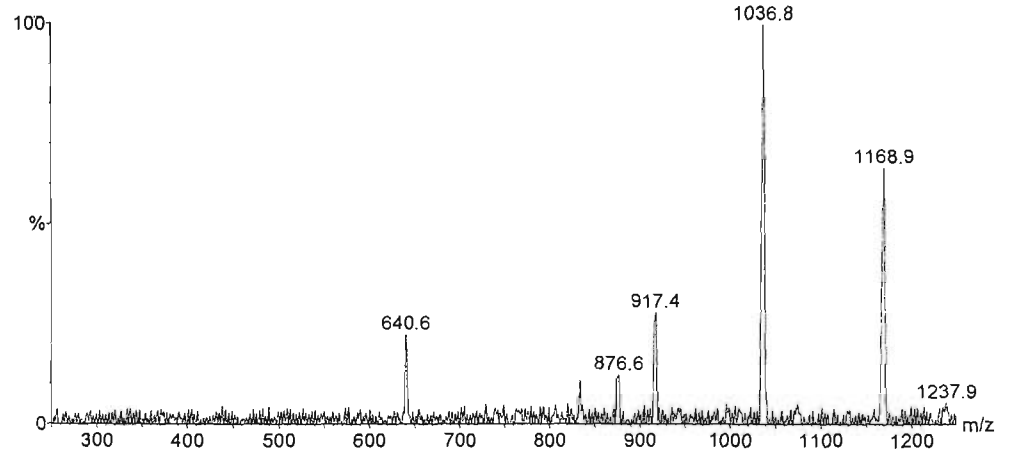
	Precursor Ions		Product Ions				Binding Site	
	Species	m/z^b	Species	m/z^b	Species	m/z^b	PyA	PyB
ON7[a]	$[M+Py-4H]^{4-}$	1036.1	$M-A-3H]^{3-}$ $[M-A-4H]^{4-}$	1167.5 875.2	$[Py+A-H]^-$ $[Py+A]^0$	641.3 c		
	$[M-A-4H]^{4-}$	875.2	$[w_2]^-$	635.0	$[a_{10}-B_{10}-3H]^{2-}$	956.4	A10	A10
ON8	$[M+Py-4H]^{4-}$	1036.7	$[M-A-3H]^{3-}$ $[M-A-4H]^{4-}$ $[M-G-3H]^{3-}$ $[M-G-4H]^{4-}$	1168.8 875.2 1162.3 871.6	$[Py+A-H]^-$ $[Py+A]^0$ $[Py+G-H]^-$ $[Py+G]^0$	641.3 c 656.8 c		
	$[M-A-4H]^{4-}$ $[M-G-4H]^{4-}$	875.2	$[w_6-H]^{2-}$ $[w_5-H]^{2-}$	937.0 773.1	$[a_6-B_6-3H]^{2-}$ $[a_7-B_7-3H]^{2-}$	814.5 971.3	A6 G7	A6 G7
ON8 & ON9	$[M+Py-4H]^{4-}$	1036.4	$[M-A-3H]^{3-}$ $[M-A-4H]^{4-}$	1168.4 877.4	$[Py+A-H]^-$ $[Py+A]^0$	640.8 c		
	$[M-A-4H]^{4-}$	877.4	$[w_6-H]^{2-}$ $[w_1]^-$	937.5 -	$[a_6-B_6-3H]^{2-}$ $[a_{11}-B_{11}-3H]^{2-}$	- 1066.8	A6(1) A11(2)	d
ON10	$[M+Py-4H]^{4-}$	1036.5	$[M-A-3H]^{3-}$ $[M-A-4H]^{4-}$	1168.6 874.5	$[Py+A-H]^-$ $[Py+A]^0$	641.3 c		
	$[M-A-4H]^{4-}$	874.5	$[w_7-H]^{2-}$	1081.1	$[a_5-B_5-3H]^{2-}$	-	A5	d
ON11	$[M+Py-5H]^{5-}$	952.6	$[M-G-4H]^{4-}$	1027.1	$[Py+A-H]^-$	657.5		
	$[M-G-4H]^{4-}$	1027.1	$[w_7-H]^{2-}$	1082.4	$[a_7-B_7-3H]^{2-}$	970.2	G7	d

a. ON 7 = 5′-d(CGTTTATAAACG)-3′ ON 8 = 5′-d(CGCAAAGCTTTG)-3′
ON 9 = 5′-d(GCGTTTCGAAAC)-3′ ON 10 = 5′-d(CGAAAGCTTTCG)-3′
ON 11 = 5′-d(CGCAAAGCTTTGGC)-3′

b. Observed m/z

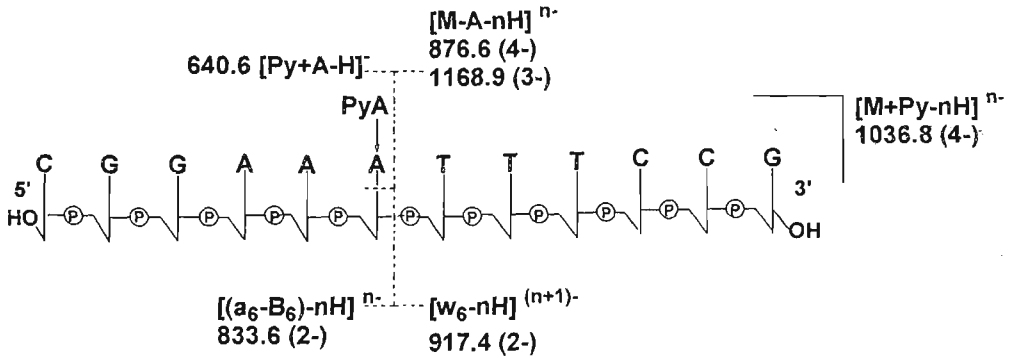
c. This species is neutral and therefore not detected

d. These oligonucleotides were not reacted with pyrindamycin B



5'-d(CGCAAATTTGCG)-3' (ON1) and 5'-d(CGGAAATTTCCG)-3' (ON2)

Figure 3.12: ESI-MS/MS spectrum of the $[M+Py-4H]^{4-}$ ion of the major adduct formed between pyrindamycin A and 5′-d(CGGAAATTTCCG)-3′ (ON2)

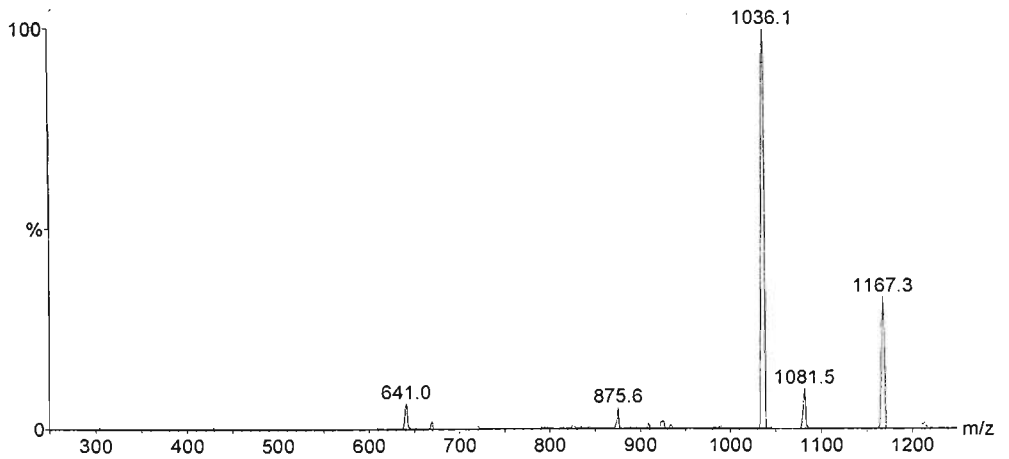


Scheme 3.3: Fragmentation of the $[M+Py-4H]^{4-}$ ion of 5′-d(CGGAAATTTCCG)-3′ (ON2) and PyA.

The MS/MS spectrum of the precursor ion, $[M+Py-4H]^{4-}$, formed in the reaction between pyrindamycin A and 5′-d(CGGAAATTTCCG)-3′ is shown in figure 3.12. The major pathway is depurination which gives rise to the $[M-A-4H]^{4-}$ and $[M-3H]^{3-}$ ions at *m/z* 876.6 and 1168.9 and the $[Py+A-H]^{-}$ ion at *m/z* 640.6. The $[M-A-4H]^{4-}$ ion then decomposes to

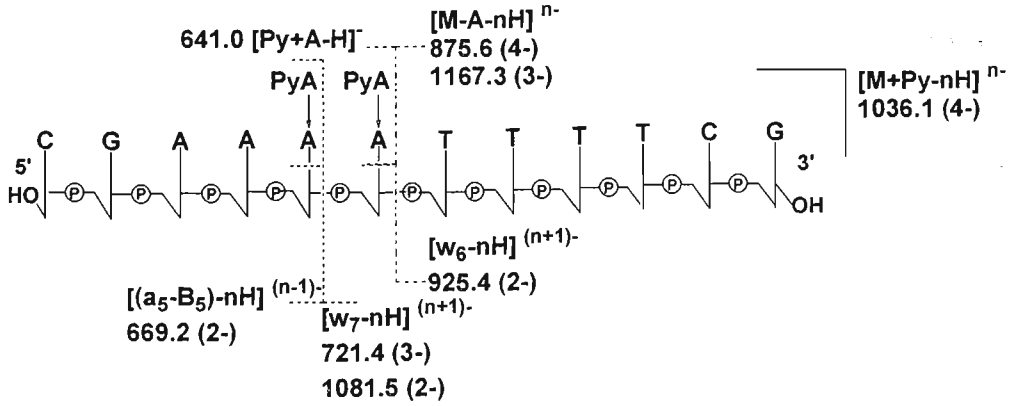
yield the sequence ions $[w_6\text{-}H]^{2-}$ and $[a_6\text{-}B_6\text{-}3H]^{2-}$ ions, which are observed at *m/z* 917.4 and 833.6 respectively. These data enables the identification of the site of alkylation to be the A6 as the subsequent labilisation of the 3′-glycosidic bond results in the formation of the above-mentioned sequence ions. It is notable that the spectrum of the $[M\text{-}A\text{-}4H]^{4-}$ species at *m/z* 876.6 showed the same sequence ions as that of the intact percursor $[M+Py\text{-}4H]^{4-}$. This provides support for the notion that loss of the [Py+A] either as a neutral or anion are intermediates in the fragmentation pathway(s) leading to the formation of the sequence ions.

For the oligonucleotides, 5′-d(CGCAAATTTGCG)-3′ (ON1) and 5′-d(CGGAAATTTCCG)-3′ (ON2), alkylation occurred at A6, the preferred binding site within the 5′-(AAA)-3′ sequence, based on earlier sequencing studies. (Boger *et al.*, 1990) With ON2, however, the intact adduct is the major ion in the ESI mass spectrum (table 3.4) indicating that the resulting adducts are more stable when there is a guanine in the fourth position preceding the binding site, rather than a cytosine.



5′-d(CGAAAATTTTCG)-3′ (ON3) and 5′-d(GCAAAATTTTGC)-3′ (ON4)

Figure 3.13: ESI-MS/MS spectrum of the $[M+Py-4H]^{4-}$ ion of the major adduct formed between pyrindamycin A and 5′-d(CGAAAATTTTCG)-3′ (ON3)



Scheme 3.4: Fragmentation of the $[M+Py-4H]^{4-}$ ion of 5′-d(CGAAAATTTTCG)-3′ (ON3) and PyA

In some cases, more than one binding site was evident from the MS/MS spectra. For example, figures 3.13 and 3.14 shows the ESI-MS/MS spectra of the $[M+Py-4H]^{4-}$ ion of

the major adduct formed when pyrindamycin A was reacted with 5′-d(CGAAAATTTTCG)-3′ (ON3) and 5′-d(GCAAAATTTTGC)-3′ (ON4) respectively. In these examples, there were three fragmentation pathways evident, namely decomposition to yield the $[M\text{-}A\text{-}3H]^{3-}$ ion (*m/z* ~1168) and $[Py+A]^{-}$ (*m/z* ~641) ion and loss of neutral (Py+A) to yield the $[M\text{-}A\text{-}4H]^{4-}$ ion (*m/z* ~875), which are not sequence specific. The latter ion decomposes to either the complementary $[w_6\text{-}H]^{2-}$ (*m/z* 926) and $[a_6\text{-}B_6\text{-}3H]^{2-}$ ions or the $[w_7\text{-}H]^{2-}$ (*m/z* 1081) and $[a_5\text{-}B_5\text{-}3H]^{2-}$ ions. These sequence ions show that the ligand is binding to both A6 and A5 respectively and therefore that the mass-selected precursor ions arise from two different positional isomers. The intensities of the w-ions resulting from the two binding sites were comparable, however, it would be unwarranted to attach too much significance to this in terms of the relative abundances of each adduct in the reaction mixture (more below). Binding to both these sites was also observed for the reactions of pyrindamycin B with the same oligonucleotide (*i.e.* ON3) and for pyrindamycin A with 5′-d(GCAAAATTTTGC)-3′ (ON4).

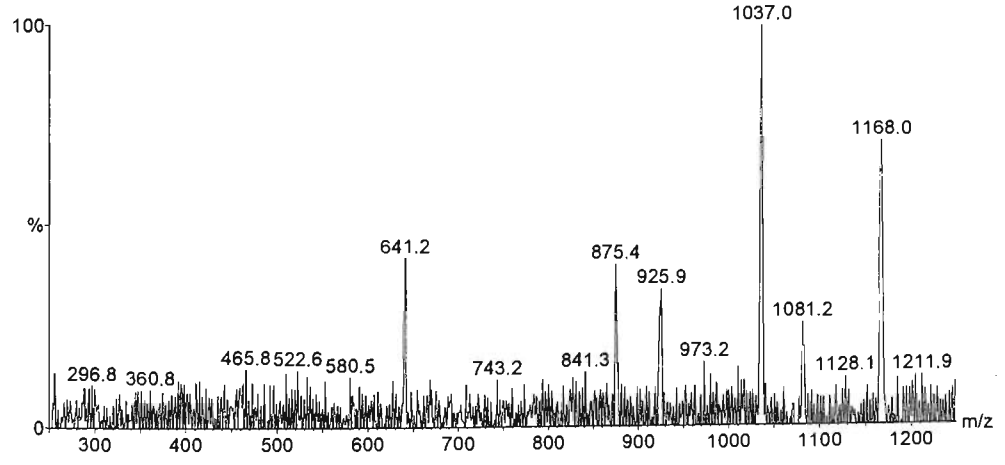
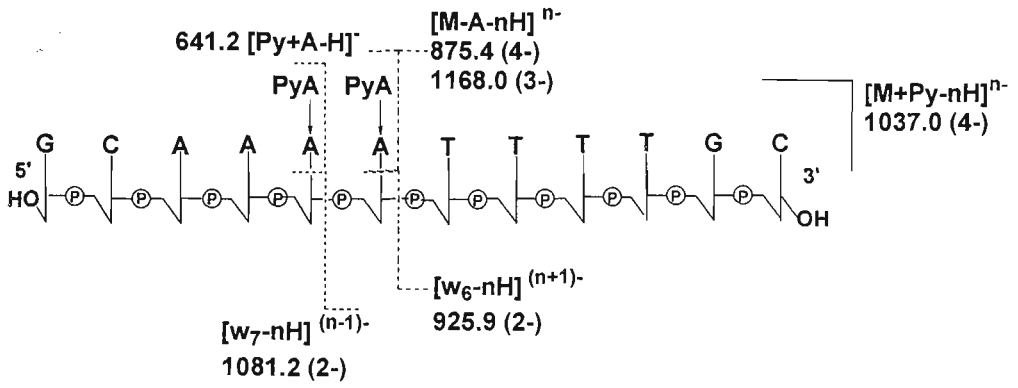


Figure 3.14: ESI-MS/MS spectrum of the $[M+Py\text{-}4H]^{4-}$ ion of the major adduct formed between pyrindamycin A and 5′-d(GCAAAATTTTGC)-3′ (ON4)

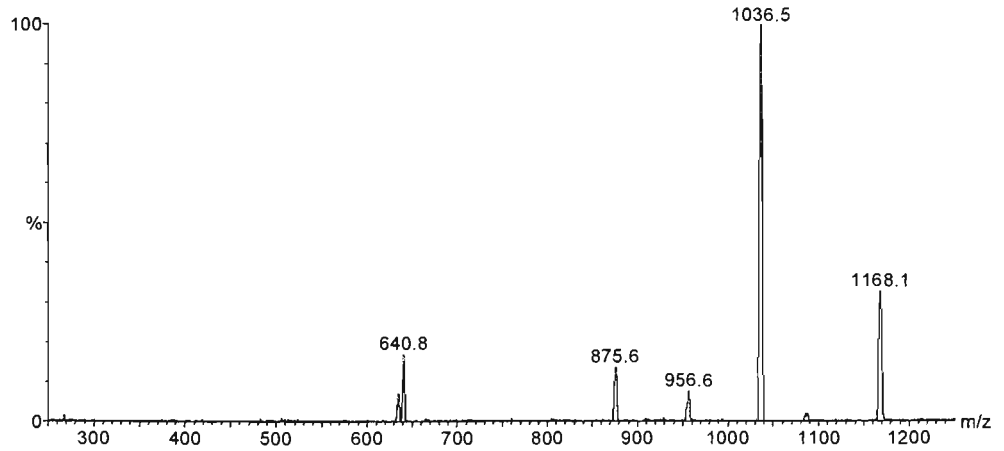


Scheme 3.5: Fragmentation of the $[M+Py-4H]^{4-}$ ion of 5′-d(GCAAAATTTTGC)-3′ (ON4) and PyA

The oligonucleotides 5′-d(CGAAAATTTTCG)-3′ (ON3) and 5′-d(GCAAAATTTTGC)-3′ (ON4) each incorporate a run of 4 adenines *i.e.* 5′-AAAA and in each case it was found that two major adducts were formed on reaction with pyrindamycin A (with similar results for ON3 and pyrindamycin B). The resulting adducts have the same mass and were not separated by HPLC. The evidence for the presence of two adducts is based on the presence of two pairs of sequence ions in the MS/MS spectra $[M+Py-4H]^{4-}$ ions of each oligonucleotide (table 3). The sequence ions $[w_7-H]^-$ and $[a_5-B_5-3H]^{2-}$ indicate that binding occurred at A5 and the $[w_6-H]^-$ and $[a_6-B_6-3H]^{2-}$ ions show that binding to A6 also occurs.

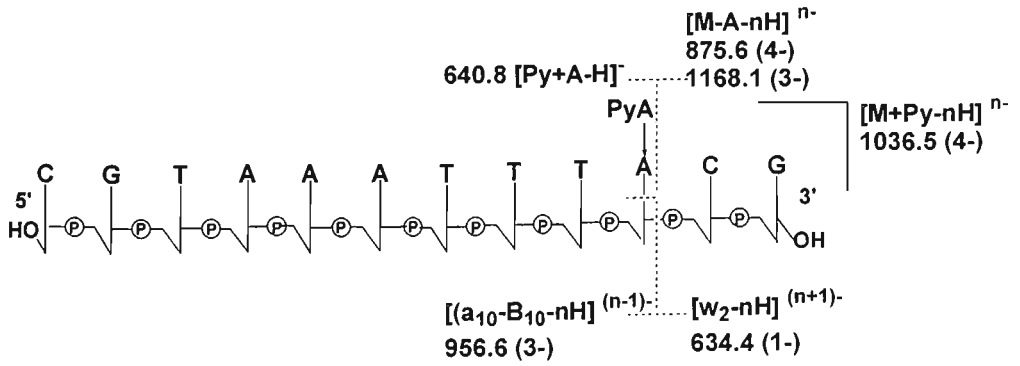
In the case of the MS/MS spectra of ions from adducts formed with ON3, the sequence ions from the A6 adduct were of much lower abundance. There are two possible explanations for this. First, that the A5 adduct is the major species produced from the reaction, thus resulting in more abundant sequence ions. Alternatively, the higher abundance of cleavage products from this adduct may indicate it has lower stability relative to the A6 adduct (*i.e.*

the former adduct fragments more readily). Given that other work (Boger *et al.*, 1990) suggests that when adenine is the base in the fourth 5′-position the non-covalent interactions are stronger than when guanine is in that position, the latter explanation is more likely.



5′-d(CGTAAATTTACG)-3′(ON5) and 5′-d(CGTTTATAAACG)-3′(ON7)

Figure 3.15: ESI-MS/MS spectrum of the $[M+Py-4H]^{4-}$ ion of the major adduct formed between pyrindamycin A and 5′-d(CGTAAATTTACG)-3′ (ON5)



Scheme 3.6: Fragmentation of the $[M+Py-4H]^{4-}$ ion of 5′-d(CGTAAATTTACG)-3′ (ON5) and PyA

As shown in figure 3.15, when the $[M+Py-4H]^{4-}$ ion (*m/z* 1036.5) was selected as the precursor ion, the resulting MS/MS spectrum is relatively simple, indicating only two reaction pathways. In the one pathway, the depurinated $[M-A-3H]^{3-}$ ion is evident at m/z 1168.1 together with the complementary $[Py+A-H]^-$ ion at *m/z* 640.8. In the second pathway, the precursor ion fragments to yield the $[M-A-4H]^{4-}$ species at *m/z* 875.6 and a neutral [Py+A] adduct. The depurinated ion then further fragments to give the singly charged w_2-ion ($[(d(CG)+H_2O)-H]^-$) at *m/z* 634.4 and the complementary 5′-sequence (a_{10}-B_{10})-ion ($[M-A-(d(CG)+H_2O)-3H]^{3-}$) at *m/z* 956.6. These ions enabled A10 to be identified as the site of binding, rather than the expected binding site of A6 (more below).

The oligonucleotide, 5′-d(CGTAAATTTACG)-3′ (ON5), was reacted with pyrindamycin A to examine the effect that thymine in the fourth 5′-position from the alkylation site would have on the sequence selectivity of the ligand. Thus, binding was expected to occur to the high affinity site, 5′-(AA<u>A</u>)-3′ at A6. The MS/MS spectra of this adduct showed the depurinated species and the w_2 (*m/z* 634.4) and (a_{10}-B_{10})-sequence ions (*m/z* 956.6). This was unexpected and indicated that the site of alkylation was A10, a lower affinity 5′-(TT<u>A</u>)-3′ site. On the basis of other sequencing studies, (Boger *et al.*, 1990) the pyrindamycins are only expected to bind this site in the absence of a 5′-(AA<u>A</u>)-3′ binding site.

To explore the influence of the location of the binding site within the sequence, pyrindamycin A was also reacted with the oligonucleotide, 5′-d(CGTTTATAAACG)-3′

(ON7). This sequence was designed to incorporate the same two binding sites in reversed positions within the oligonucleotide.

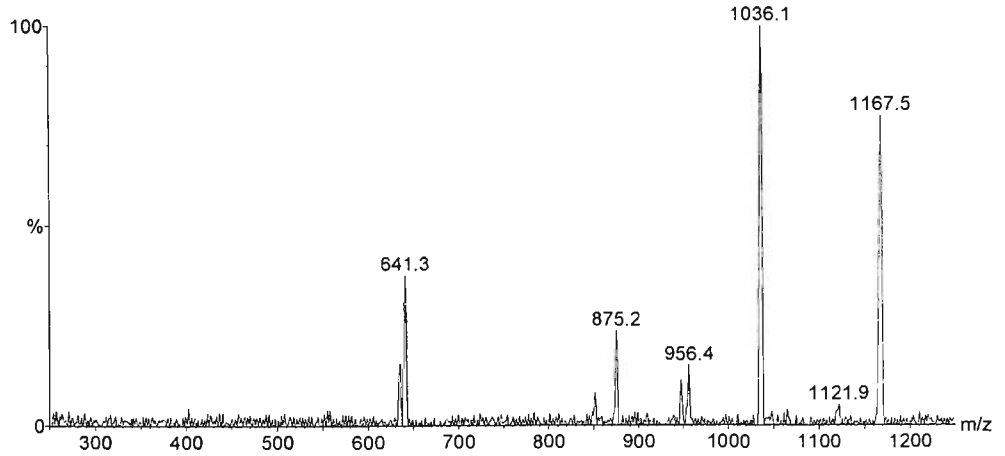
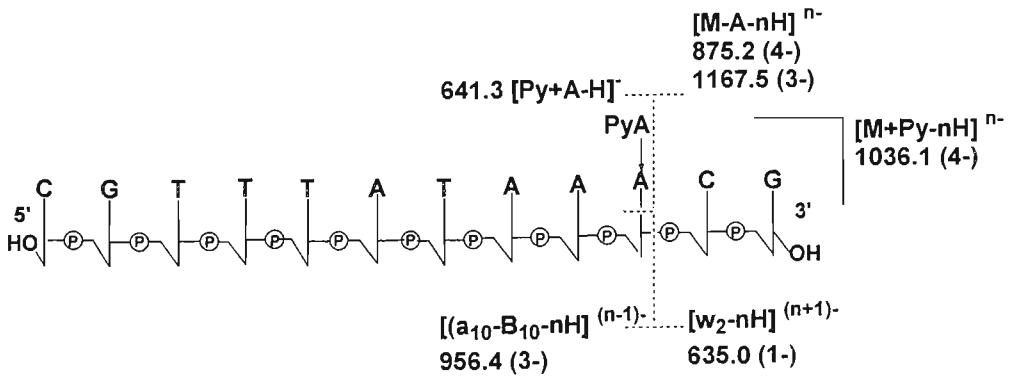


Figure 3.16: ESI-MS/MS spectrum of the $[M+Py-4H]^{4-}$ ion of the major adduct formed between pyrindamycin A and 5′-d(CGTTTATAAACG)-3′ (ON7)



Scheme 3.7: Fragmentation of the $[M+Py-4H]^{4-}$ ion of 5′-d(CGTTTATAAACG)-3′ (ON7) and PyA

Figure 3.16 shows the ES-MS/MS spectrum of the $[M+Py-4H]^{4-}$ ion formed upon reaction of pyrindamycin A and 5′-d(CGTTTATAAACG)-3′ (ON7). The $[M+Py-4H]^{4-}$ ion yielded

the depurinated form of ON7 as is evidenced by the peaks at *m/z* 875.2 and 1167.5. The sequence ions $[w_2]^-$ and $[a_{10}\text{-}B_{10}\text{-}4H]^{3-}$ at *m/z* 635.0 and 956.4 were also present.

The alkylation site was again determined to be A10, owing to the presence of the w_2 and $(a_{10}\text{-}B_{10})$-sequence ions form the depurinated oligonucleotide (table 3.5). Hence, in this case binding occurs to the higher affinity 5′-(AAA)-3′ at A10 rather than to the 5′-(TTA)-3′ at A6. This shows that the position of the alkylation site within these short oligonucleotides may influence relative sequence selectivities. This is presumably because sites closer to the terminii are more accessible owing to the "fraying" of the ends of the duplex. Alternatively, these data may reflect sequence-dependent conformational changes on the minor groove. These data suggest that the sequence selectivity observed for short oligonucleotides, such as those used in this study, may differ from that observed in longer segments of DNA and may indicate the need for longer sequences of DNA in any subsequent studies arising from this work.

5'-d(CTAAAATTTTAG)-3' (ON6)

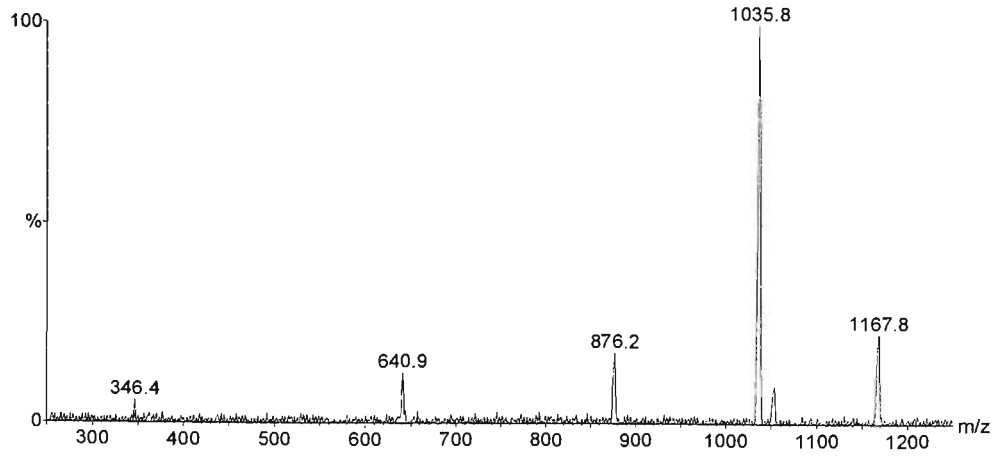
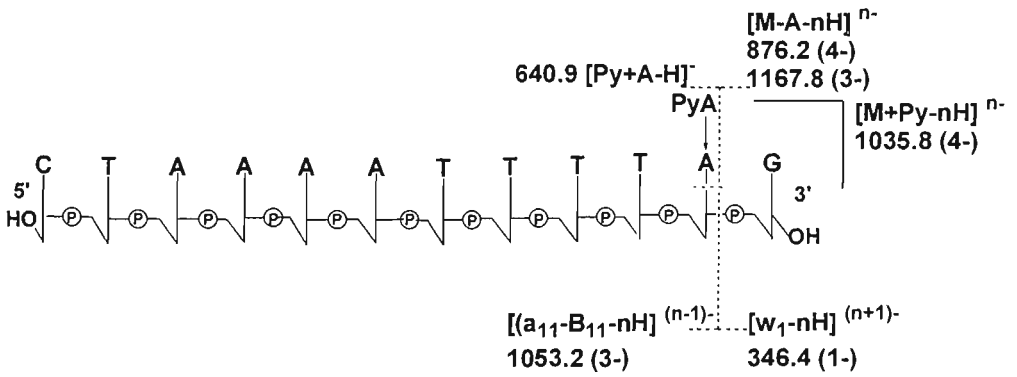


Figure 3.17: ESI-MS/MS spectrum of the $[M+Py-4H]^{4-}$ ion of the major adduct formed between pyrindamycin A and 5'-d(CTAAAATTTTAG)-3' (ON6)



Scheme 3.8: Fragmentation of the $[M+Py-4H]^{4-}$ ion of 5'-d(CTAAAATTTTAG)-3' (ON6) and PyA

Figure 3.17 shows the MS/MS spectrum of the $[M+Py-4H]^{4-}$ ion of 5'-d(CTAAAATTTTAG)-3' (ON6) and PyA. The MS/MS spectrum shows two clear fragmentation pathways. The first of which involves depurination yielding the $[M-A-4H]^{4-}$

ion at *m/z* 876.2. The second (and presumably subsequent) pathway then gives rise to the $[w_1]^-$ and $[a_{11}\text{-}B_{11}\text{-}4H]^{3-}$ sequence ions at *m/z* 346.4 and 1053.2 respectively. Thus the reaction of pyrindamycin A with the oligonucleotide, 5′-d(CTAAAATTTTAG)-3′ (ON6) resulted in alkylation occurring at A11, and hence binding occurred to the 5′-(TTA)-3′ site rather than to the higher affinity 5′-(AAA)-3′ site (Boger *et al.*, 1990) at A5 and A6. This was similar to the binding observed to the oligonucleotide, 5′-d(CGTAAATTTACG)-3′ (ON5), where alkylation also occurred at A10 and again reinforces the suggestion that the influence of the position of the binding site within these short oligonucleotides is of great importance.

5′-d(CGCAAAGCTTTG)-3′ (ON8)

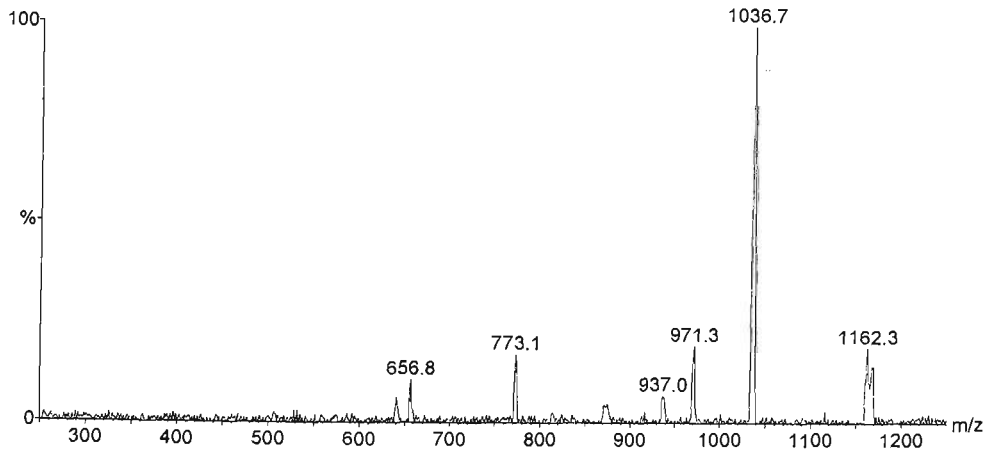
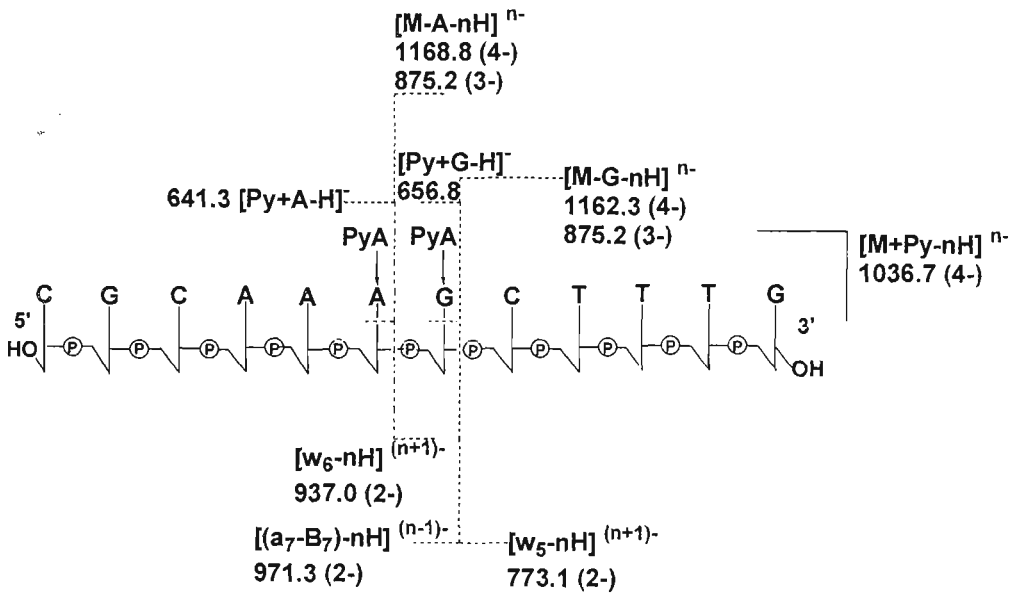


Figure 3.18: ESI-MS/MS spectrum of the $[M+Py\text{-}4H]^{4-}$ ion of the major adduct formed between pyrindamycin A and 5′-d(CGCAAAGCTTTG)-3′ (ON8)



Scheme 3.9: Fragmentation of the $[M+Py-4H]^{4-}$ ion of 5′-d(CGCAAAGCTTTG)-3′ (ON8) and PyA

Figure 3.18 shows the MS/MS spectrum of the $[M+Py-4H]^{4-}$ ion of 5′-d(CGCAAAGCTTTG)-3′ (ON8) and PyA. This experiment was designed to explore the binding of pyrindamycin A to single-stranded DNA as the sequence described here is non-self-complementary. The spectrum shows ions formed via four simple reaction pathways. In the one pathway, the depurinated $[M-A-3H]^{3-}$ ion is evident at m/z 1168.8 together with the complementary $[Py+A-H]^-$ ion at *m/z* 641.3. In the second pathway, the precursor ion fragments to yield the $[M-A-4H]^{4-}$ species at *m/z* 875.2 and a neutral [Py+A] adduct. The expected pathways for fragmentation for binding at A6 would yield the w_6 and (a_6-B_6)-ions, which were observed at *m/z* 937.0 and 814.5. However, the intact adduct ion, $[M+Py-4H]^{4-}$ ion also fragmented to yield the depurinated $[M-G-3H]^{3-}$ ion at *m/z* 1162.3 along with the complementary $[Py+G-H]^-$ ion at *m/z* 656.8. The $[M-G-4H]^{4-}$ species was observed at *m/z* 871.6 and further fragmented to give the w_5-ion at *m/z* 773.1 and corresponding (a_7-B_7)-ion

at *m/z* 971.3. In this case, the ions corresponding to alkylation at the guanine at position G7 in the oligonucleotide were of greater intensity than those owing to the A6 adduct.

This was the first observation of guanine alkylation for any of the sequences studied and is presumably because this particular oligonucleotide can not adopt a double-stranded conformation in solution as the sequence was non-self-complementary. The alkylation of guanine can be explained in terms of the nucleophilicity of the nucleobases, which follows the order: G > A > C > T, U (*i.e.* ligands show a greater affinity for guanine as it is more nucleophilic than the other bases). (Barton & Lippard, 1980) Nucleophilic sites within DNA at physiological pH follow the order: guanine-N7, adenine-N7, adenine-N1, cytosine-N3. (Martin & Mariam, 1979; Marzilli, 1977) The N3 of the purines is sterically hindered by the sugar moiety but may become available upon changes in the glycosidic torsion angles. (Kazakov, 1996)

ON8 was designed to be non-self-complementary, the binding that occurred to the adenine at A6 was unexpected. It was postulated that this binding resulted from the formation of a double-stranded conformation by the overlap of two strands of the oligonucleotide as follows:

5′-CGCAAAGCTTTG-3′
3′-GTTTCGAAACGC-3′

To test this hypothesis we designed a new oligonucleotide, which would not be able to adopt the same overlapping conformation. To achieve this guanine and a cytosine were added to the 3′-end of a new oligonucleotide, *i.e.* 5′-CGCAAAGCTTTGGC-3′. The

additional bases would not be able to form base-pairs at the termini causing fraying and destabilisation of the double-stranded conformation as is shown below.

$$\begin{array}{l} 5'\text{-}CG_{CAAAGCTTTG}{}^{GC\text{-}3'} \\ \rightarrow \qquad\qquad\qquad\qquad \leftarrow \\ 3'\text{-}CG^{GTTTCGAAAC}{}_{GC\text{-}5'} \end{array}$$

5'-d(CGCAAAGCTTTG)-3' (ON8) and 5'-d(GCGTTTCGAAAC)-3' (ON9)

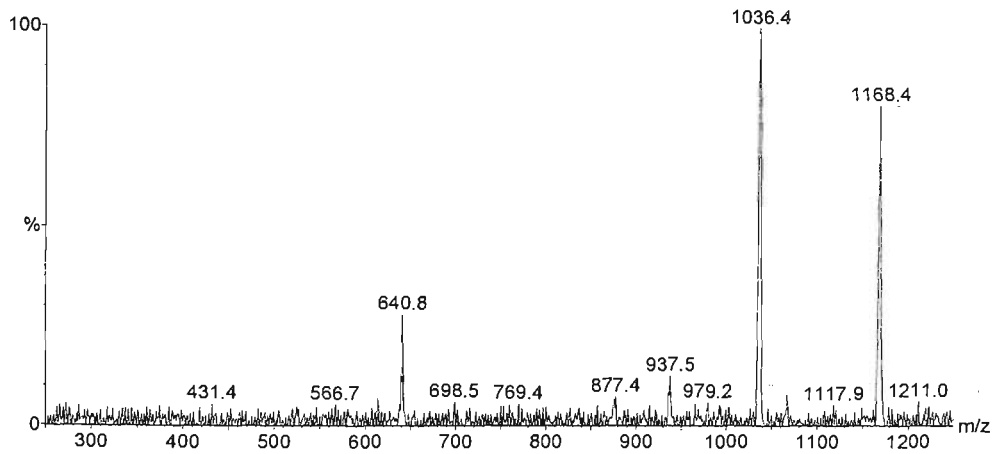
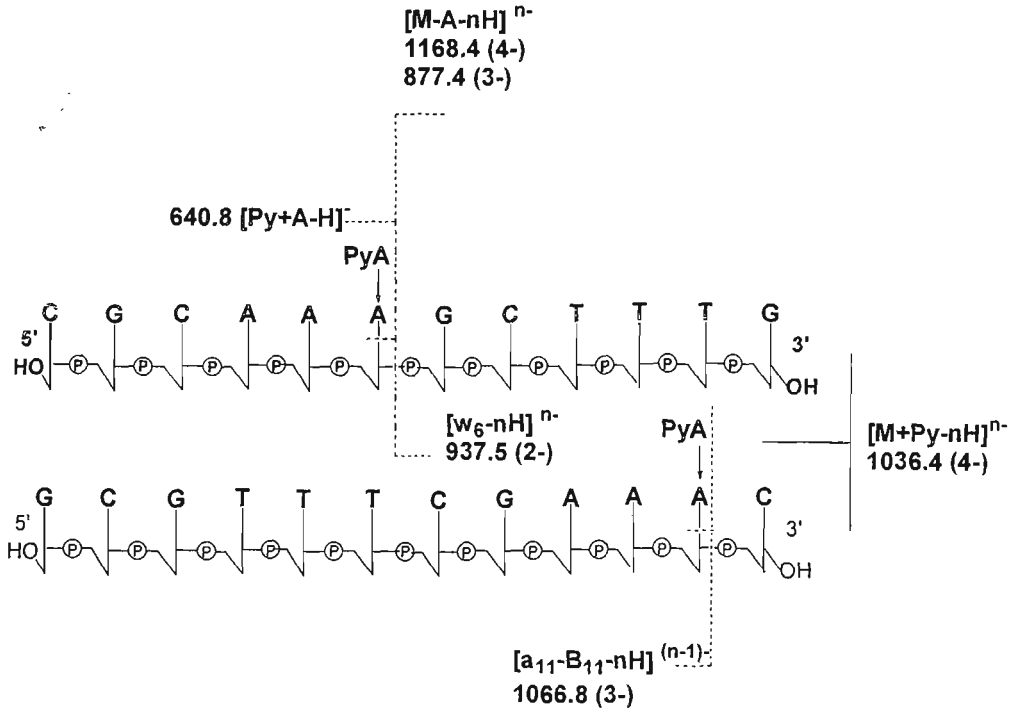


Figure 3.19: ESI-MS/MS spectrum of the $[M+Py-4H]^{4-}$ ion of the major adduct formed between pyrindamycin A and 5'-d(CGCAAAGCTTTG)-3'/5'-d(GCGTTTCGAAAC)-3' (ON8/ON9)



Scheme 3.10: Fragmentation of the $[M+Py-4H]^{4-}$ ion of 5′-d(CGCAAAGCTTTG)-3′/5′-d(GCGTTTCGAAAC)-3′ (ON8/ON9) and PyA

Figure 3.19 shows the ESI tandem mass spectrum of the $[M+Py-4H]^{4-}$ ion of either 5′-d(CGCAAAGCTTTG)-3′ (ON8) and PyA or 5′-d(GCGTTTCGAAAC)-3′ (ON9) and PyA. It is impossible to distinguish between the two as the oligonucleotide-pyrindamycin adducts have the same nominal mass. In the MS/MS spectrum, however, there were three fragmentation pathways evident, namely decomposition to yield the $[M-A-3H]^{3-}$ ion at *m/z* 1168.4 and $[Py+A]^-$ ion at *m/z* 640.8 (which is not sequence specific) and loss of neutral (Py+A) to yield $[M-A-4H]^{4-}$ at *m/z* 877.4. The latter ion decomposes to either the complementary $[w_6-H]^{2-}$ (*m/z* 937.5) and $[a_6-B_6-3H]^{2-}$ (not observed) ions of ON8 or the $[w_1]^-$ (not observed) and $[a_{11}-B_{11}-4H]^{3-}$ (*m/z* 1066.8) ions of ON9. These show that the ligand is binding to both A6 of ON8 and A11 of ON9 respectively. No evidence of guanine

alkyation was observed. This confirms that the ligands alkylate adenine rather than guanine when a double-stranded conformation is present in solution.

5'-d(CGAAAGCTTTCG)-3' (ON10)

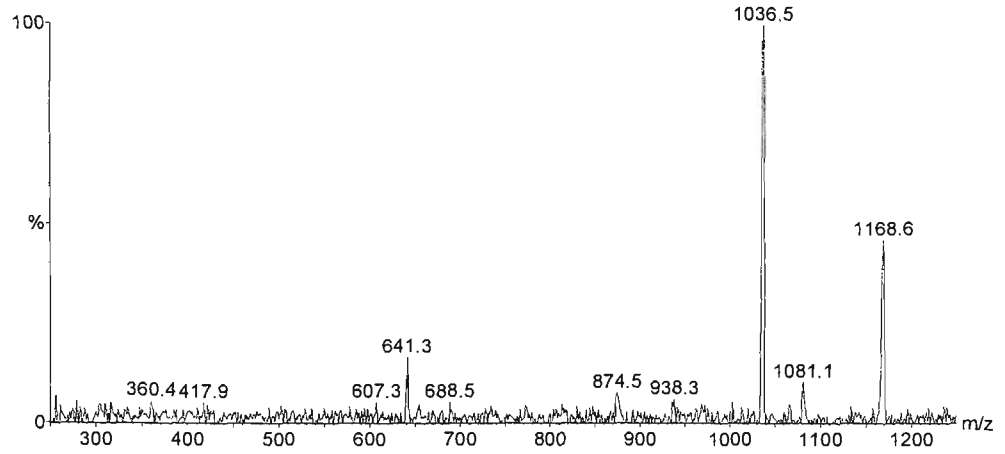
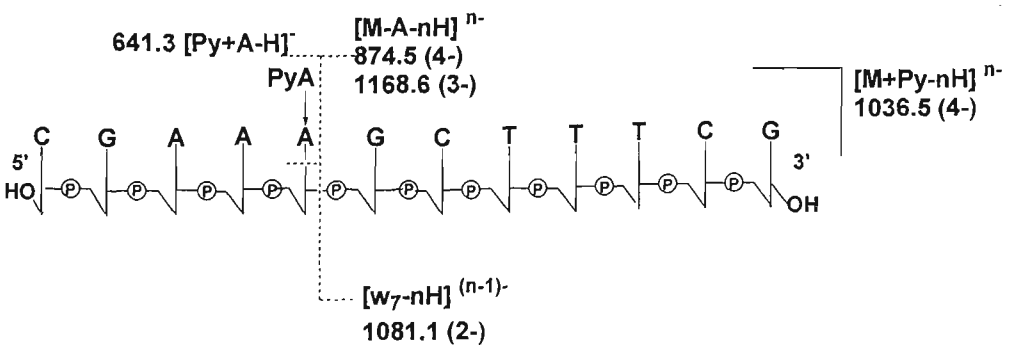


Figure 3.20: ESI-MS/MS spectrum of the $[M+Py-4H]^{4-}$ ion of the major adduct formed between pyrindamycin A and 5'-d(CGAAAGCTTTCG)-3' (ON10)



Scheme 3.11: Fragmentation of the $[M+Py-4H]^{4-}$ ion of 5'-d(CGAAAGCTTTCG)-3' (ON10) and PyA

ON10 (5′-CGAAAGCTTTCG-3′) was designed to incorporate an identical binding site to ON8, *i.e.* 5′-AAAG-3′. Binding to ON8 occurred at both the underlined adenine and guanine bases. It was thought that binding to guanine only occurred as a result of the absence of the double-stranded complex required for the minor groove interaction. As such, the self-complementary oligonucleotide, ON10, was used to examine whether guanine alkylation would be observed in the presence of double-stranded conformation of the oligonucleotide. The ESI mass spectrum (not shown) of the reaction mixture yielded a number of ions similar to that shown for ON8 (figure 3.15). The subsequent MS/MS spectrum of the $[M+Py-4H]^{4-}$ species at *m/z* 1036.5, as shown in figure 3.20 was relatively simple. As was the case for the other oligonucleotides studied, depurination occurred, yielding only the pyrindamycin-adenine adduct and both the $[M-A-3H]^{3-}$ and $[M-A-4H]^{4-}$ ions at *m/z* 1168.6 and *m/z* 874.5 respectively. Only the $[w_7-H]^{2-}$ ion at *m/z* 1081.1 and of low intensity (and not the corresponding $[a_5-B_5-3H]^{2-}$ ion) was observed suggesting that the adduct was more stable than previous adducts observed with different sequences. There was no evidence of guanine alkylation for this self-complementary oligonucleotide.

5'-d(CGCAAAGCTTTGGC)-3' (ON11)

An additional non-self-complementary oligonucleotide, 5'-CGCAAAGCTTTGGC-3' (ON11), was also reacted with pyrindamycin A in order to clearly establish that guanine alkylation occurred to single-stranded DNA and hence that the double-stranded conformation was a prerequisite for adenine alkylation by these ligands. Figure 3.21 shows the ESI-MS/MS spectrum of the $[M+Py-5H]^{5-}$ ion at *m/z* 952.6. The first fragmentation pathway yielded the $[M-G-4H]^{4-}$ ion at *m/z* 1027.1 and the corresponding $[Py+G-H]^{-}$ ion at *m/z* 657.5. The latter ion was observed to undergo neutral guanine loss to give the $[Py-H]^{-}$ ion at *m/z* 506.4. The depurinated $[M-G-4H]^{4-}$ species fragmented by a second pathway to produce the $[w_7-H]^{2-}$ and $[a_7-B_7-3H]^{2-}$ sequence ions at *m/z* 1082.4 and *m/z* 970.2 respectively. These ions show that alkylation by pyrindamycin A occurred at the G7 as was postulated. No evidence of adenine alkylation was observed for this single-stranded non-self-complementary oligonucleotide.

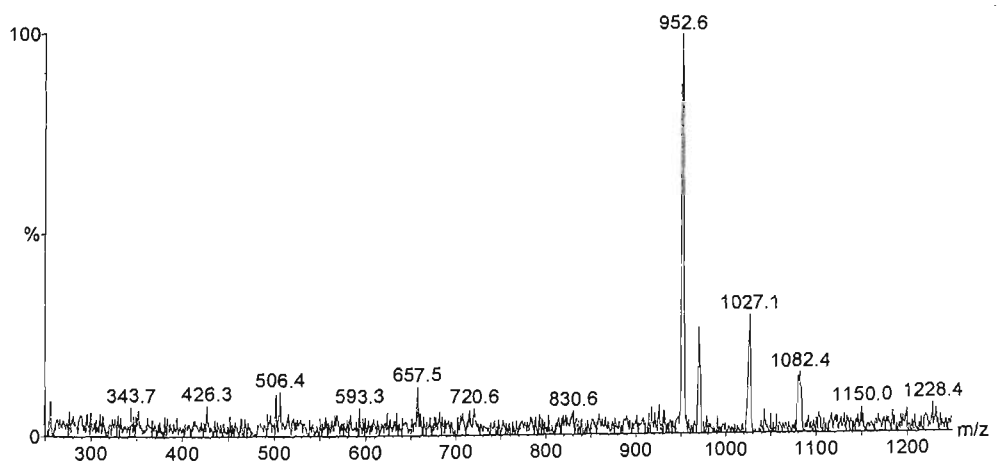
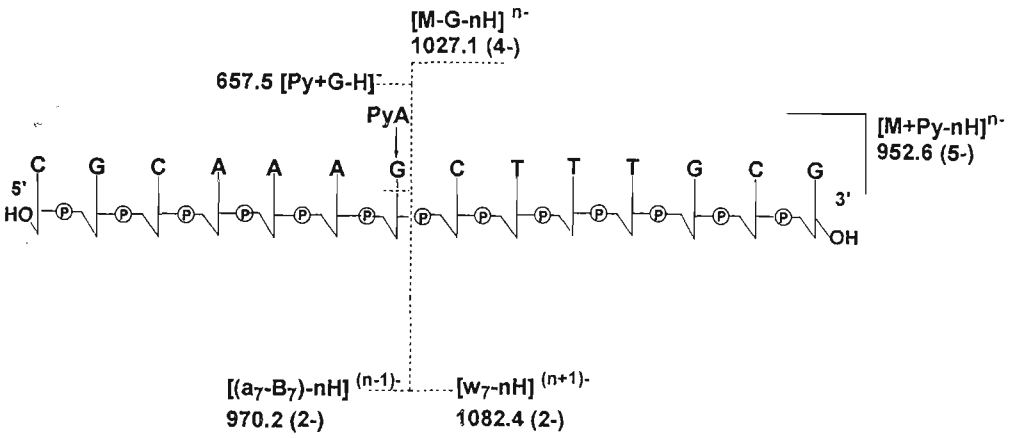


Figure 3.21: ESI-MS/MS spectrum of the $[M+Py-4H]^{4-}$ ion of the major adduct formed between pyrindamycin A and 5'-d(CGCAAAGCTTTG)-3' (ON8)



Scheme 3.12: Fragmentation of the $[M+Py-4H]^{4-}$ ion of 5′-d(CGCAAAGCTTTG)-3′ (ON8) and PyA

3.7 Conclusions

The ESI mass spectra of each of the reaction mixtures clearly establish that the pyrindamycins bind to each of the oligonucleotides examined. The ESI mass spectra also provide some indication of the stability of the adduct(s) formed, and therefore indirectly a measure of the strength of the non-covalent interactions occurring between the ligand and the oligonucleotide. The adducts formed between pyrindamycin A and ON1-ON4 and ON7 involved alkylation at a 5′-AA<u>A</u> site within the oligonucleotides, whilst ON5 and ON6 involved alkylation at a 5′-TT<u>A</u> site. ON5 and ON6 were less stable than ON1-ON4 as the relative intensity of sequence ions produced were greater for both sequences. These data can be explained in terms of the greater affinity of the ligand for the 5′-AA<u>A</u> site over the 5′-TTA site. ON7, however, showed a similar trend to ON5 and ON6. Hence, the lower stability is postulated to be a consequence of the position of the alkylation site (terminal versus internal) within these short oligonucleotides. The results obtained for adducts of each of the pyrindamycins A or B were comparable, thus providing further support for the proposal that these ligands interact with DNA via a mechanism involving conversion to a common intermediate as shown in scheme 1.2.

As a consequence of the gentle nature of the ESI ionisation process, sequence ions were only observed in ESI mass spectra of some reaction mixtures. This necessitated the use of ESI-MS/MS to identify the site(s) of alkylation of the ligands on the full range of sequences examined.

MS/MS spectra were obtained for $[M+Py-4H]^{4-}$ and $[M-A-4H]^{4-}$ ions of all the adducts formed. Each spectrum showed relatively few product ions indicating that only two or three

fragmentation pathways dominated decomposition of the adduct. This contrasts with the MS/MS spectra of free oligonucleotides (Iannitti, 1999) in which cleavages across the entire oligonucleotide backbone are generally observed. In the case of ligands that alkylate guanine (Iannitti *et al.*, 1997; Iannitti-Tito *et al.*, 2000) it was proposed that the dominance of one or two major fragmentation pathways is a consequence of alkylation labilising the glycosidic bond in the alkylated purine residue. The present study has shown that alkylation of adenines has the same influence on the fragmentation in MS/MS spectra. In all cases, the same dominant pathway is evident in the MS/MS spectra of the alkylated oligonucleotides, namely depurination followed by cleavage of the 3′-CO bond in the alkylated residue to yield complementary 3′ and 5′ sequence ions. The MS/MS spectra were also much simpler than the corresponding ESI mass spectra since the charge of the precursor is defined. In contrast, in the ESI mass spectra, a number of ions with different charge were observed for each the species present in the reaction mixtures. Ions resulting from cleavage to the 3′-side of the alkylation site (w-ions) were found to be generally more intense than the corresponding cleavage to the 5′-side (a-B_n ions). The same trend is observed in the MS/MS spectra of free oligonucleotides for which w-ions are generally more abundant. (Iannitti, 1999)

The work described further demonstrates the potential of electrospray tandem mass spectrometry for the characterisation of chemically modified oligonucleotides. This work has extended earlier studies to demonstrate that alkylation of adenine occurs to double stranded oligonucleotide sequences at the high affinity sites, however, binding was observed to lower affinity terminal sites, suggesting that longer oligonucleotides need to be

examined in order to mimic conditions in intact DNA. ESI-MS and ESI-MS/MS provides considerable detail about the binding and selectivity of these ligand-DNA interactions. The binding of the ligands occurred for all oligonucleotides and subsequent studies by tandem mass spectrometry allowed the elucidation of the binding site. Adenine (presumably N3) is the primary target for the alkylation event, however, guanine alkylation by the pyrindamycins was observed when the single-stranded, non-self-complementary oligonucleotide used. Hence, the requirement that DNA is in a duplex form for adenine alkylation to occur has been further confirmed. It should be noted that no attempt was made to determine the absolute sensitivity of these techniques, which could have been improved by the use of nanoelectrospray or capillary LC-MS on the Q-TOF, an approach that should be pursued in any subsequent studies resulting from this work.

Bibliography

Barry, J.P.; Vouros, P.; van Schepdael, A. and Law, S-J. (1995) *J. Mass Spec.* **30**, 993-1006

Barton, J.K. and Lippard, S.J. (1980) in *Metal Ions in Biology* (Spiro, T.G. Ed.) Vol. 1, Wiley, New York, pp 31

Boger, D.L.; Ishizaki, T.; Zarrinmayeh, H.; Munk, S.A.; Kitos, P.A. and Suntornwat, O. (1990) *J. Am. Chem. Soc.* **112**, 8961-8971

Boschenok, J. and Sheil, M.M. (1996) *Rapid Commun. Mass Spectrom.* **10**, 144-149

Caruthers, M.H.; Barone, A.D.; Beaucage, S.L.; Dodds, D.R.; Fisher, E.F.; McBride, L.J.; Matteuchi, M.; Stabinsky, Z. and Tang, J.Y. (1987) *Methods Enzymol.* **154**, 287-313

Iannitti, P. (1999) *Electrospray tandem mass spectrometry of oligonucleotides and ligand-oligonucleotide adducts.* PhD thesis, University of Wollongong, Wollongong

Iannitti-Tito, P.; Weimann, A.; Wickham, G. and Sheil, M.M. (2000) *The Analyst* **125**, 627-634

Kazakov, S.A. (1996) in *Bioorganic Chemistry: Nucleic Acids* (Hecht, S.M. Ed.) Oxford University Press, New York

Martin, R.B. and Mariam, Y.H. (1979) *Metal Ions in Biological Systems* Marcel Dekker, New York

Marzilli, L.G. (1977) *Prog, Inorg. Chem.* **23**, 255-378

McLuckey, S.A. and Habibi-Goudarzi, S. (1993) *J. Am. Chem. Soc.* **115**, 12085-12095

Metzler, D.E. (1977) *Biochemistry: The Chemical Reactions of Living Cells* Academic Press, New York

Sambrook, J.; Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (2nd Edition) Cold Spring Harbor Laboratory Press, New York

Sinha, N.D.; Biernat, J.; McManus, J. and Koster, H. (1984) *Nucleic Acids Research* **12**, 4539-4557

CHAPTER FOUR: *The Novel DNA Structures and NMR*

4.0 Introduction

DNA exhibits a wide range of structural topologies (duplex, triplex, quadruplex, bulged structures and hairpin loops), which have been associated with specific biological function(s) and are formed in a highly sequence-dependent manner. These structures may be recognised or stabilised by small molecule ligands, which in turn has significant bearing on the efficacy of these compounds as drug molecules. For example, some intercalators have been shown to bind more tightly to bulged structures than normal B-DNA with the possibility of perturbing the recognition of DNA defects by repair enzymes, or inducing a stabilising frame shift. Similarly, DNA lesions and strand breaks are also susceptible to the influence of ligand binding. Many DNA intercalators inhibit the activity of topoisomerase I and II, though it is not known whether binding to the DNA cleavage site is involved directly in inhibiting the strand religation step in the replication cycle. The next two chapters detail structural studies of ligand binding to a range of unusual DNA structures that have been engineered to include the unusual features of hairpin loops, strand breaks in the form of a nicked backbone and single base bulges to examine the effect that ligand binding may have on similar features in *in vitro* DNA.

4.1 The Nicked Site Analogues: Hairpins and Loops

In the current study, the formation of a double hairpin structure from the oligonucleotide, 5′-d(GCGAAGCACGAAGT)-3′ has been examined. As in the case of hairpin structures described previously, (Hirao *et al.* 1992) there is a single-stranded loop region closed by a base paired stem. However, these oligonucleotides were designed to incorporate two of

each, such that they would form a double hairpin and as such were named 'loop sequences'. The loop region consisted of a 5'-GAA, which has been previously found to be of greatest stability. (Hirao *et al.,* 1992) The stem region contained predominantly G-C base pairs, although an A-T base pair was also included as there have been no reports to date on hairpins containing A-T base pairs in the stem region. It was envisaged that the sequence 5'-d(GCGAAGCACGAAGT)-3' would fold to give two hairpins that would be mutually stabilised by co-axial stacking, but with a single break at the centre of the double stranded stem region. Thus, the double hairpin mimics a DNA damage site (single strand break), providing a model system for drug recognition, with potential implications for the DNA repair process.

The binding of a well-known and characterised intercalating agent, nogalamycin (figure 1.7) to this oligonucleotide, was explored for a number of reasons. First, the question of whether nogalamycin further stabilises the hairpin structures was addressed as nogalamycin and closely-related compounds have been shown to stabilise duplex DNA. (Searle *et al.*, 1988; Searle & Wakelin, 1990) Secondly, it was hoped to further probe the molecular basis for binding this intercalator to the double loop structure. It is well established that the binding kinetics of nogalamycin are very slow, (Fox & Waring, 1984) which is thought to be a consequence of the requirement for the DNA duplex to be disrupted to enable the bulky sugar substituents to thread through the structure. By providing a broken or 'nicked' backbone, it was hypothesised that intercalation of the agent would be facilitated, thereby enabling an investigation of the importance of the phosphate backbone in nogalamycin binding to DNA.

4.2 Sample Preparation

The nogalamycin were generously donated by Upjohn Corporation and was used without further purification. A stock solution of 5 mM nogalamycin was prepared in D_2O. The concentration was achieved by acidifying the suspension using 0.1 M DCl until the ligand dissolved. The pH was then adjusted to ca. 7 using 0.1 M NaOD.

The oligonucleotides used in this study were purified by reverse-phase HPLC using a similar method to that described in chapter 3. The column used was an YMC packed C18 semi-preparative column. A Hewlett Packard (Agilent) Series 1100 HPLC equipped with HP Chemstation™ software and a variable wavelength detector set to 260 nm was utilised. After the initial stage of HPLC, the fractions were pooled and excess acetonitrile was removed with the aid of a rotary evaporator. The volume of solvent was then reduced to ca. 10-15 mL by freeze-drying. An equal volume of 50% acetic acid was added to remove the DMT-protecting group. Ether extraction (3 x 200 mL diethyl ether) was then undertaken to separate the oligonucleotide from the protecting group and acid. Any remaining ether was removed with the aid of the rotary evaporator. The oligonucleotide solution was then dialysed using Cellusep™ dialysis tubing with a molecular weight cut-off of 1000 Da. Five solutions were used during the dialysis. Solutions 1, 2, 4 and 5 were distilled water, whilst solution 3 was composed of a high salt buffer (100 mM NaCl / 10 mM NaH_2PO_4). Each solution was allowed to equilibrate for eight hours at 4 °C with constant stirring. The resulting solution after dialysis was then freeze-dried.

Oligonucleotide samples were dissolved in 99.9% D_2O that contained 10 mM NaH_2PO_4, 100 mM NaCl and trimethylsilylpropionate (TSP) was added to give a final concentration of ~1 mM. The samples were ultrasonicated and freeze-dried, redissolved in 99.9% D_2O to a final volume of 700 μL and then placed in an NMR tube under a stream of dry nitrogen gas.

4.3 Titration of Nogalamycin with 5'-d(GCGAAGCACGAAGT)-3'

A 1:1 complex between nogalamycin and 5'-d(GCGAAGCACGAAGT)-3' was formed by titrating 40 μL aliquots of a 5 mM solution of nogalamycin into a solution of ca. 5 mM 5'-d(GCGAAGCACGAAGT)-3'. NMR spectra up to and including the 1:1 complex were acquired immediately after the addition of each aliquot of the nogalamycin solution. The pH of the samples was adjusted to pH ~7 after completion of the titration. The samples were then lyophilised several times from solutions of D_2O. The sodium salt of trimethylsilylpropionate was used as an internal reference compound. Sodium azide was added to act as an anti-bacterial agent, whilst EDTA was utilised to complex any heavy metal ions present.

4.4 Data Acquisition

4.4.1 1D ^{1}H-NMR Experiments

The progress of the titrations was monitored by 1D ^{1}H-NMR experiments at 298 K (25 °C) on a Bruker DRX500 NMR spectrometer. A spectral width of 9.997 ppm (4988.5 Hz) over 32k (32768) data points was employed. The transmitter frequency was placed at the centre

of the spectrum at the position of the D_2O solvent resonance. The solvent resonance was suppressed using low power irradiation (60 Db) during the relaxation delay of 1.5 s. The titration was undertaken via the addition of the ligand (nogalamycin) to the oligonucleotide in D_2O. The ligand was added in small aliquots until a 1:1 complex was formed. Formation of the complex was judged by the changes to the 1D ^{1}H-NMR spectrum, in particular by the appearance and disappearance of the cytosine H5 resonances. The complex was allowed to equilibrate for 24 hours. A series of 1D-^{1}H-NMR spectra were also acquired on the resulting complex between 276 K (3 °C) and 328 K (55 °C) on the same instrument.

4.4.2 2D NMR Experiments

Two-dimensional spectra of 5′-d(GCGAAGCACGAAGT)-3′ and its resulting complex with nogalamycin were recorded at 500 MHz. These were acquired at 288 K (15 °C), with a spectral width of 9.977 ppm in both dimensions. Nuclear Overhauser Spectroscopy (NOESY) spectra were acquired over 1024 complex data points in the t2 dimension and ≥ 400 t1 increments. Generally, 64 scans were summed to give the final spectra. The D_2O resonance was suppressed using low power irradiation during the mixing time and the relaxation delay of ca. 1.5 s. NOESY spectra were acquired at mixing times of 75, 100, 120 and 300 ms. Double Quantum Filtered Correlation Spectroscopy (DQF-COSY) and Total Correlation Spectroscopy (TOCSY) experiments were acquired with 1024 data points with up to 64 scans per increment. A spin lock of 84 ms was employed for the TOCSY experiment.

All spectra were processed on a Silicon Graphics Indy workstation using XWINNMR™ software.

4.5 Nomenclature

The nucleotide residues within this oligonucleotide were numbered as follows: 5′-d($G_1C_2G_3A_4A_5G_6C_7A_8C_9G_{10}A_{11}A_{12}G_{13}T_{14}$)-3′ and 5′-d($A_8C_9G_{10}A_{11}A_{12}G_{13}T_{14}$)-3′. The hairpin was numbered starting from nucleotide 8 for ease of comparison with the loop sequence.

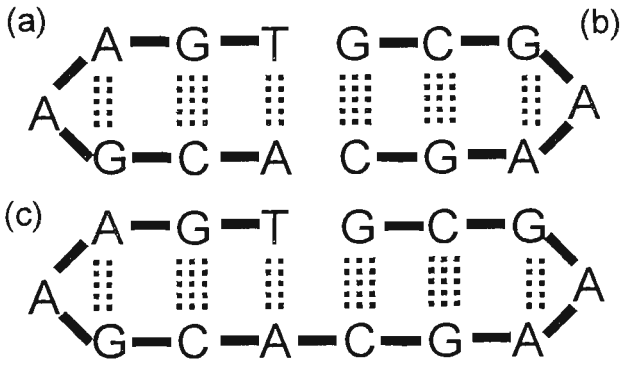


Figure 4.1: Structure of the hairpins and loop sequence studied: (a) 5′-d(ACGAAGT)-3′; (b) 5′-d(GCGAAGC)-3′ (Hirao et al., 1994) and (c) 5′-d(GCGAAGCACGAAGT)-3′

4.6 Evidence for the Formation of these Unique Structures

The formation of the 5'-GAA loop involves a non-Watson-Crick anti-anti base pair between G_3 and A_5 as is shown in figure 4.2. G-A mismatches have been investigated extensively because this is a common structural element in RNA folding. (Gao & Patel, 1988) Structural studies demonstrated that the sheared G(anti)-A(anti) conformation was adopted at neutral pH. (Kan *et al.*, 1983; Patel *et al.*, 1984; Prive *et al.*, 1987)

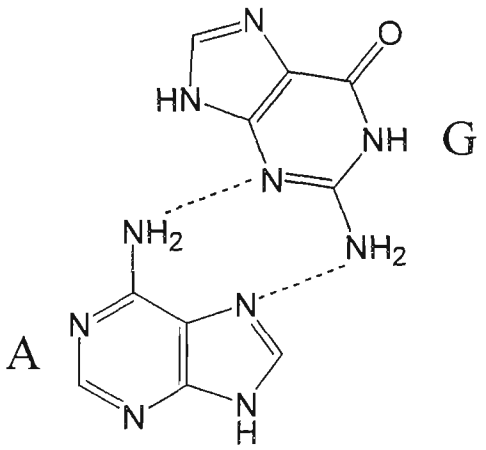


Figure 4.2: Anti-anti base pairing between G_{10} and A_{12} in the sequence d($G_{10}A_{11}A_{12}$). Hydrogen bonds are formed between the N3 and C2-amino of guanine and the N7 and C6-amino of adenine respectively. In Watson-Crick base pairs, the O6, N1 and C2-amino of guanine and the N1 and C6-amino of adenine are involved in hydrogen bonding interactions.

Three base pairs are involved in formation of the hairpins, although it appears that the G-A base pair is critical for the unusual stabilisation that is observed. The central A appears to be less critical and has been replaced with other bases in previous studies (Yoshizawa *et al.*, 1997) and relatively little effect on the stability was noted as a result.

In general, the chemical shifts of oligonucleotide protons fall within particular ranges depending on the type of proton. The H1' protons of the sugar occur between 5.0 and 6.5 ppm, whilst the H2' and H2" protons generally occur at around 2 ppm. The chemical shifts of the H8 and H6 protons of the bases occur between 7.0 and 8.5 ppm. Additional peaks are observed in the H1' region corresponding to the cytosine H5 protons. These are J coupled to the cytosine H6 protons and are good markers that can be used in the identification of the CH6 protons. The method for sequential assignment of the DNA backbone is illustrated in figure 4.3. This method allows the unambiguous assignment by stepping from base H8 or H6 proton to the H1' of its internucleotide sugar to the next base proton and so on. The same can be done for the H2' and H2" of the sugar. Figure 4.3 shows the sequential connectivity between the H8/H6-H1'.

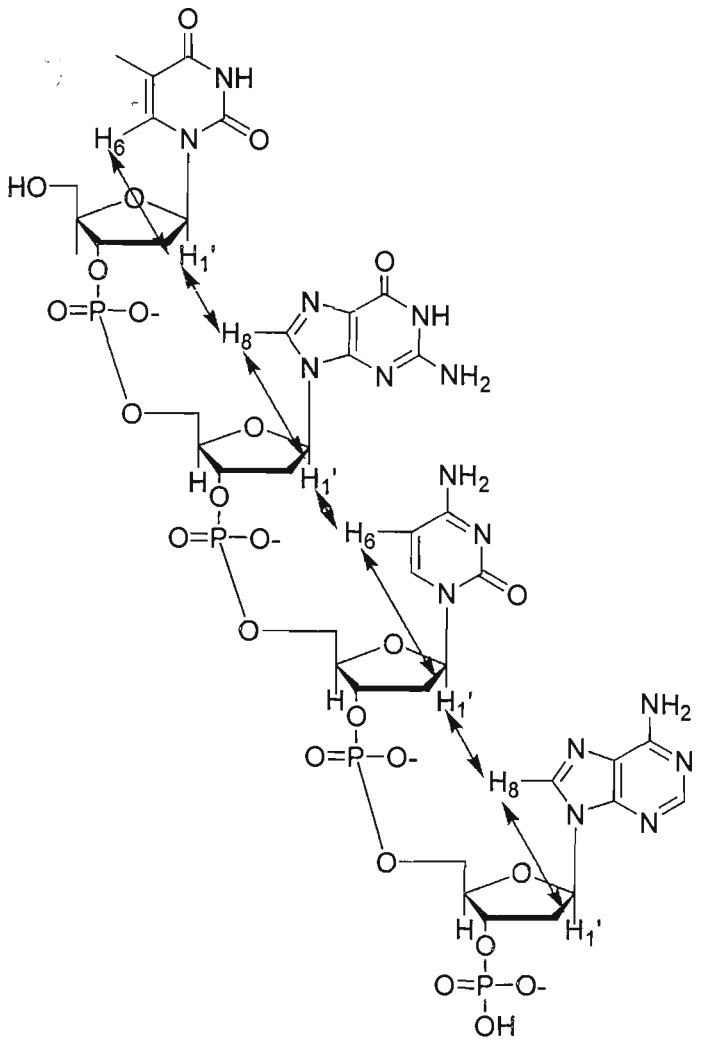


Figure 4.3: Sequential assignment of the DNA backbone. NOE crosspeaks are observed between intranucleotide sugar H1' protons and aromatic base H8/H6 protons and between adjacent internucleotide sugar H1' protons and aromatic base H8/H6 protons as shown.

Figure 4.4(a) shows the sequential assignment of $d(A_8C_9G_{10}A_{11}A_{12}G_{13}T_{14})$ between the H8/H6 of the bases and the H2'/H2" of the sugars. Connectivities were observed along the $A_8C_9G_{10}A_{11}$ and $A_{12}G_{13}T_{14}$ moieties of the oligonucleotide, but no NOEs were observed between the A_{11} and A_{12}. This connectivity can also be followed through the H8/H6 of the bases to the H1' of the sugars as is shown in figure 4.4(b).

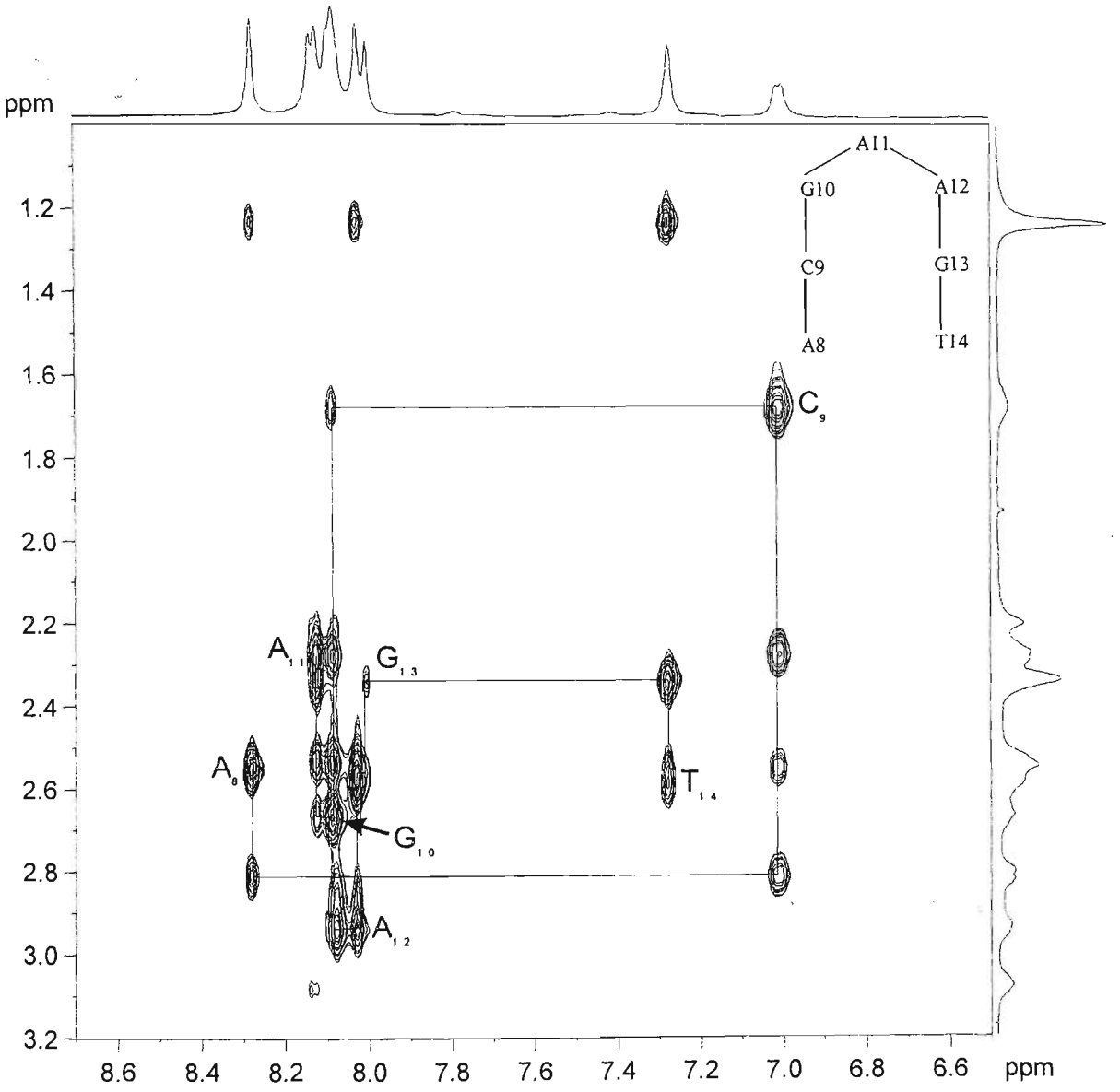


Figure 4.4(a): 300 ms NOESY spectrum showing H8/H6-H2'/H2" connectivities for 5'-d(ACGAAGT)-3' (3 mM) in D_2O containing 0.1 M NaCl/100 mM NaH_2PO_4 (pH 7.0). The solid lines represent the sequential connectivity observed for the A_8-C_9-G_{10}-A_{11} strand and through the A_{12}-G_{13}-T_{14} strand. There is a loss in connectivity owing to the unusual 5'-GAA loop.

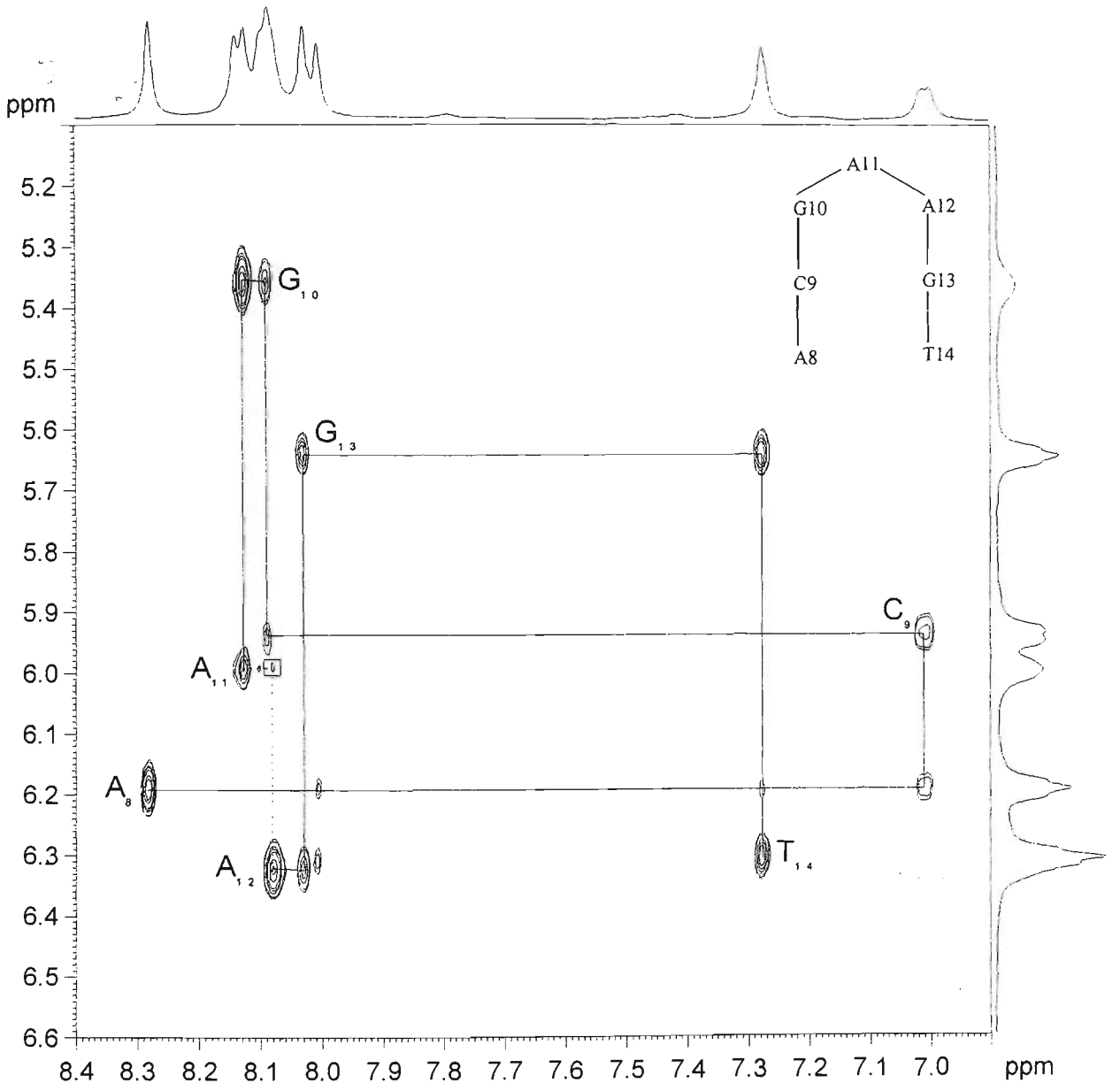


Figure 4.4(b): 300 ms NOESY spectrum showing H8/H6-H1' connectivities for 5'-d(ACGAAGT)-3' (3 mM) in D_2O containing 0.1 M NaCl/100 mM NaH_2PO_4 (pH 7.0). The solid lines represent the sequential connectivity observed for the DNA backbone. There is a break in connectivity owing to the unusual 5'-GAA loop and is represented by the dashed lines.

Owing to the relative simplicity of the NOESY spectra of the hairpin compared to the double hairpin or loop structure, the assignment of the hairpin facilitated the subsequent assignment of the loop sequence and is shown in figure 4.5. Figure 4.5(a) shows the connectivity observed along the G_6-C_7-A_8-C_9-G_{10} and A_{12}-G_{13}-T_{14}-G_1-C_2-G_3 strands. In NOESY data recorded at 15 °C, there were NOEs observed across the break in the DNA backbone, *i.e.* between G_{13} and T_{14}. This acts as evidence toward the formation of the loop sequence in which the two hairpins are involved in co-axial stacking. There were no NOEs observed along the backbone for the loop regions, owing to the unusual conformations adopted by the nucleotides in this region.

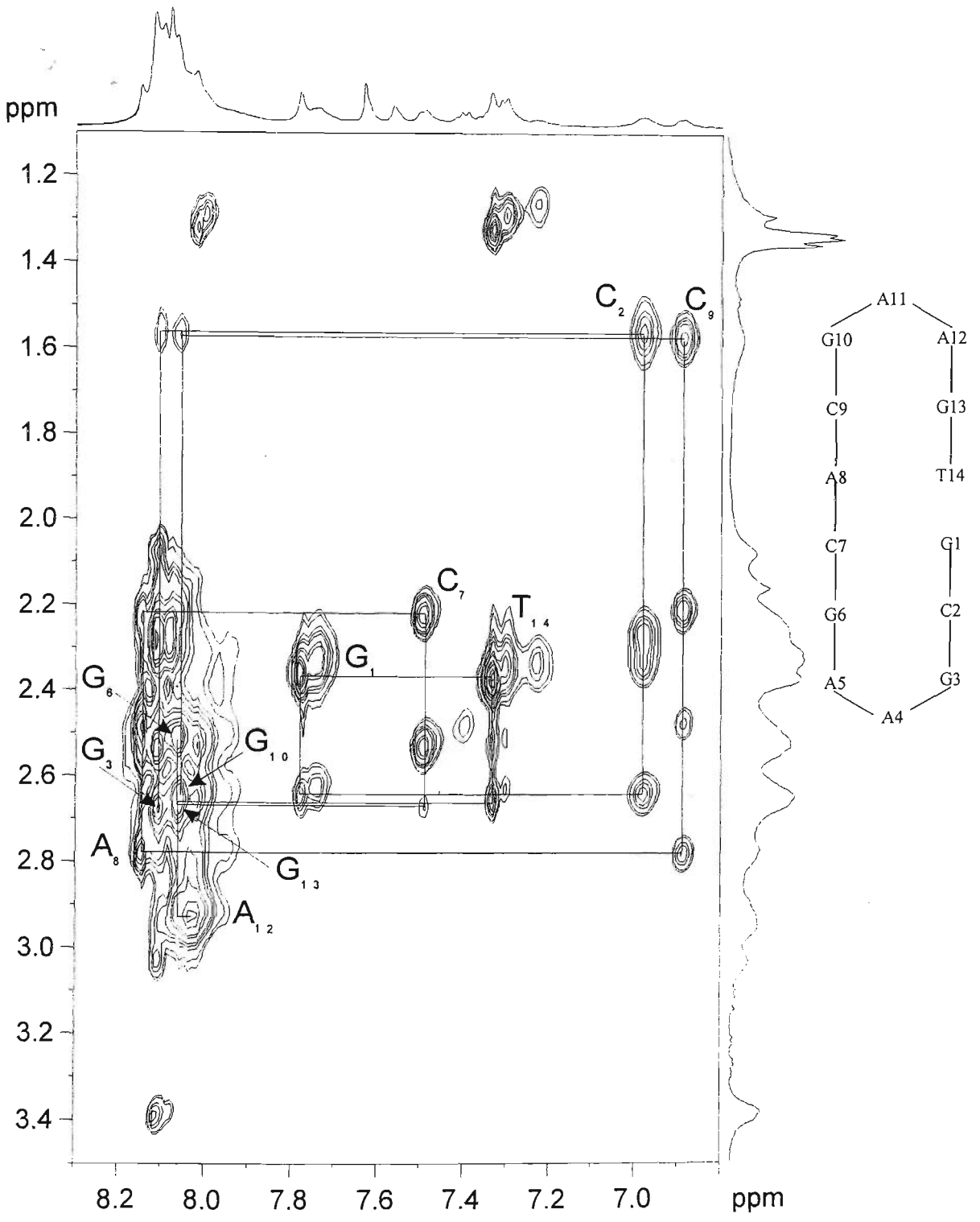


Figure 4.5(a): Section of the 300 ms NOESY spectrum showing H8/H6-H2'/H2" connectivity for 5'-d(GCGAAGCACGAAGT)-3' (3 mM) in D_2O containing 0.1 M NaCl/100 mM NaH_2PO_4 (pH 7.0). The intranucleotide H8/H6-H2'/H2" NOEs are labelled.

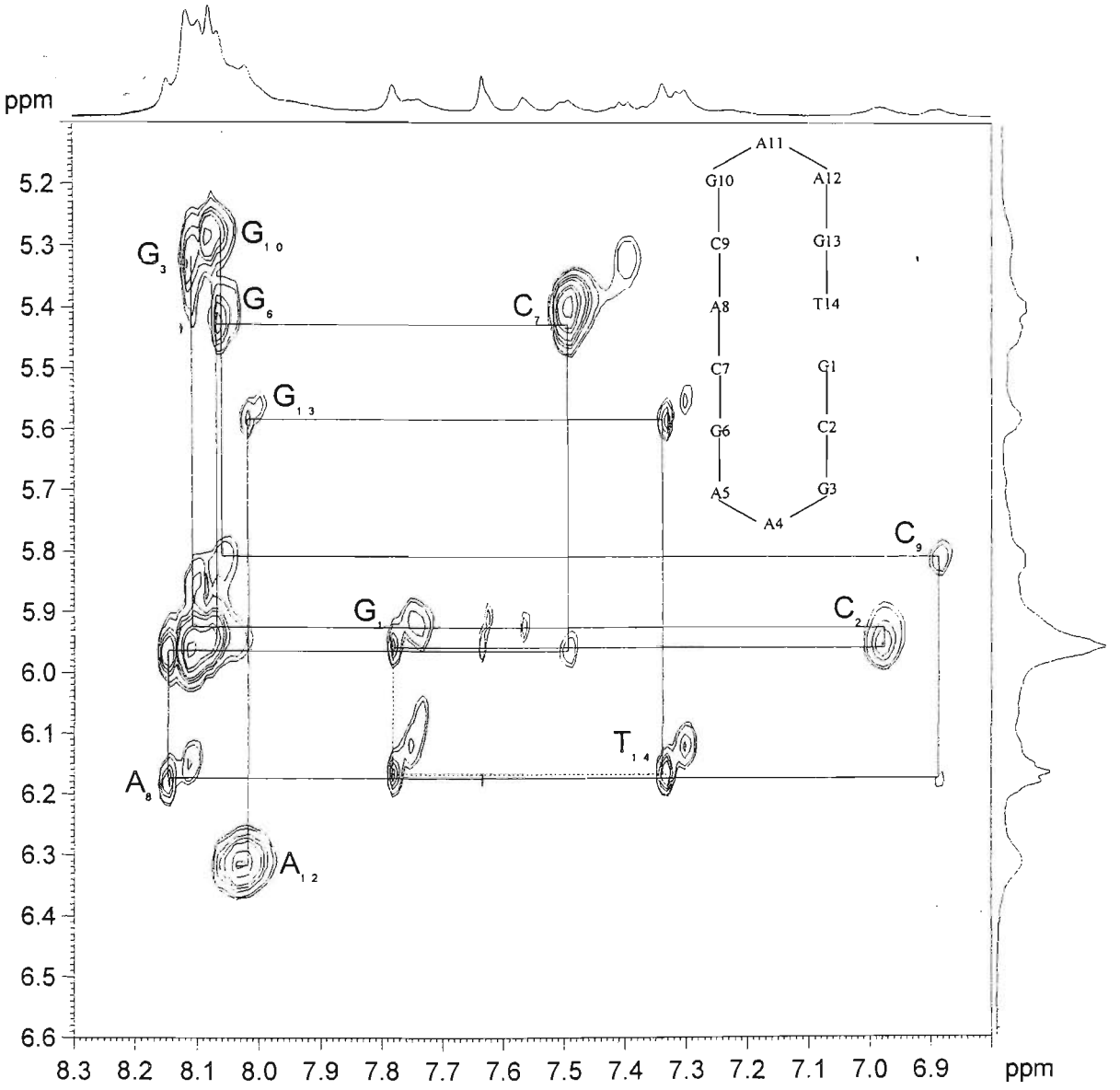


Figure 4.5(b): Section of the 300 ms NOESY spectrum showing the H8/H6-H1' connectivity for 5'-d(GCGAAGCACGAAGT)-3' (3 mM) in D_2O containing 0.1 M NaCl/100 mM NaH_2PO_4 (pH 7.0).

Figure 4.5(b) shows the H8/H6-H1' region of the NOESY spectrum for 5'-d(GCGAAGCACGAAGT)-3'. The solid lines show the connectivity along the G_1-C_2-G_3 strand and A_{12}-G_{13}-T_{14} strand. It is interesting to note that NOEs are observed across the break in the DNA backbone, *i.e.* between G_1H8 and T_{14}H1', as is indicated by the

continuation of the dashed lines between these resonances. The continuation of the solid lines show the connectivity along the G_6-C_7-A_8-C_9-G_{10} strand. There is a break in connectivity along the DNA backbone at each loop region. The observation of NOEs across the break site provides unambiguous evidence for a significant population of co-axially stacked hairpins.

The oligonucleotides used in this study, however, did not adopt typical double stranded conformations and as such many resonances occurred at significantly different chemical shifts to those observed for double stranded DNA. The initial NMR studies of the heptamer sequence 5′-d(GCGAAGC)-3′ by Hirao *et al.* provided some insights into features characteristic of hairpin formation. For example, the H4' of A_4 in the loop undergoes a large up-field shift as a consequence of stacking interactions in the loop region. (Hirao *et al.*, 1994) The chemical shift of the H4' was used as one of a number of indicators that the hairpins are folded. The H4', H5' and H5" resonances of the adenines thought to be in the loop were observed up-field shifted by $\geq$ 2 ppm compared to the other H4' signals from nucleotides in the stem regions. The H2' and H2" sugar protons of A_4 and A_{11} are observed to be down-field shifted by 0.13 and 0.06 ppm respectively. These shifts are analogous to those identified in the isolated hairpins (Hirao *et al.*, 1994; Yoshizawa *et al.*, 1997) and can be explained by the close proximity of the H4' of A_5 and A_{11} to the non-Watson-Crick base pairs of G_3-A_5 and G_{10}-A_{12} respectively. Further, the cytosine H2' and H2" resonances of C_2 and C_9 were observed up-field of the thymine methyl resonances at 1.56 and 1.57 ppm respectively. These data suggest that the loops are well formed. The similarity of the shifts

for the two 5′-GAA loop components of 5′-d(GCGAAGCACGAAGT)-3′ suggest that they are populated to similar extents.

Figure 4.6 shows the change in chemical shift occurring as a result of end-stacking, *i.e.* the difference in chemical shift between the protons of the hairpin and those of the loop sequence. Detailed analysis of the NMR data for the isolated hairpin 5′-d(ACGAAGT)-3′ and the same sequence within the loop sequence revealed differences in chemical shifts of up to 0.2 ppm for nucleotides involved in end-stacking. Similar results (chemical shift changes up to 0.2 ppm) were obtained when the chemical shift values reported for the 5′-d(GCGAAGC)-3′ hairpin studied by Hirao *et al.* were compared to the values obtained for the same sequence within 5′-d(GCGAAGCACGAAGT)-3′.

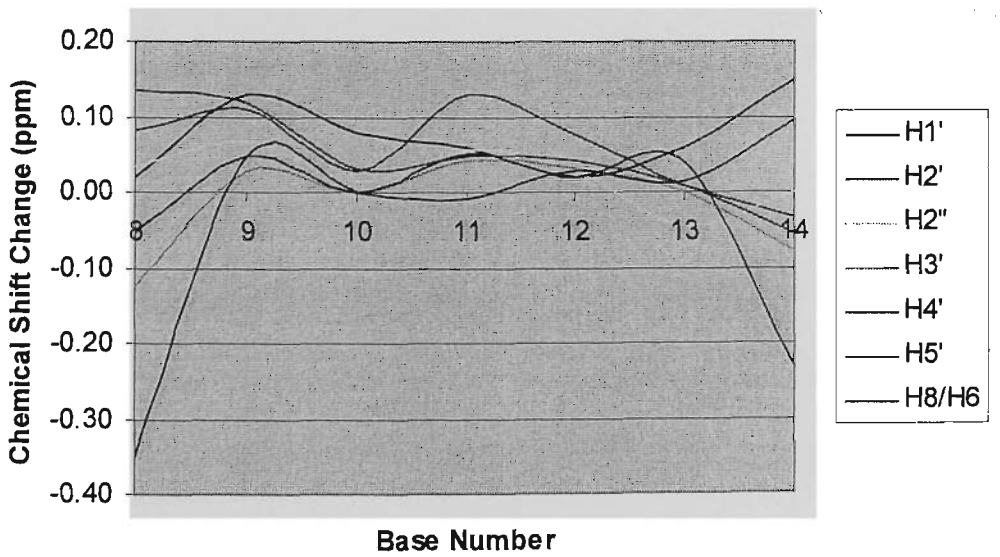


Figure 4.6: Change in chemical shift observed in the 1D-^{1}H-NMR spectra as a consequence of end-stacking. Comparison of the isolated hairpin 5′-d(ACGAAGT)-3′ and the same sequence within the double hairpin.

Further evidence for the sheared G_3(anti)-A_5(anti) conformation is provided by the fact that the G_3H1'-G_3H8 and A_5H1'-A_5H8 crosspeaks are both weak in comparison to the NOEs observed between the H5 and H6 of cytosine, consistent with anti-glycosidic torsion angles for G_3 and A_5 in the mismatch site.

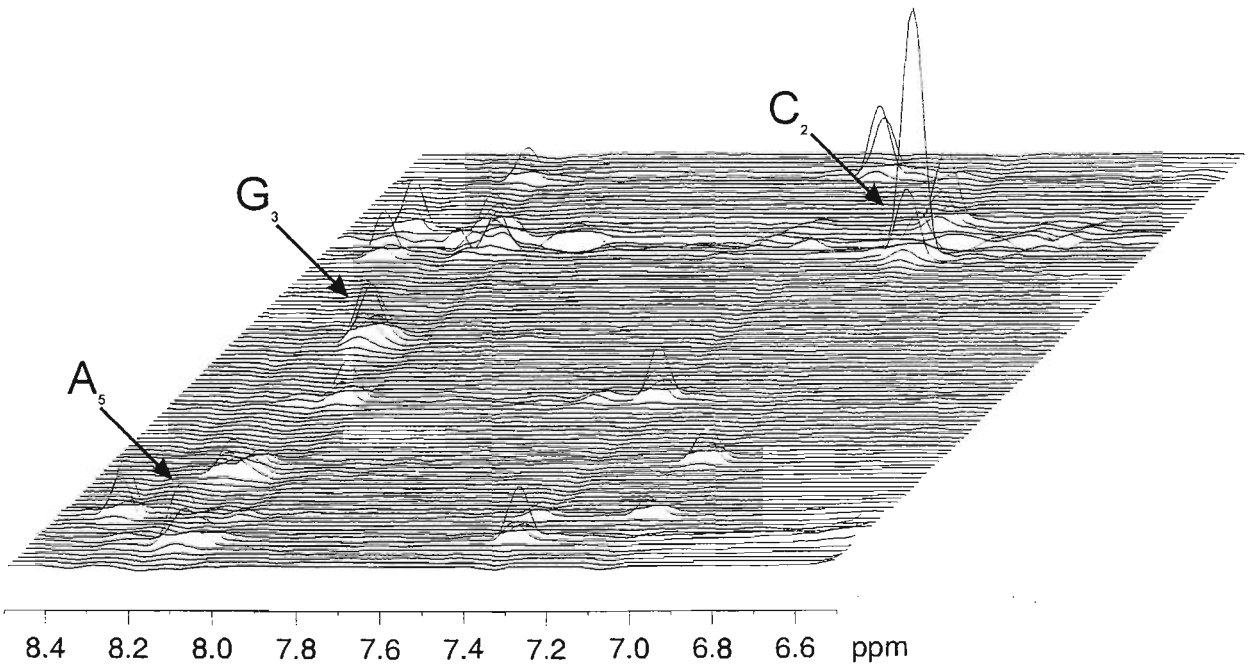


Figure 4.7: Expanded stacked plots of the 300 ms NOESY spectra of 5'-d(ACGAAGT)-3'. This region establishes distance connectivities between the base protons (6.5-8.5 ppm) and the sugar H1' and cytosine H5 protons (4.5-6.5 ppm). The cytosine H5-H6 crosspeak and the base to its own sugar H1' crosspeaks are labelled in the figure.

Figure 4.8 illustrates part of the deoxyribose spin systems for the H4' and H2'/H2" sugar protons of the adenines involved in the loop region (A_4 and A_{11}). The solid lines show the

pattern of NOEs observed for A_4, while the dashed lines represent the A_{11} sugar protons. The 1D NMR spectrum has also been shown on the x- and y-axes.

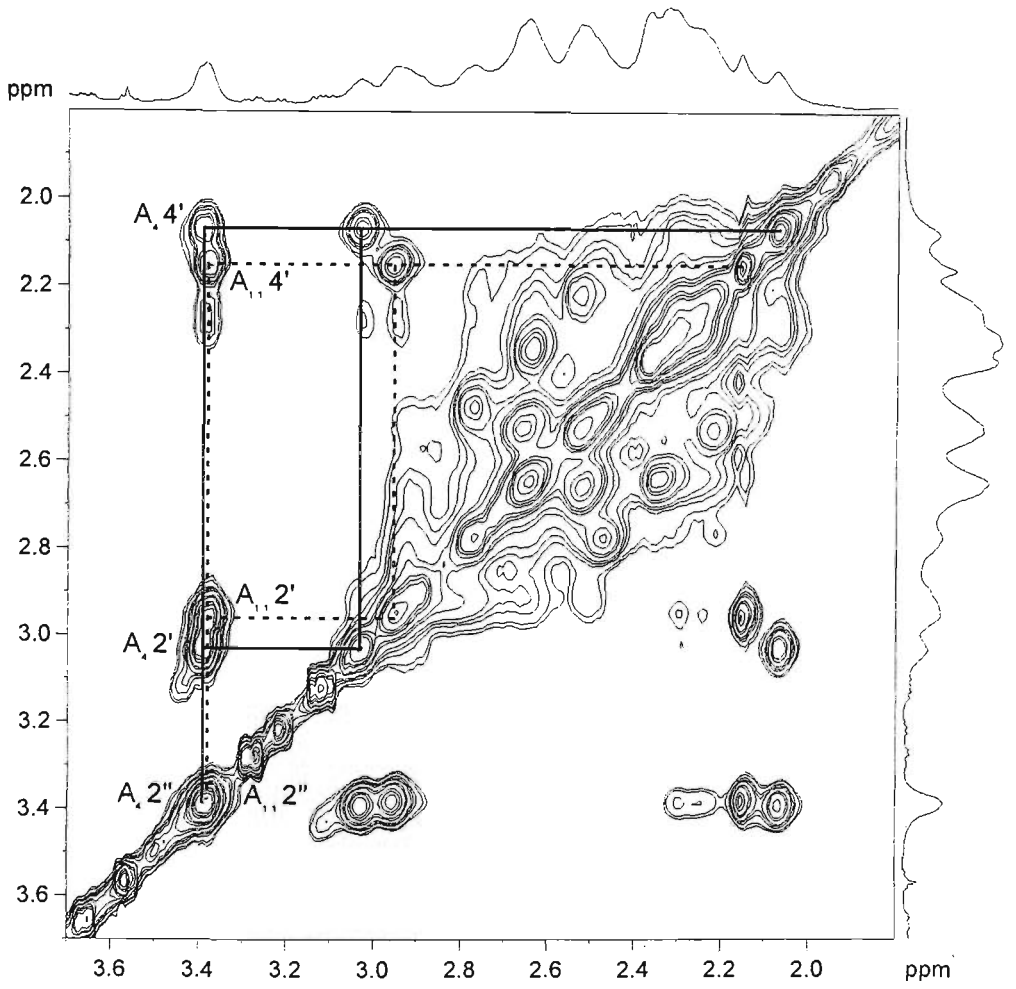


Figure 4.8: H4' to H2'/H2" section of the 300 ms NOESY spectrum for the loop sequence 5′-d(GCGAAGCACGAAGT)-3′ (3 mM) in D_2O containing 0.1 M NaCl/100 mM NaH_2PO_4 (pH 7.0). The connectivity between the H4' and H2'/H2" sugar protons are shown for the adenines involved in the 5′-GAA loop (A_4 and A_{11}).

Figure 4.9 shows the ^{1}H-NMR spectra in the H1'/H5 region of 5′-d(ACGAAGT)-3′ at a range of different temperatures. Upon examination of the H1' sugar protons in the 1D NMR spectrum, some resonances were observed to exhibit selective broadening effects.

For example, the H1' sugar protons of A_8, C_9, G_{13} and T_{14} gave sharp peaks, whilst those of G_{10}, A_{11} and A_{12}, *i.e.* those involved in the loop, gave broad signals in the spectrum recorded at 25 °C. This was postulated to be due to dynamic processes occurring in the loop region. The A_{11} base stacks on the guanine (G_{10}) as the oligonucleotide folds back on itself, however, the adenine appears to be in equilibrium between two conformations. This postulate was confirmed by the sharpening of these peaks at higher temperatures (figure 4.9 b-e). These results are consistent with those obtained by Hirao *et al.* who proposed that the adenine involved in the loop did not adopt any single conformation and "wobbles" around anti and high-anti conformations. (Hirao *et al.*, 1994)

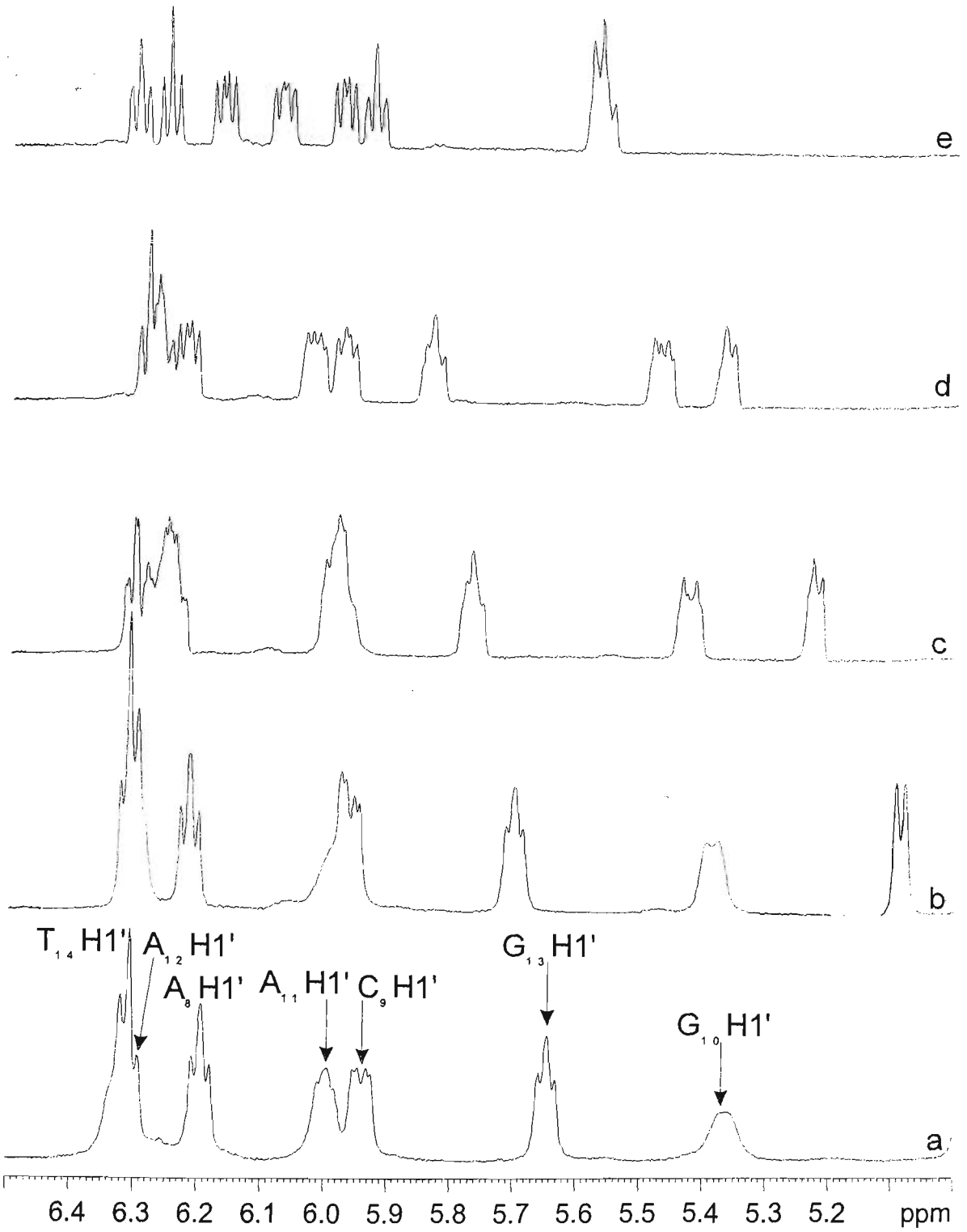


Figure 4.9: ^{1}H NMR spectra in the H1'/H5 region of 5'-d(ACGAAGT)-3' (3 mM) at 15 °C (a), 25 °C (b), 35 °C (c), 45 °C (d) and 55 °C (e) in D_2O containing 0.1 M NaCl / 100 mM NaH_2PO_4 (pH 7.0)

4.7 Melting Temperature Studies

The melting transition of both the free oligonucleotides (5′-d(ACGAAGT)-3′ and 5′-d(GCGAAGCACGAAGT)-3′) and the complex with nogalamycin were determined using a Pharmacia Biotech Ultrospec 2000 UV-Vis melting temperature apparatus. The concentration of the solutions were ~7 μM (0.5 μM per base) and the buffer used was 300 mM NaCl / 30 mM NaH_2PO_4. The absorbance at 260 nm was recorded as the annealed structure was heated from 25 °C to 100 °C at a rate of 1.0 °C / minute. The resulting change in absorbance with temperature is shown in figure 4.10. The hairpin was observed to undergo a broad melting transition with a T_m of 48.5 °C. UV melting studies on the intact double hairpin exhibited a broad, but intense absorbance change with a melting temperature calculated to be 65.3 °C. The hairpin containing the A-T base pair, 5′-d(ACGAAGT)-3′, was expected to have a lower T_m than the T_m reported for 5′-d(GCGAAGC)-3′ by Hirao *et al.* (Hirao *et al.*, 1994) owing to the weaker H-bonding interaction between A-T base pairs compared to G-C base pairs. The observed melting temperature of 65.3 °C was lower than had been previously reported for the hairpin, 5′-d(GCGAAGC)-3′ (*i.e.* 76 °C). (Yoshizawa *et al.*, 1997) The absence of the low temperature transition seen for the isolated hairpin provides further evidence for stabilisation through co-axial stacking between hairpins. It is postulated that the overlap of the melting transitions of the two hairpins is responsible for the unusual breadth of the melting transition for the loop sequence. Thus, the hairpins appear to behave independently with flexibility at the site of the nick.

The stem region of the 5′-ACGAAGT-3′ hairpin is not stable on its own as evidenced by the low, broad melting transition and the rapid attenuation of the imino protons in the 1D

spectra (more below in chapter 5). However, the co-axial stacking of the two hairpins dramatically stabilises the structure despite the break in the DNA backbone.

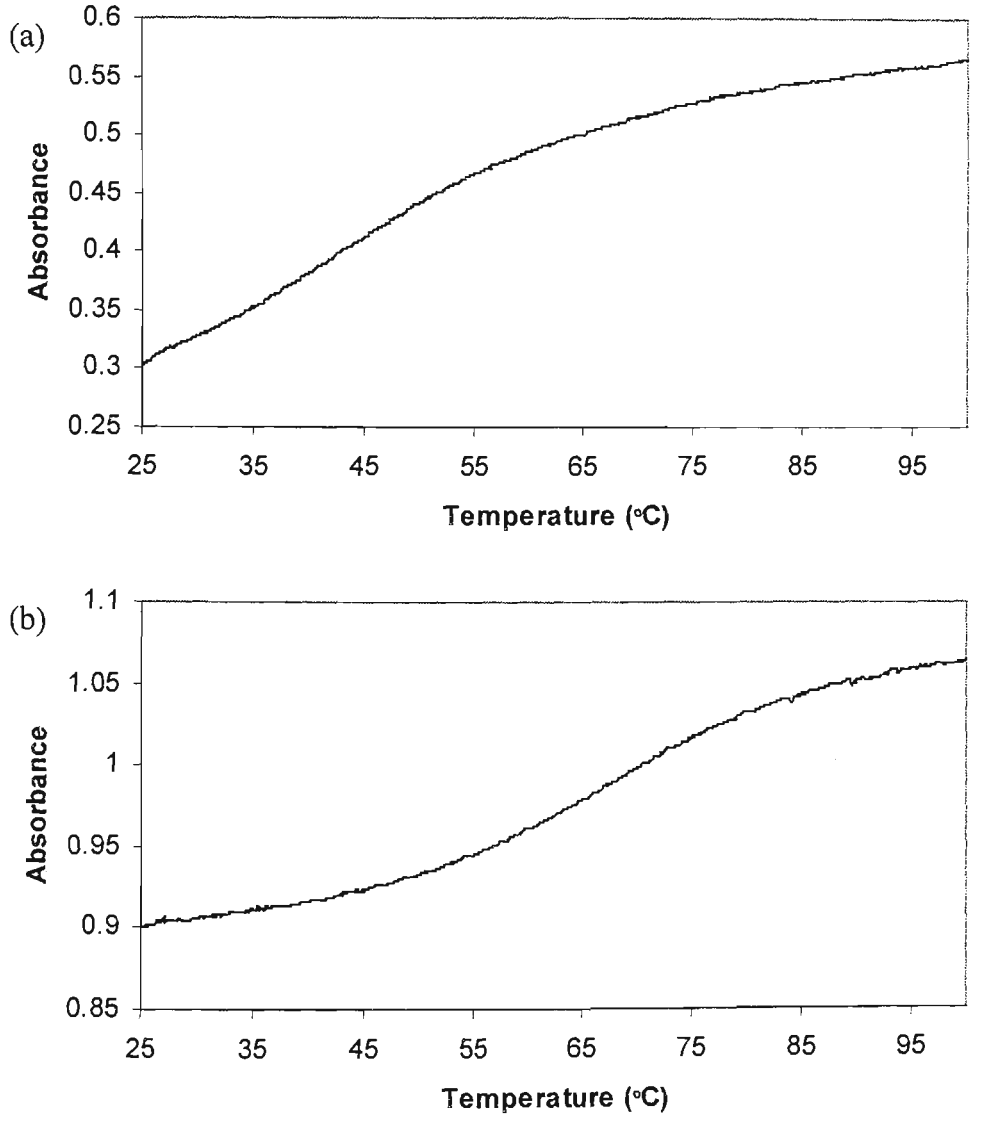


Figure 4.10: UV melting curves of (a) hairpin 5′-d(ACGAAGT)-3′ (3.5 μM) and (b) loop sequence 5′-d(GCGAAGCACGAAGT)-3′ (7 μM) in 300 mM NaCl/30 mM NaH_2PO_4 (pH 7.0).

4.8 Imino Protons as Evidence for the Formation of the Stem

Figure 4.11 shows the imino region of the 1D spectra for the hairpin (a) and loop sequence (b). There is one strong peak and a very broad, low intensity peak at 5 °C for 5′-d(ACGAAGT)-3′ corresponding to the G_{13} and T_{14} imino protons respectively. As the A_8-T_{14} base pair is terminal, it has less protection, hence the broad nature of the peak in the spectrum. The observation of these highly exchange broadened imino proton resonances for the Watson-Crick hydrogen bonded base pairs suggest that the stem region of 5′-d(ACGAAGT)-3′ is not particularly stable with the G_6NH only visible at low temperature. When the data for the isolated hairpin were compared with that for the double hairpin sequence, it was evident that for the latter, there were four sharp peaks resolved at 25 °C corresponding to the four Watson-Crick base pairs of 5′-d(GCGAAGCACGAAGT)-3′. The assignment of the imino protons was carried out by examining the sequential NOEs (more below). The thymine imino (T_{14}NH) at the break site showed some evidence of exchange broadening. Figure 4.22 shows a study on the temperature dependence of the line widths, which illustrates that the 5′-d(ACGAAGT)-3′ hairpin melts first with the overall process being biphasic, showing that the substitution of an A-T base pair in the stem significantly destabilises the structure. The co-axial stacking of the two hairpins at the nicked site may account for the increased stability of the terminal base pairs (A_8-T_{14} and G_1-C_7) and hence the 5′-d(ACGAAGT)-3′ sequence.

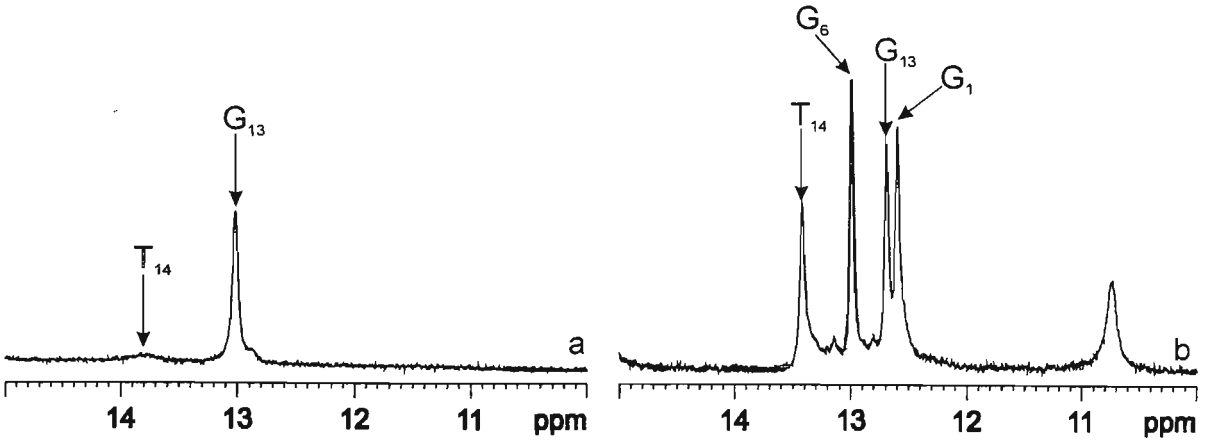


Figure 4.11: 1D NMR spectra at 5 °C of hairpin (a) and loop sequence (b) showing the imino protons. The identities of the imino protons were confirmed by 2D-Watergate-NOESY experiments as detailed in section 4.15.

Further evidence for the formation of the stem region comes from the observation of the protected imino protons when the 1D ^{1}H-NMR is acquired in water (10% D_2O). These peaks are only observed if the imino protons are involved in hydrogen bonding interactions and hence are protected from rapid exchange with the water.

The absence of the G_3 and G_{10} imino protons involved in the G-A base pairs in the loops indicate that the base paired conformation does not protect the guanine imino protons from exchange. The exchange-broadened peak observed at 10.8 ppm is thought to correspond to the amino protons of the adenines involved in the loops and is consistent with the proposed structure of the G-A base pair.

4.9 Chemical Shift Change with Temperature

The melting of a nucleic acid double helix into a single stranded molecule is a well-known phenomenon and the melting behaviour can also be followed by a temperature-dependent study of NMR spectra. The temperature could only be increased to a maximum of approximately 70 °C owing to NMR technical limitations and as such the full melting curve could not be obtained. However, these experiments showed agreement with the results of the UV spectroscopy experiments (figure 4.10), *i.e.* the oligonucleotides began to melt at ~ 35 °C.

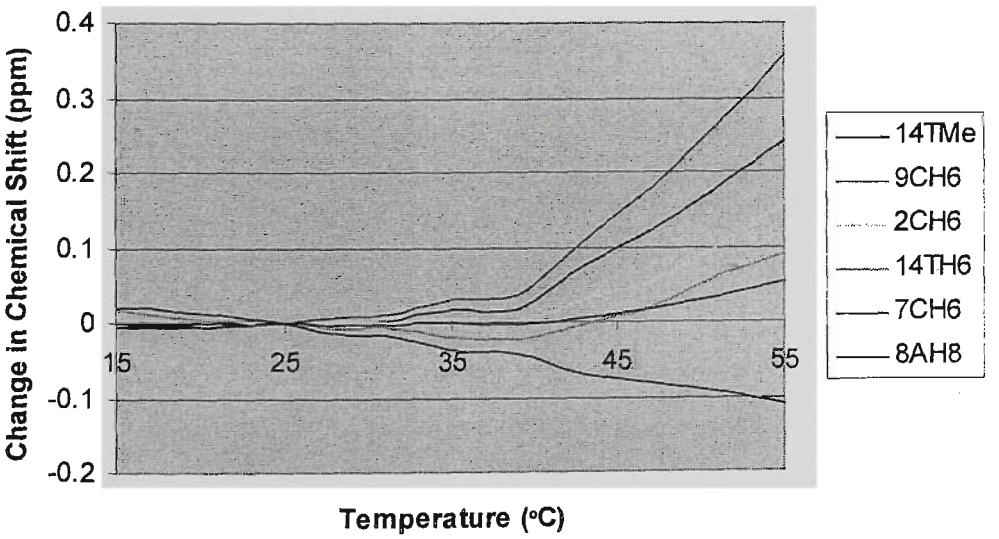


Figure 4.12: Chemical shift change of a number of DNA resonances with increased temperature. The data presented here has been normalised to the data collected for 25 °C.

Figure 4.12 shows the change in chemical shift of a number of selected DNA resonances observed in the NMR spectra with increasing temperature. The chemical shifts of the aromatic base protons are affected significantly as the oligonucleotide undergoes the

transition from a double stranded to a single stranded conformation. The methyl resonances of thymine in ssDNA are shifted down-field by 0.24 ppm relative to those found in dsDNA.

4.10 Sugar Conformation

The deoxyribose sugar units serve as a semi-flexible linkage between the nucleobases and the phosphate backbone. The conformation of the sugar is described in terms of pseudorotation and glycosidic torsion angles. Two stable ring conformations have been identified: the North (N) conformation, with a pseudorational phase angle (P) of 0°, and the South (S) conformation at 180°. The deoxyribose adopts a C3'-*endo* arrangement for the N-conformation and a C2'-*endo* arrangement for the S-conformation. In practice, nucleic acids contain deoxyribose sugar in N- or S-type conformations, *i.e.* a narrow range of pseudorotational phase angles are allowed (3-34° and 139-175° respectively).

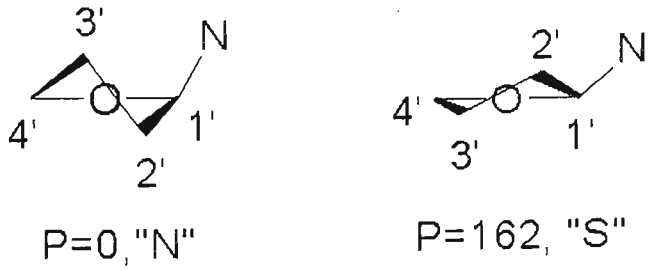


Figure 4.13: Stable conformations of deoxyribose: (a) North and (b) South

The experimentally determined coupling constants (table 4.1), $J_{H1', H2'}$ and $J_{H1', H2''}$, or alternatively, the combined value, ΣH1', show that the sugar residues, A_8, C_9, A_{11} and G_{13} within the hairpin, 5'-d(ACGAAGT)-3', take the C2'-*endo* sugar pucker conformation (table 4.1). (Altona & Sundaralingam, 1972) This indicates that even these unusual DNA

structures exist in the B-form. It was difficult to determine all coupling constants experimentally, however, owing to overlap of the resonance signals in the 1D NMR spectra. The results obtained here agree with those in an earlier study by Hirao and co-workers. (Hirao *et al.* 1994) The sugar geometries for all nucleotides were also determined following the method of Hosur *et al.* (Hosur *et al.* 1986) utilizing DQF-COSY data. The intensities of the peaks in the COSY spectra are approximately proportional to the values of the coupling constants (*J*) between the two nuclei. The three bond (vicinal) coupling constants between sugar protons H1', H2', H2", H3' and H4' depend, in a predictable manner, on the sugar geometry. Thus, from the analysis of the relative intensities of the cross peaks, it was possible to determine the sugar pucker conformation. The conformation was found to be C2'-*endo* for all nucleotides with the exception of the terminal T_{14}, which was observed to take the C3'-*endo* conformation, typical of terminal sugar residues.

Table 4.1: Determination of sugar pucker conformation based on coupling constants for H1' to H2'/H2".

Sugar H1'	ΣH1' (Hz)[a]	Sugar Pucker[b]	% S[c]
8A	14.74	S	84.9
9C	15.37	S	91.9
11A	15.10	S	88.9
13G	14.46	S	81.8

[a]*Sum of all J couplings contributing to the H1' ΣH1' = $J_{H1',H2'} + J_{H1',H2''}$. If ΣH1' > 13.3 Hz, there is a predominance of South (pS > 60%) or if ΣH1'< 12.5 Hz, there is a predominance of North (pS < 40%). (Altona & Sundaralingam, 1973)*

[b]*The predominant sugar conformation based on the values obtained for ΣH1'.*

[c]*Percentage of S confomer (Altona & Sundaralingam, 1973)*

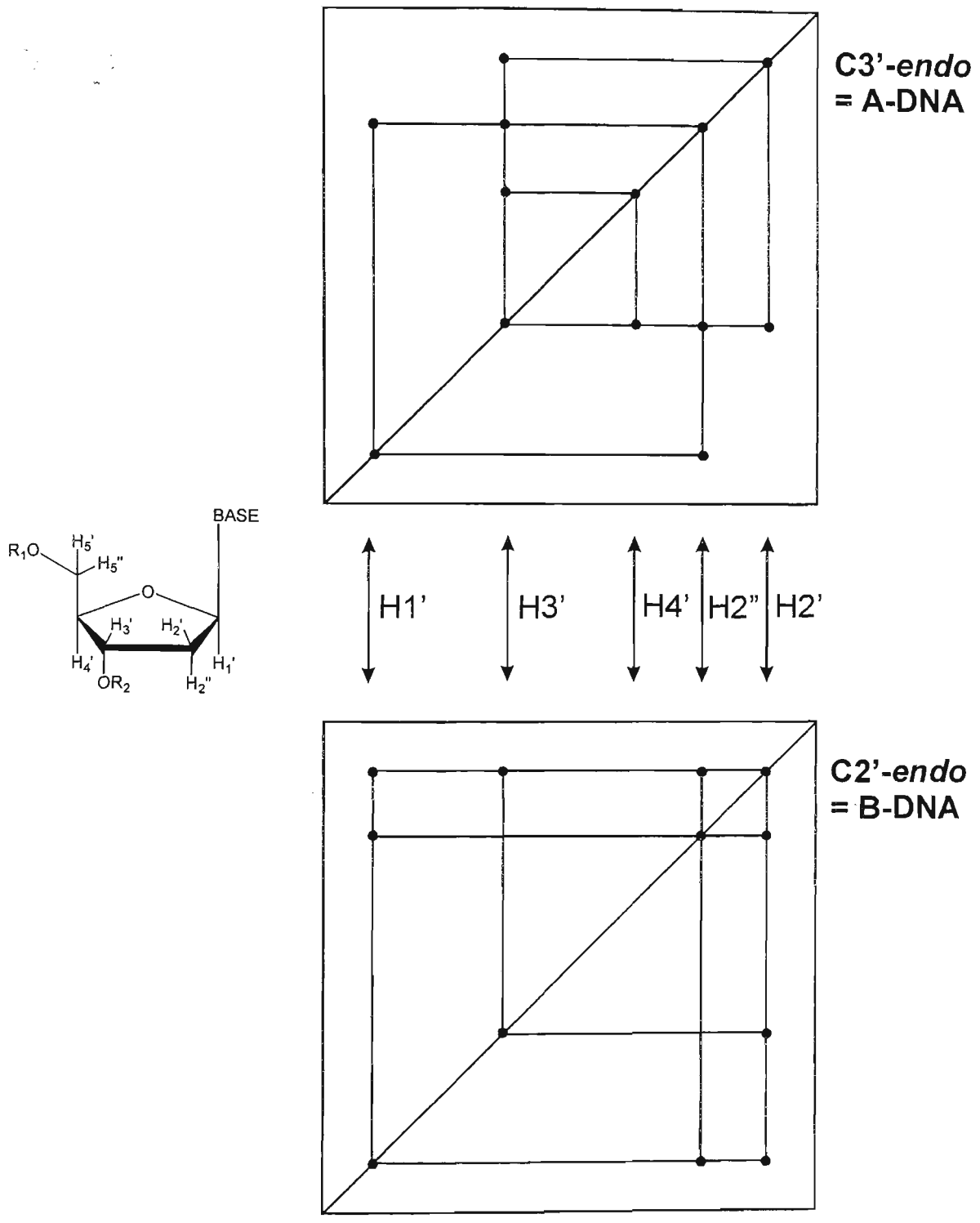


Figure 4.14(a): Theoretical DQF-COSY patterns for the coupling observed between sugar protons in A-form and B-form DNA.

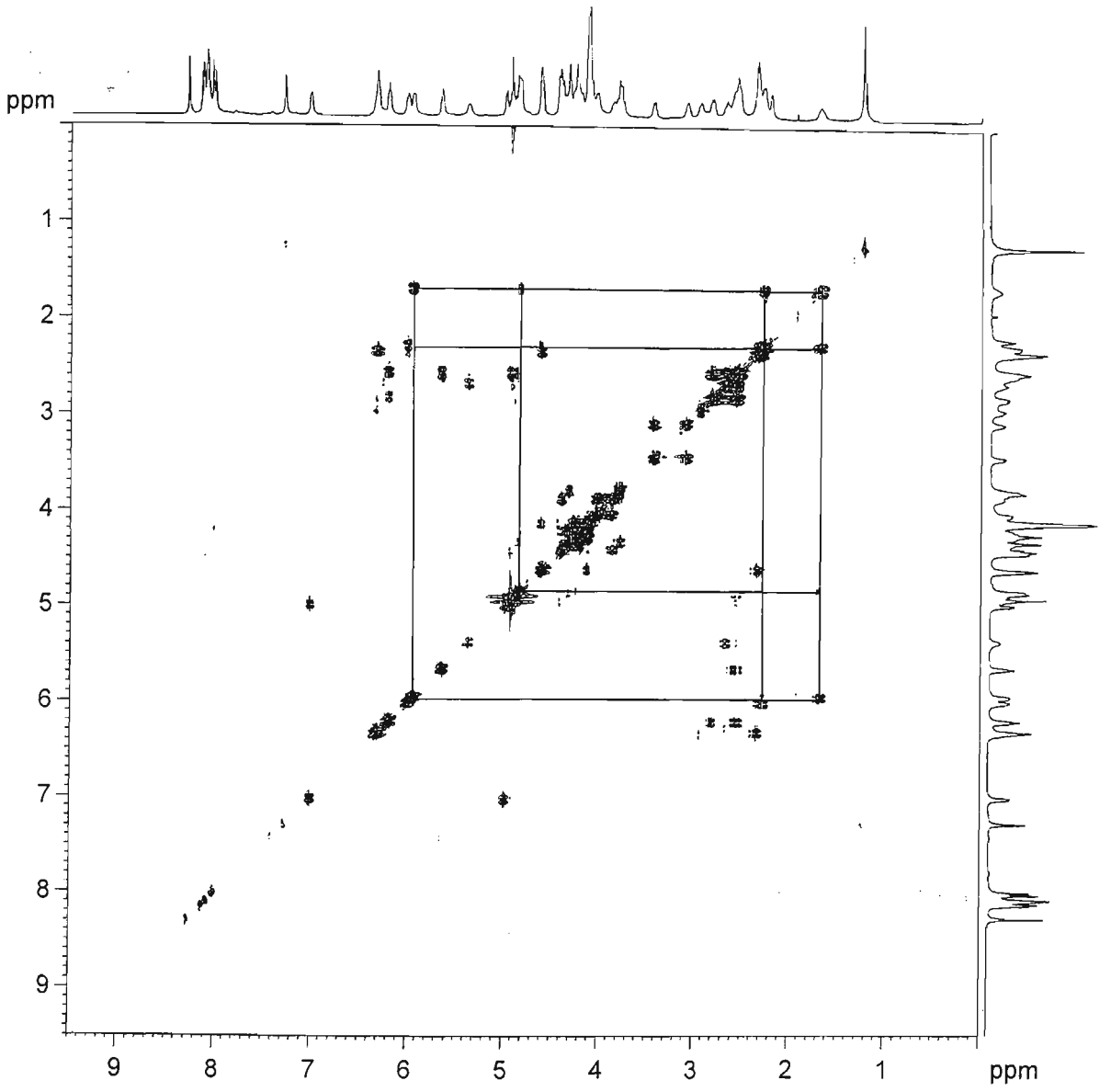


Figure 4.14(b): DQF-COSY spectrum of 5′-d(ACGAAGT)-3′ (3 mM) in D_2O containing 0.1 M NaCl/100 mM NaH_2PO_4 (pH 7.0) showing coupling of sugar protons for C_2.

An example of the method for determining sugar conformation from the DQF-COSY data (Hosur *et al.*, 1986) is shown in figure 4.14. The theoretical patterns for the C2'-*endo* and 3'-*endo* conformations are illustrated in figure 4.14(a), whilst the DQF-COSY spectrum acquired for 5′-d(ACGAAGT)-3′ is shown in figure 4.14(b). The pattern observed for the

C_2 sugar resonances have been highlighted as an example of this method. The figure clearly shows that the C_2 sugar protons give the same pattern as the C2'-*endo* conformation and is further evidence that the stem sections of the loop sequence are B-form like. This method was also applied to the determination of the sugar conformation for the loop sequence. The sugar protons involved in Watson-Crick base pairs in the stem region gave similar patterns (not shown) to those shown in figure 4.14(b), which is further evidence that the stem region is B-form like. The bases involved in the loop could not be analysed by these methods owing to the unusual chemical shift values of the H4' and H2'/H2" and increased peak widths.

4.11 Titration of 5'-d(GCGAAGCACGAAGT)-3' with Nogalamycin

The titration experiment resulted in the formation of a 1:1 complex between nogalamycin and 5'-d(GCGAAGCACGAAGT)-3'. The following section discusses the changes in the oligonucleotide spectra upon addition of the ligand. Subsequent to the titration, 2D NMR experiments allow the identification of ligand-DNA contacts and any changes in DNA conformation upon ligand incorporation.

Figure 4.15 shows the 1D NMR spectra resulting from the addition of aliquots of nogalamycin up to and including the formation of a 1:1 complex between 5'-d(GCGAAGCACGAAGT)-3' and nogalamycin. The addition of nogalamycin to 5'-d(GCGAAGCACGAAGT)-3' causes significant changes to the ^{1}H-NMR spectrum. Immediately upon addition of the ligand, a number of the free DNA resonances diminish in intensity and new peaks appear in the ^{1}H-NMR spectrum. This is most easily observed in

the aromatic region (figure 4.15). The free and bound signals are in slow exchange at 25 °C with the change in complexity of the NMR spectrum indicative of the formation of a single bound species.

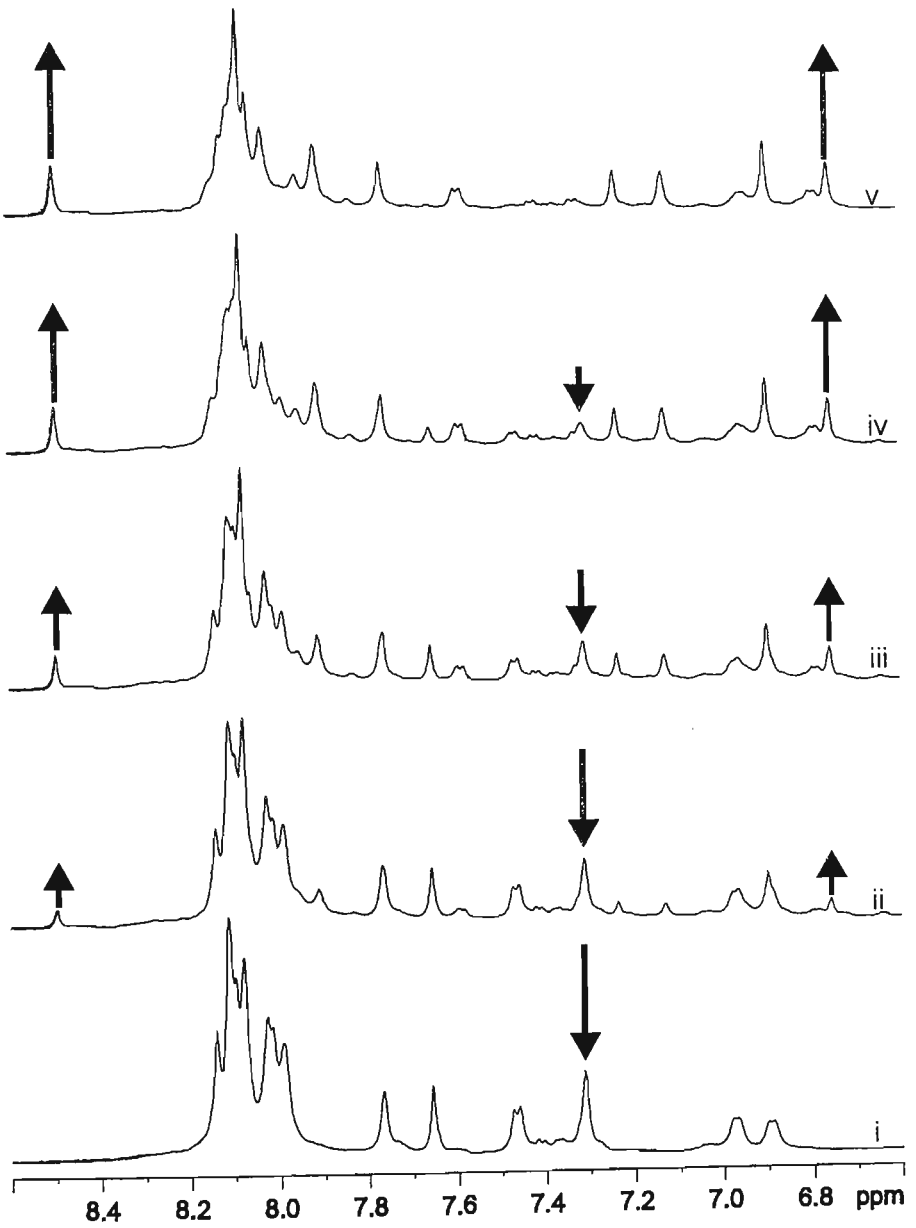


Figure 4.15(a): 1D stacked plot spectra (6.6-8.6 ppm) acquired after the addition of aliquots of a 5 mM solution of nogalamycin: (i) 0:1 (ii) 0.25:1 (iii) 0.5:1 (iv) 0.75:1 (v) 1:1. Arrows highlight a number of clearly defined signals, owing to free DNA signals that decrease, as well as increasing signal intensity owing to complex formation.

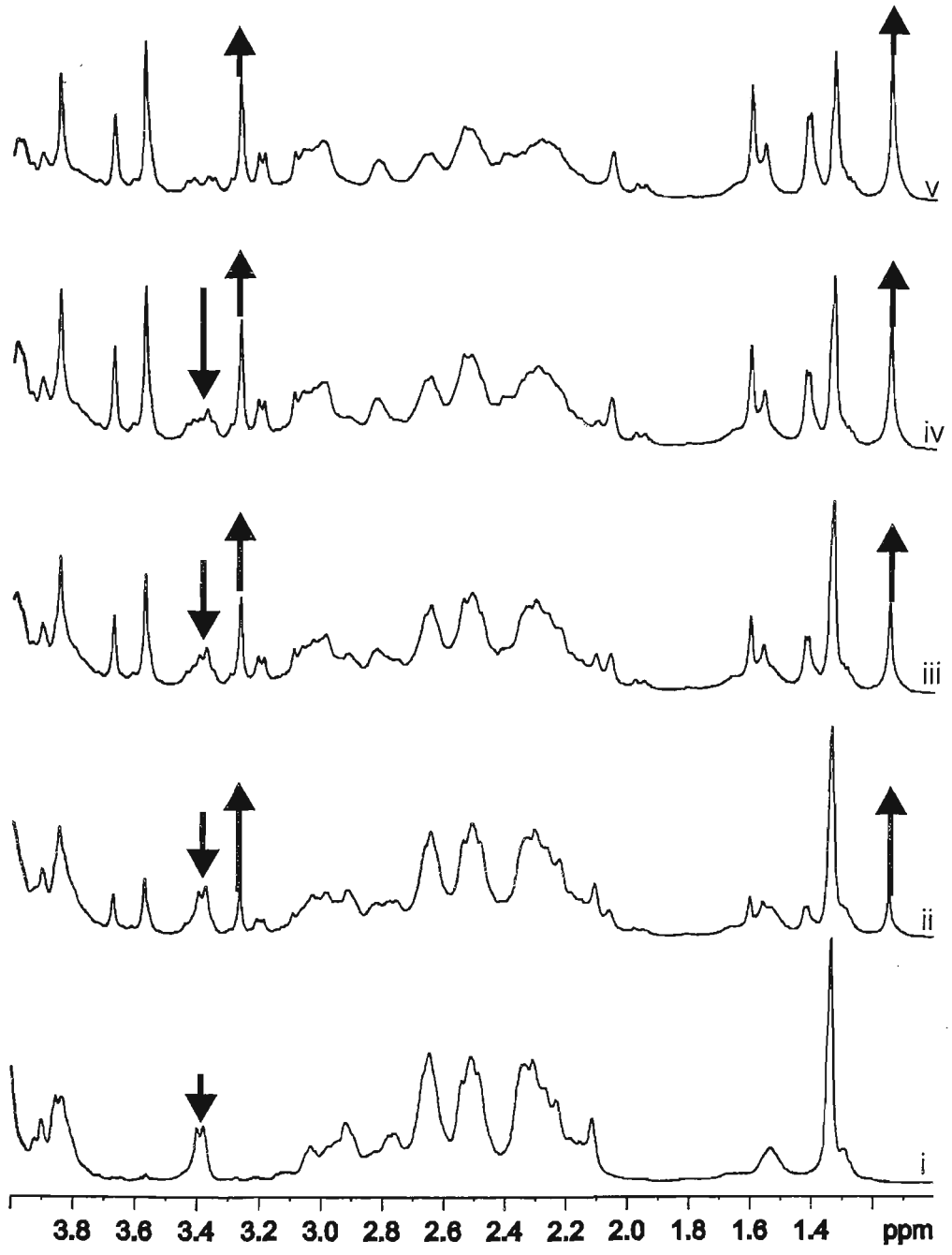


Figure 4.15(b): 1D stacked plot spectra (1.0-4.0 ppm) acquired after the addition of aliquots of a 5 mM solution of nogalamycin: (i) 0:1 (ii) 0.25:1 (iii) 0.5:1 (iv) 0.75:1 (v) 1:1. Arrows highlight a number of clearly defined signals, owing to free DNA signals that decrease, as well as increasing signal intensity owing to complex formation.

Figure 4.15(a) shows the aromatic region of the 1D spectra acquired after titration of the antibiotic nogalamycin into the loop sequence, 5′-d(GCGAAGCACGAAGT)-3′. The peak at ~8.5 ppm corresponds to the aromatic H8 proton of A_{11}. The chemical shift change is typical of the H8 of the adenine at the 5′-CpA intercalation site seen in previous studies. (Williams & Searle, 1999) Another peak highlighted at ~6.75 ppm corresponds to the H11 proton of the nogalamycin. In addition to the emergence of new peaks as the drug is titrated into the solution, a number of peaks corresponding to proton resonances from the single stranded oligonucleotide are attenuated. For example, the peak at ~7.3 ppm represents the H6 proton of T_{14} in the free oligonucleotide. The peak appearing at ~7.15 ppm represents the same proton from the complex. It was observed that the intensity of the latter peak increased as the intensity of the former diminished and that the free and bound forms are in slow exchange with each other.

Figure 4.15(b) shows the stacked plot of the aliphatic region of the 1D spectra incorporating the methyl and methoxy resonances. In the methyl region, the single methyl resonance from $T_{14}CH_3$ at ~1.15 ppm decreases in intensity upon binding of the ligand. In addition a number of new peaks appear between 1.3 and 1.6 ppm corresponding to the appearance of the nogalamycin methyl peaks. The ligand methyl peaks do not correspond with the chemical shifts of the free nogalamycin methyl resonances and thus must arise as a result of the bound ligand. The nogalamycin methoxy resonances also appear between 3.2 and 3.9 ppm. The appearance of the 3'-methoxy of the nogalose sugar of nogalamycin at ~3.3 ppm has been highlighted in the 1D spectra.

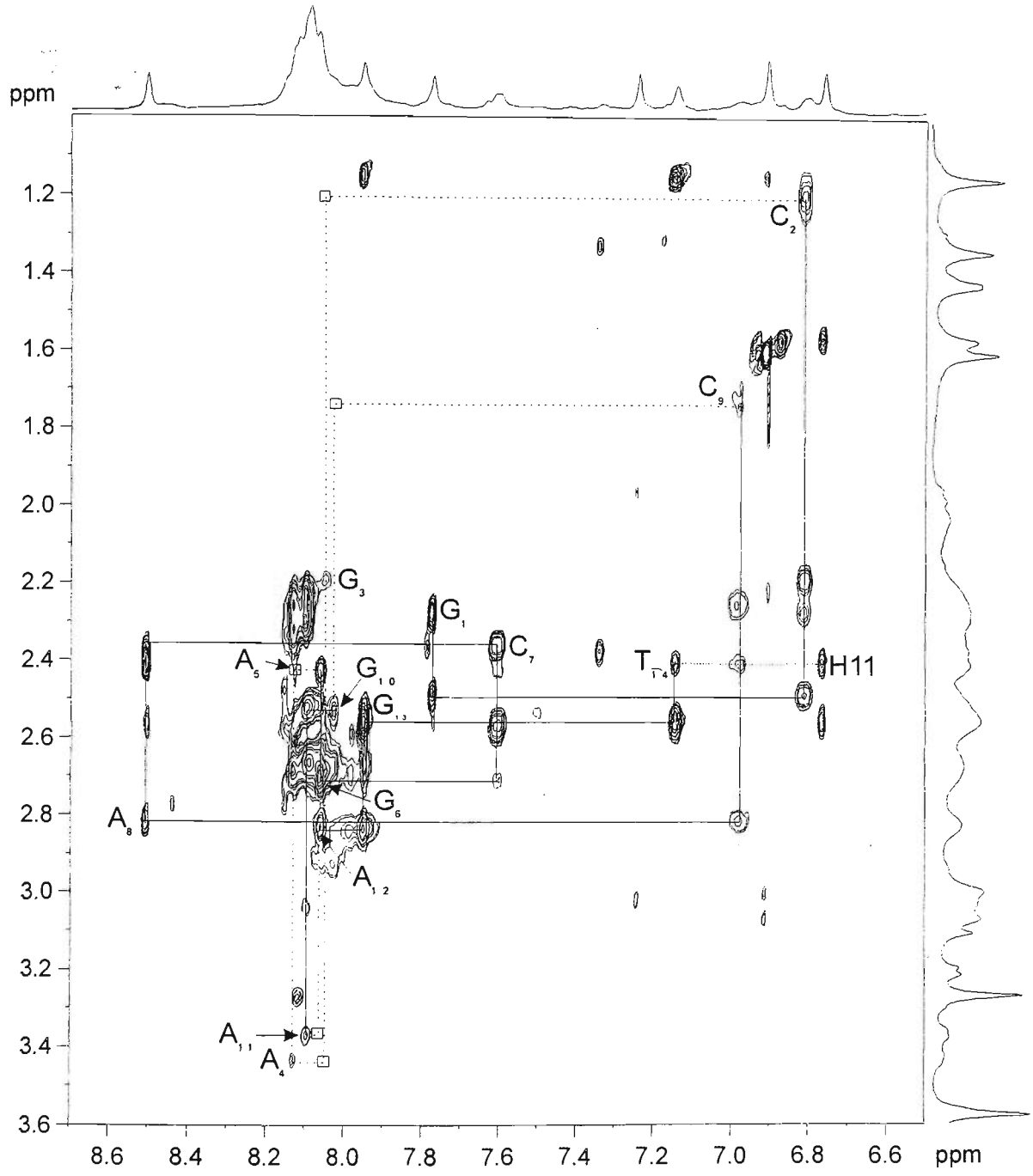


Figure 4.16: Section of the 300 ms NOESY spectrum of the 5'-d(GCGAAGCACGAAGT)-3'-Ng complex (3 mM) in D_2O containing 0.1 M NaCl/100 mM NaH_2PO_4 (pH 7.0) showing the sequential H8/H6-H2'/H2" connectivity along the DNA backbone.

Figure 4.16 shows the H8/H6-H2'/H2" sequential connectivity observed (solid lines) for the 5'-d(GCGAAGCACGAAGT)-3'-Ng complex. The dashed lines indicate the loss of sequential connectivity as the backbone is contorted in the loop regions. The squares show where the NOEs would have been observed if the sequence had adopted a 'normal' double stranded conformation. The dotted lines show the interaction of the T_7 nucleobase with the H11 proton of nogalamycin. The spectrum here differs from that of the free oligonucleotide in that there are a number of resonances, which have been up-field or down-field shifted as a result of ligand binding. In addition, internucleotide NOEs across the intercalation site between G_1 and T_{14} as observed for the free oligonucleotide are no longer evident. The drug, however, acts as a surrogate base pair with the aromatic H11 proton giving NOEs to both G_1H1' and T_{14}H1', consistent with intercalation between the hairpins.

4.12 Chemical Shift Perturbations upon Ligand Binding

The examination of chemical shift perturbations to the DNA resonances can provide an indication of the binding site of a ligand. Generally, the largest perturbations were noted for the T_{14}, G_1, A_8 and C_7, *i.e.* at the intercalation site.

Table 4.2 shows the changes in chemical shift for the aromatic base protons (H8/H6) and sugar H1' resonances upon binding of the ligand. A negative value is indicative of a down-field shift, whilst a positive value indicates an up-field shift. The table is arranged in the loop-type conformation thought to be adopted by the oligonucleotide to allow direct

comparison of chemical shift differences and location within the sequence and hence structure.

The A_8H8 and C_7H6 resonances were observed to be down-field of the shifts noted for the free loop sequence, whilst the $T_{14}H6$ was shifted up-field and the G_1H8 showed no change (table 4.2). This is consistent with drug aglycone stacking preferentially with the T_{14} and G_1 bases. (Searle & Wakelin, 1990) Deoxyribose sugar protons show similar trends with the H1' protons best resolved. The largest change was observed for the minor groove-oriented protons localised to the base pairs either side of the intercalation site. The C_7H1' and A_8H1' protons were all shifted down-field, whilst the G_1H1' and $T_{14}H1'$ were up-field shifted. These data are consistent with drug aglycone stacking preferentially with the T_{14} and G_1 bases. (Searle & Wakelin, 1990)

Table 4.2: Change in chemical shift upon nogalamycin binding: (a) H8/H6 and (b) H1'.

(a) H8/H6	12A	13G	14T	1G	2C	3G	
11A	-0.056	-0.076	-0.195	-0.010	-0.176	-0.061	4A
0.013	10G	9C	8A	7C	6G	5A	0.016
	0.000	0.000	0.355	0.108	-0.004	0.000	

(b) H1'	12A	13G	14T	1G	2C	3G	
11A	-0.049	-0.071	-0.051	-0.395	-0.069	-0.082	4A
0.028	10G	9C	8A	7C	6G	5A	0.048
	0.025	0.129	0.269	0.147	0.891	0.000	

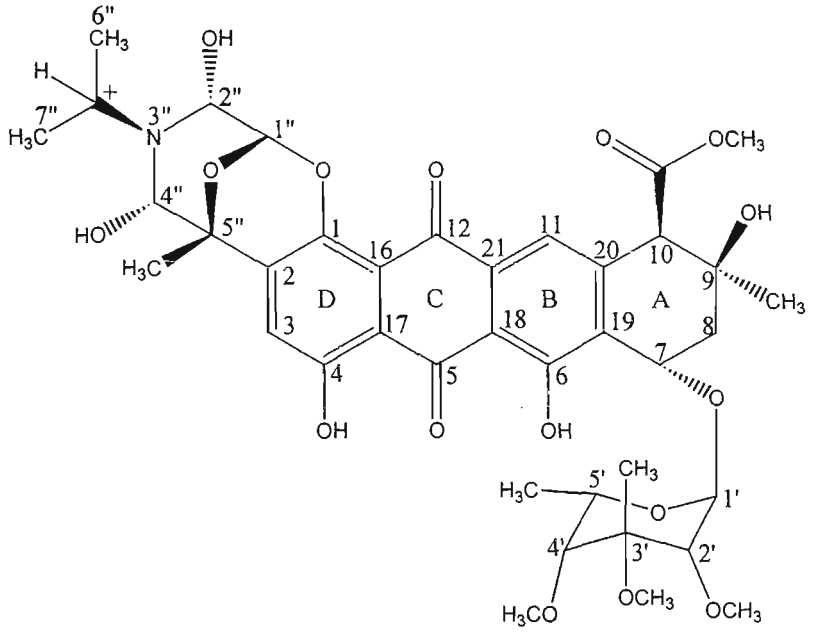


Figure 4.17: Structure of nogalamycin showing the numbering scheme of nogalamycin

The assignment of the ^{1}H-NMR spectrum of nogalamycin was undertaken using a combination of NOESY, TOCSY and DQF-COSY prior to the complex in order to gain an understanding of the drug-drug NOEs that were observed in the NMR spectra. In this way, the resulting NMR spectra of the complex were simplified by elimination of the DNA-DNA and drug-drug NOEs, leaving only the drug-DNA NOEs. Table 4.3 shows the NOEs that were observed in the NOESY spectra and have been separated into the three distinct regions of the ligand: the nogalose sugar, the aglycone chromophore and the bicyclic sugar. Generally, drug-drug NOEs were observed between the protons of the nogalose sugar (table 4.3a), between the protons of the aminoglucose sugar (table 4.3c) and between protons present on the aglycone chromophore (table 4.3b), although some drug-drug NOEs were noted between these three distinct regions of the ligand. When referring to protons on the ligand, the nomenclature as shown in figure 4.17 was used. Where two protons or groups,

i.e. methyl or methoxy groups, were present, the proton or group extending up from the plane of the chromophore was given the identity (1) and those pointing down were assigned as (2). For example, on the 3' carbon of the nogalose sugar, two groups are present (methyl and methoxy groups). The methyl protons are referred to as NgH3'(1) while the methoxy protons are referred to as NgH3'(2) owing to their spatial orientation.

Table 4.3(a): Drug-drug NOEs observed in NMR spectra: nogalose sugar protons

	NgH1'	NgH2'(1)	NgH2'(2)	NgH3'(1)	NgH3'(2)	NgH4'(1)	NgH4'(2)	NgH5'(1)	NgH5'(2)
NgH1'		•	•	•	•		•	•	•
NgH2'(1)	•		•	•	•				
NgH2'(2)	•	•		•	•				
NgH3'(1)	•	•	•		•	•		•	
NgH3'(2)	•	•	•	•					
NgH4'(1)				•			•		•
NgH4'(2)	•			•		•		•	•
NgH5'(1)	•			•			•		•
NgH5'(2)	•					•	•	•	
NgH7	•	•	•		•			•	•
NgH8(1)	•			•				•	•
NgH8(2)	•			•				•	•
NgH9	•	•		•				•	

Table 4.3(b): Drug-drug NOEs observed in NMR spectra: protons on the chromophore

	NgH3	NgH7	NgH8(1)	NgH8(2)	NgH9	NgH10	NgH11
NgH3							
NgH7			•	•	•	•	
NgH8(1)		•		•	•	•	
NgH8(2)		•	•		•	•	
NgH9			•	•		•	•
NgH10					•		•
NgH11					•	•	

Table 4.3(c): Drug-drug NOEs observed in NMR spectra: aminoglucose protons

	NgH1"	NgH2"	NgH3"	NgH4"	NgH5"	NgN3"(1)	NgN3"(2)
NgH1"		*	*	*	*	*	*
NgH2"	*			*	*	*	*
NgH3"	*				*	*	*
NgH4"	*	*			*	*	*
NgH5"	*	*	*	*			
NgN3"(1)	*	*	*	*			
NgN3"(2)	*	*	*	*			
Other					NgH3	NgH3	

The assignment of the drug and DNA resonances provides the starting point for identifying intermolecular NOEs that define the position and orientation of the bound antibiotic.

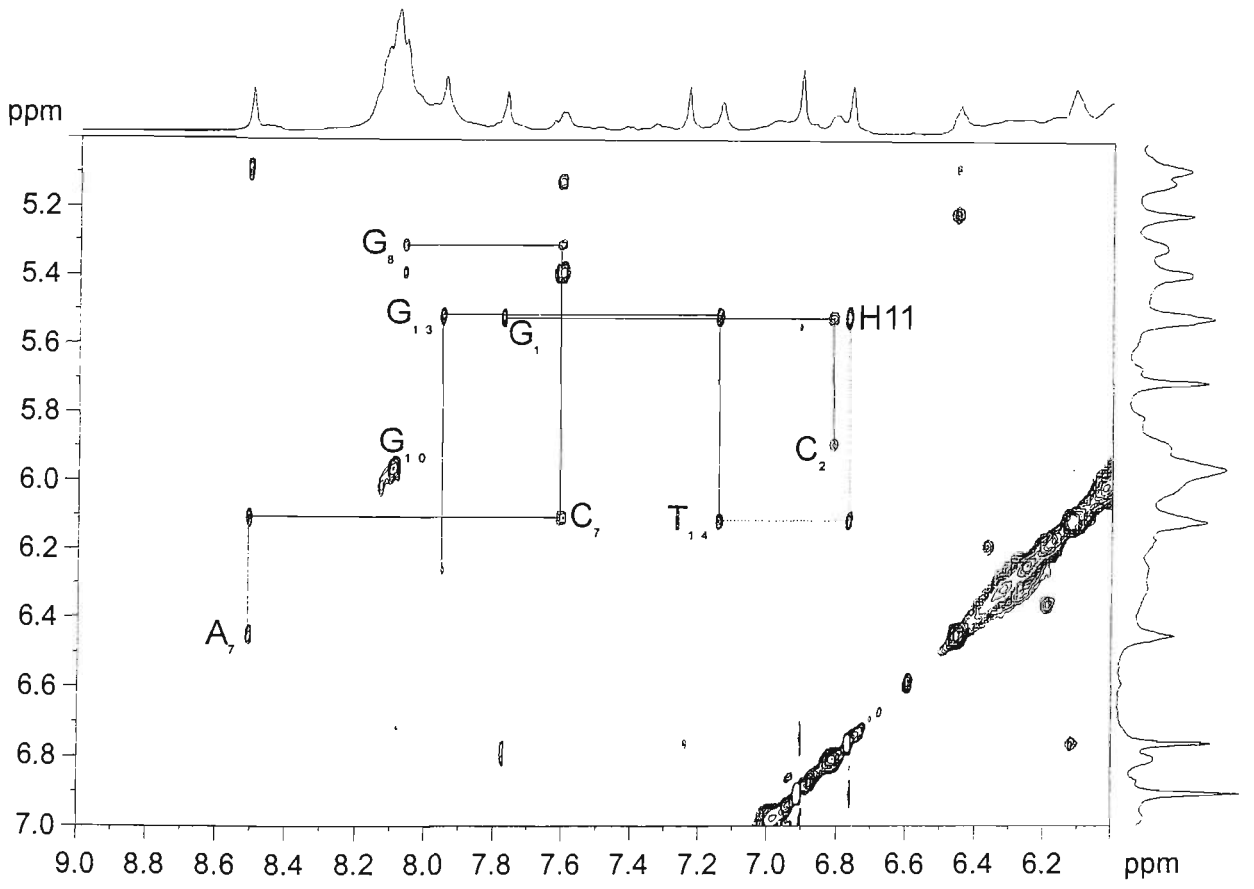


Figure 4.18: Section of the 300 ms NOESY spectrum of 5′-d(GCGAAGCACGAAGT)-3′ (3 mM) in D_2O containing 0.1 M NaCl/100 mM NaH_2PO_4 (pH 7.0) highlighting drug-DNA NOEs that clearly show the intercalation site.

Figure 4.18 shows a portion of the NOESY spectrum highlighting the NOEs that are important in defining the position and orientation of the ligand within the intercalation site. The solid lines represent the connectivity through G_{13}-T_{14} and G_1-C_2, whilst the dotted lines show the connectivity through the ligand, *i.e.* T_{14}-Ng-G_1. The solid lines show the connectivity through the stem region on the other strand, *i.e.* G_6-C_7-A_8. The data in figure 4.18 shows the sequential connectivity along the DNA backbone is lost. There are several reasons for this. First, the unusual geometry of the loops causes a loss of sequential connectivity in these regions. The intercalation of nogalamycin between the TpG(CpA) bases in the stem region also results in no NOE crosspeaks being observed between T_{14} and G_1 as were observed in the free oligonucleotide. NOE crosspeaks were observed, however, between T_{14}H1' and the H11 of nogalamycin and between Ng-H11 and G_1H1' as shown in figure 4.18.

The antibiotic aglycone H11 provides an unambiguous indicator of the position and orientation of the drug. NOEs are identified to the deoxyribose protons T_{14}H1' (T_{14}H1'-NgH11) and G_1H1' (NgH11-G_1H1') and can only be satisfied by intercalation of the drug at the 5'-TpG step such that the nogalose sugar points toward the centre of the duplex. A large number of other intermolecular NOEs support this conclusion. For example, the methoxy and methyl groups on the nogalose and ring A sugar provide evidence for hydrophobic interactions in the minor groove and enable the precise determination of the position and orientation of the nogalose with respect to the intercalation site. For example, NOEs from drug methyl protons on the nogalose sugar show specific hydrophobic interactions with G_1H1' and C_2H1' on the floor of the minor groove. NOEs are detected

from the nogalose H1' and H2' protons to the sugar and aromatic protons of C_7 and A_8, *i.e.* NgH1'-A_8H1'/H4'/H5', NgH2'-A_8H1'/H4'/H8 and NgH2'-C_7H1'/H6. In contrast, the number of contacts in the major groove is rather limited. NOEs from the aminoglucose protons to the T_{14} protons predominantly, *i.e.* NgH1"-T_{14}H6/Me and NgH2"-T_{14}Me. These characteristic NOEs are analogous to those observed for nogalamycin binding to the 5'-TpG step in duplex DNA. (Williams & Searle, 1999)

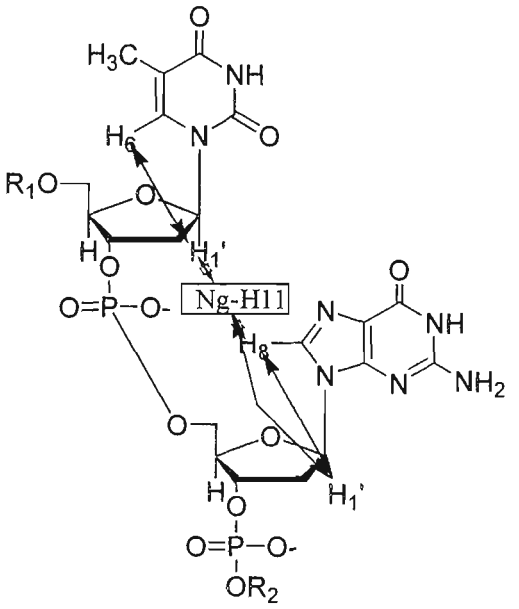


Figure 4.19: Schematic of the intercalation of nogalamycin between the 5'-TpG step.

Figure 4.19 shows the NOEs observed between the thymine and guanine bases involved in the intercalation site and the nogalamycin chromophore. A loss of sequential connectivity along the DNA backbone is observed, although connectivity can be followed via the H11 of nogalamycin (as indicated by the blue arrows).

4.13 Other Drug-DNA NOEs

Approximately 39 NOEs between the nogalamycin and the loop sequence were observed at 300 ms mixing time. This number decreased to 16 as the mixing time was reduced to 100 ms.

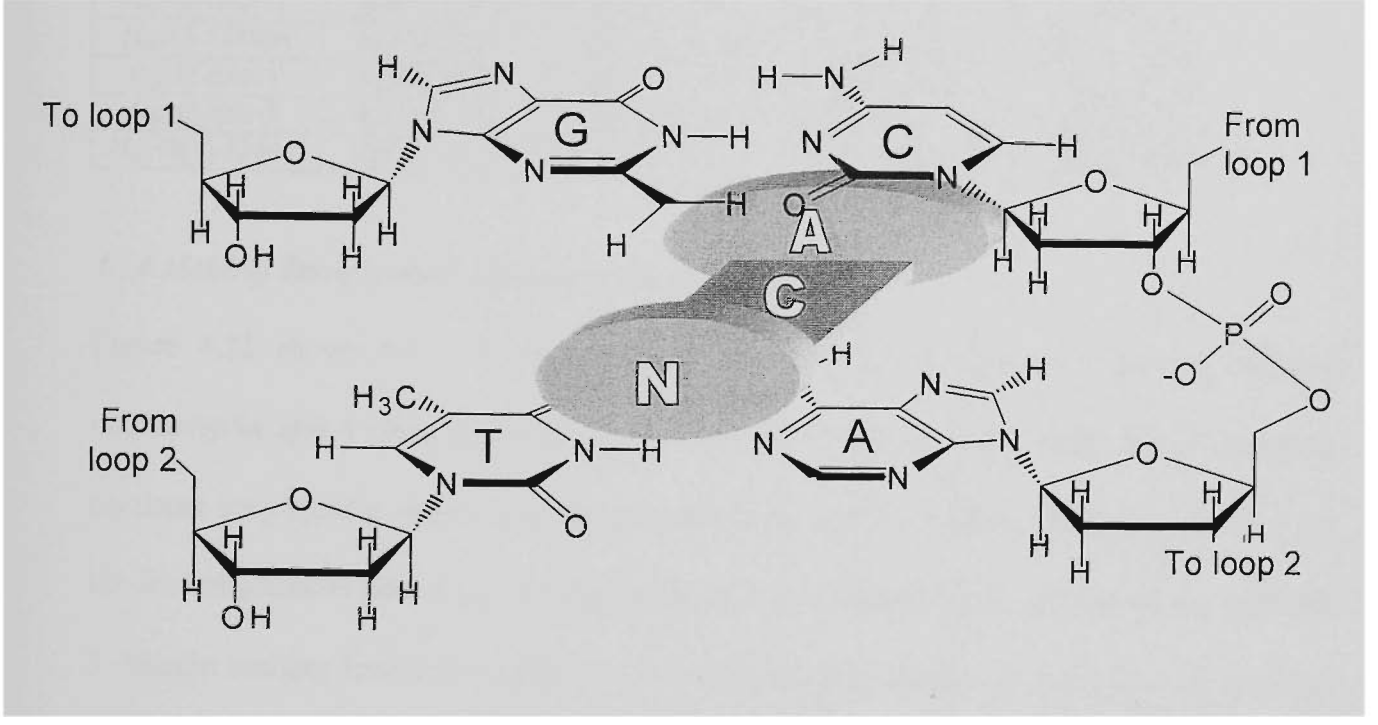


Figure 4.20: Schematic illustration of the intercalation site within the 5′-d(GCGAAGCACGAAGT)-3′-nogalamycin complex

Figure 4.20 is a schematic illustration of the nogalamycin bound at the TpG step showing the position of the nogalose and aminoglucose sugar groups in relation to the flanking base pairs. The planar chromophore stacks between the base pairs and the two sugar substituents wrap around the GC base pair.

Table 4.4: Drug-DNA NOEs observed in the 300 ms mixing time of the NOESY spectrum of the 5′-d(GCGAAGCACGAAGT)-3′-nogalamycin complex

NgH1"-14TH6	NgH8(2)-8AH1'	NgH11-1GH1'	NgH1'-8AH1'	NgH2'(2)-8AH8
NgH1"-14TMe	NgH9-1GH1'	NgH11-1G4'	NgH1'-8AH4'	NgH3'-1GH1'
NgH2"-14TMe	NgH9-1GH2"	NgH11-1GH5'	NgH1'-8AH5'	NgH3'-2CH1'
NgH5"-7CH1'	NgH9-1GH4'	NgH11-1GH5"	NgH2'(1)-8AH1'	NgH3'-2CH2"
NgH3-14TMe	NgH9-2CH1'	NgH11-13GH2'	NgH2'(2)-7CH1'	NgH3'-2CH3'
NgH7-8AH1'	NgH9-2CH3'	NgH11-14TH1'	NgH2'(2)-7CH6	NgH3'-2CH6
NgH7-8AH8	NgH10-1GH1	NgH11-14TH2'	NgH2'(2)-8AH1'	NgH3'-13GH1'
NgH8(1)-8AH1'	NgH10-7CH1'	NgH11-14TH3'	NgH2'(2)-8AH4'	NgH5'-7CH1'

4.14 Melting Temperature Studies of the Complex

Figure 4.21 shows the UV melting curve obtained for the complex formed between nogalamycin and 5′-d(GCGAAGCACGAAGT)-3′. The duplex to single strand transition occurred very rapidly over a few degrees, whereas a much broader transition was noted for the free oligonucleotide. The free oligonucleotide was found to melt in two stages, with the 3′-hairpin melting first followed by the 5′-hairpin melting shortly after. Upon drug binding, the two hairpins were stabilised and melted simultaneously as is shown by the sharp melting transition in figure 4.21.

The complex was apparently stabilised by the intercalation of nogalamycin with a T_m of 72.4 °C, relative to a T_m of 65.3 °C for the free oligonucleotide. The intercalation of nogalamycin, between the G-C and A-T base pairs at the nicked site, acts to remove the flexibility that was observed for the free oligonucleotide through strong base stacking interactions.

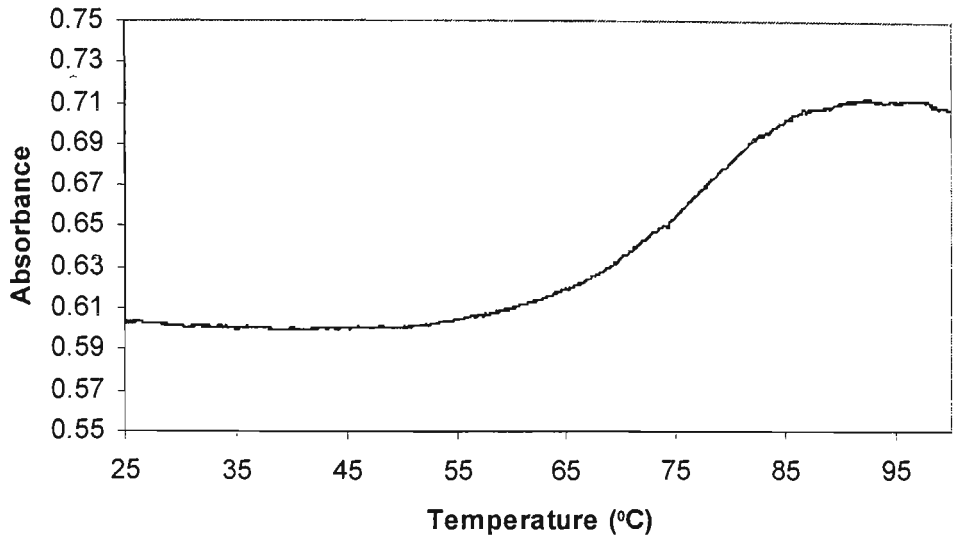


Figure 4.21: UV melting curve of a 1:1 mixture of nogalamycin (7 μM) and 5′-d(GCGAAGCACGAAGT)-3′ in 300 mM NaCl/30 mM NaH_2PO_4 (pH 7.0).

4.15 Imino Protons in the Complex

To examine further the stability of these novel structures, the samples were dissolved in water (10% D_2O) and the presence of the imino protons was detected through the use of a 1D-watergate experiment. The imino protons are only observed if they are protected from exchange through H-bonding interactions. Hence, their presence in the 1D spectrum is further evidence for the formation of the hairpin or loop structure. The A_8, G_{13}, G_1 and G_6 all form Watson-Crick base-pairs and as such their imino protons should be protected.

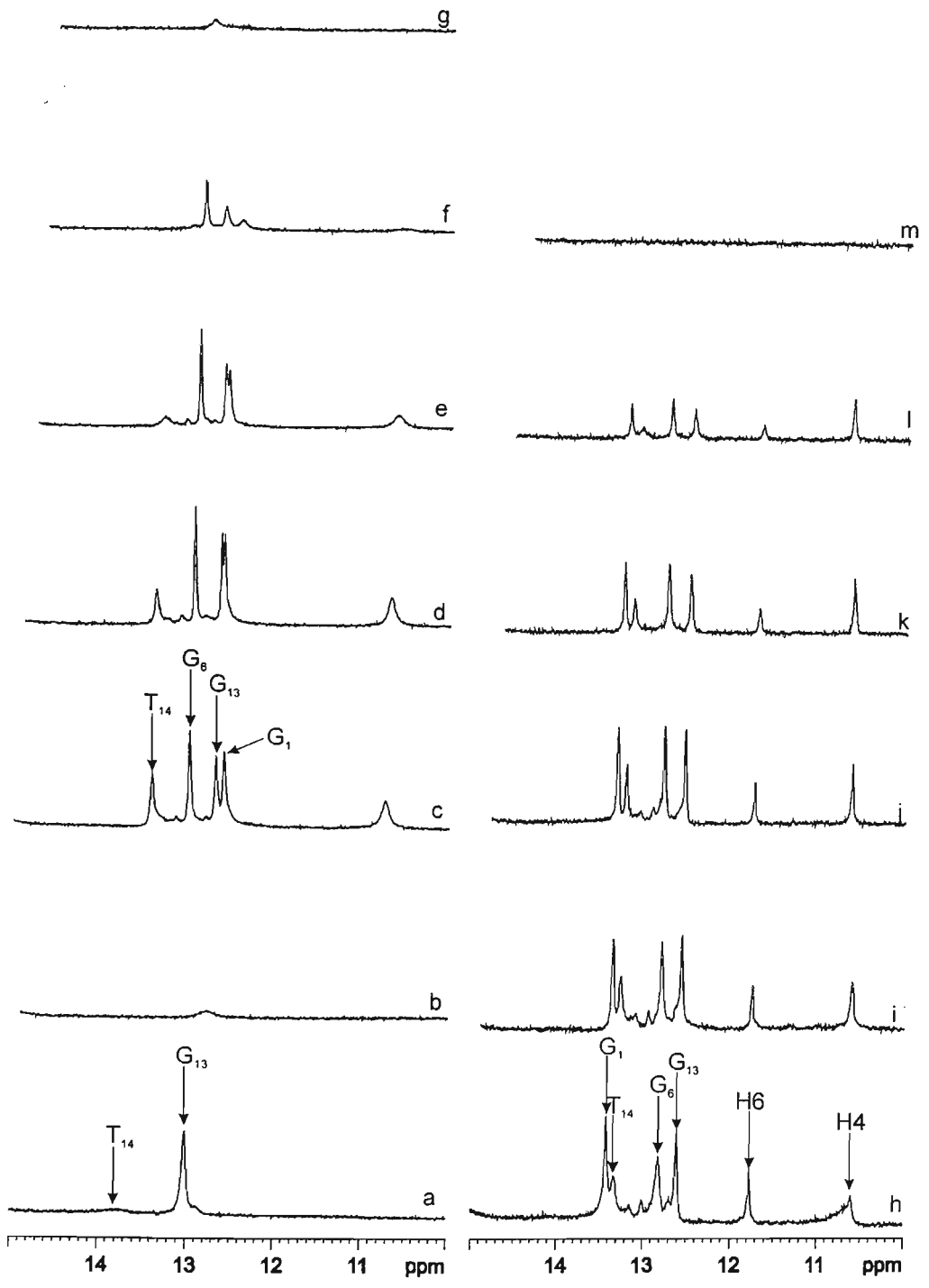


Figure 4.22: Stacked 1D plots of the 10-15 ppm range of the hairpin (a-b), loop sequence (c-g) and complex (h-m). The initial 1D spectra were recorded at 5 °C and the temperature was increased in 10 °C increments until the melting transition occurred as evidenced by the attenuation of the imino peaks.

Figure 4.22 is a section of the 1D-watergate experiment, which shows the presence of the imino protons. The experiments were carried out at 5 °C (278 K) for the hairpin, loop sequence and nogalamycin complex. The temperature was increased in 10 °C increments until the peaks disappeared, *i.e.* until the structures melted. Figure 4.22 (a-b) shows that the hairpin, 5′-d(ACGAAGT)-3′, was stable at 5 °C, with two peaks corresponding to the G_{13} and T_{14} imino protons present. At 15 °C, only a low intensity, broad peak was observed owing to the G_{13} imino proton. The thymine imino peak is noted to be of greater peak width owing to the lesser extent of H-bonding between adenine and thymine (two hydrogen bonds) relative to guanine and cytosine (three bonds) and as it was terminal and therefore less protected from hydrogen-deuterium exchange. These data agree with the low value of T_m (48.5 °C) observed for the hairpin as well as the broad melting transition.

The free loop sequence (figure 4.22 c-g), in comparison, was stable at 25 °C, although the T_{14} imino proton resonance was beginning to broaden. This peak was not observed at 35 °C and the G_{13} imino peak underwent significant broadening as the 3′-end of the loop sequence melted. The 5′-hairpin underwent a melting transition between 35 and 45 °C. These data further explain the broad melting transition observed from the UV melting studies. As the 3′-end of the loop was observed to be stable to ca. 25 °C, compared to 5 °C for the hairpin, the data suggest that base stacking interactions between the two hairpins are responsible for the additional stability. A broad peak is also observed at ca. 10.7 ppm and is thought to be from the imino protons of the non-Watson-Crick base pairs (G-A wobble pairs) formed in the loop region. The four Watson-Crick base pairs were assigned by utilising the NOE connectivities from the 2D Watergate NOESY experiment (figure 4.23), in which the imino

peaks showed NOEs to both the amino peaks and the aromatic base protons within the base pair as well as NOEs between the protons of neighbouring base pairs. Imino-amino NOEs were observed for G_1NH-G_1NH_2 and G_6NH-G_6NH_2. Amino-amino NOEs were observed for G_1, G_6 and A_8, *i.e.* G_1NH_2-G_1NH_2, G_6NH_2-G_6NH_2 and A_8NH_2-A_8NH_2. Both of the amino protons of guanine show NOE crosspeaks to the H5 protons of the cytosine involved in the G-C base pair, which allow the unambiguous identification of the amino protons and hence the imino protons. NOEs were observed between G_1NH_2-C_7H5 and also between G_6NH_2- C_2H5 as shown in figure 4.23.

The 1D spectra for the complex formed with nogalamycin are shown in figure 4.22 (h-m) and are characterised by very intense peaks for all four imino protons involved in Watson-Crick base pairing interactions. Examination of the temperature dependence of imino proton line widths showed that the base pairs forming the intercalation site are highly stabilised at 15 °C, optimising van der Waals interactions between the ligand and the oligonucleotide. These are observed in all spectra up to 45 °C, where they begin to broaden as the complex melts. In contrast to the free oligonucleotide, there appears to be a single melting process with all NHs showing the same temperature dependent line broadening, with a sudden co-operative dissociation of the complex at high temperature. Hence, both hairpins unfold simultaneously, and as such it is seen that the binding of Ng further stabilises the hairpin structures via base stacking interactions. The imino proton of G_1 is also noted to have shifted down-field by 0.8 ppm relative to the free oligonucleotide. The phenolic protons of nogalamycin also appear in the spectra at approximately 10.6 and 11.8 ppm, corresponding to H4 and H6 respectively.

Figure 4.23 shows a section of the wg-NOESY spectrum in which the amino protons have been traced across from the amino-imino crosspeaks (not shown). There are two amino protons per adenine or guanine and hence, strong NOEs are observed between the amino protons of individual bases. The amino protons also show NOE crosspeaks to the H5 protons of the cytosine involved in the base pair. These connectivities are illustrated in figure 4.23. The blue lines show the connectivity observed for the C_2-G_6 base pair, while the pink lines represent the G_1-C_7 base pair. The A_8-T_{14} base pair was identified through the examination of the NOEs seen for the amino protons of A_8 to the A_8H_2 protons. The C_9-G_{13} base pair also showed NOE crosspeaks between the $G_{13}NH$-A_8NH_2 (green lines).

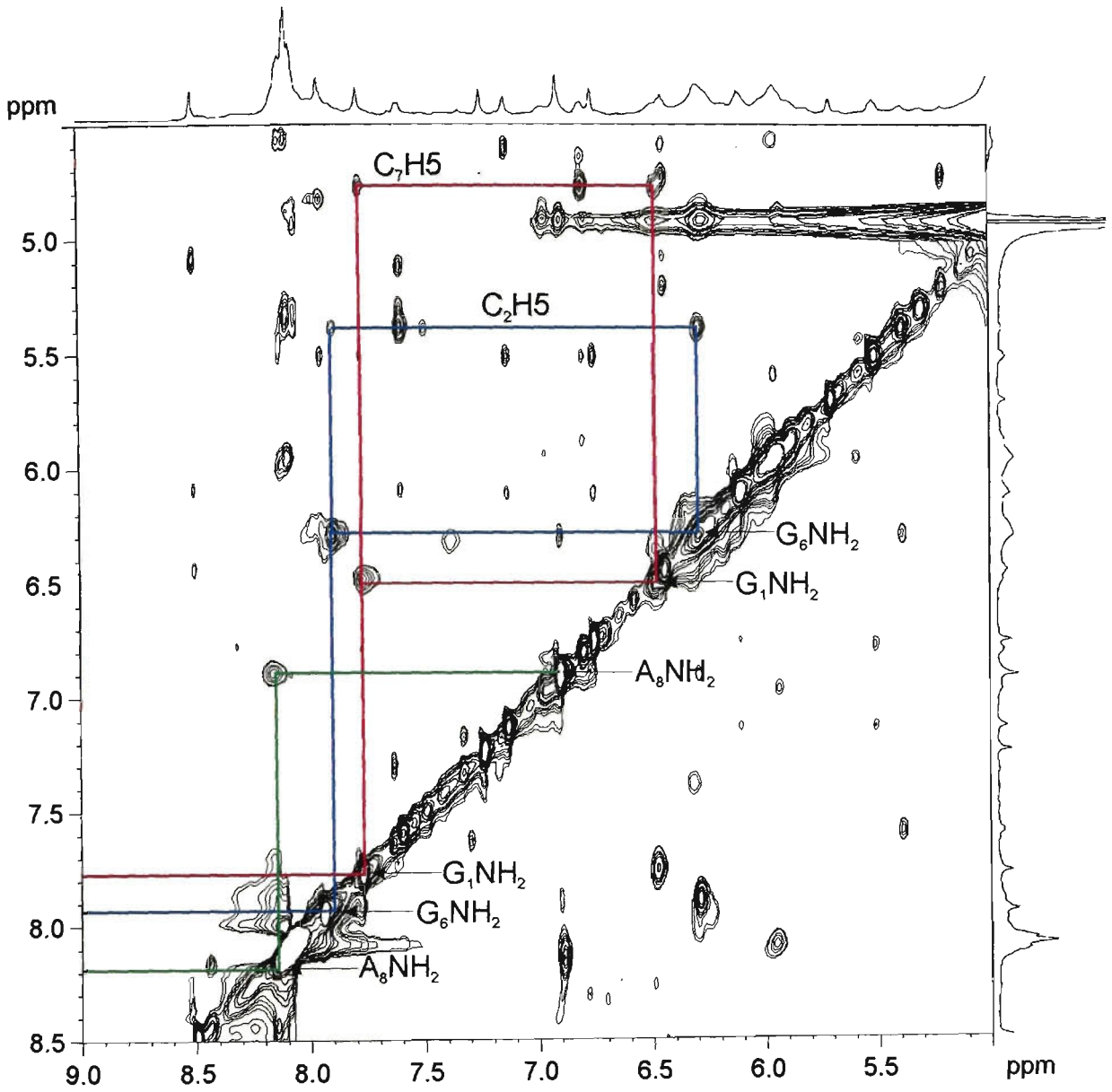


Figure 4.23: Section of the 300 ms wg-NOESY of 5′-d(GCGAAGCACGAAGT)-3′ (3 mM) in H_2O containing 0.1 M NaCl/100 mM NaH_2PO_4 (pH 7.0) showing connectivity of amino protons to amino protons and amino protons to base sugar protons.

4.16 ESI-MS Analysis

Electrospray ionisation mass spectrometry (ESI-MS) has been applied to the study of a number of non-covalent complexes of drug molecules with double-stranded DNA. (Beck *et al.*, 2001; Mazerska *et al.*, 2001; Gale *et al.*, 1995) In earlier studies using gentle ionisation conditions for the detection of the non-covalent complexes, we examined the binding of nogalamycin to a range of oligonucleotides forming duplex DNA. (Kapur *et al.*, 1999) Recently, the binding of the neocarzinostatin chromophore to unusual DNA structures including bulged DNA and RNA/DNA hybrids has been explored by ESI-MS. (Gao *et al.*, 1996; Zheng *et al.*, 1998)

The free oligonucleotide and the complex formed between nogalamycin and the oligonucleotide 5′-d(GCGAAGCACGAAGT)-3′ were both analysed by ESI-MS. NMR solutions initially contained salt concentrations of 100 mM NaCl/10 mM NaH_2PO_4 in order to stabilise these non-standard conformations. Previously, samples containing high salt concentrations have not been amenable to analysis by ESI-MS since the presence of sodium both reduces ionisation efficiency and limits mass accuracy owing to the presence of multiple species with different numbers of sodium ions bound. (Greig & Griffey, 1995) The recent use of newly introduced ESI mass spectrometers with the ion source in the Z-spray configuration, designed to minimise contamination, (Morris *et al.*, 1996) allowed the analysis of protein-DNA complexes in solutions containing high salt concentrations. (Kapur *et al.*, 2001) Hence in this work the possibility of using solutions directly from NMR without further purification or desalting was explored. The samples from NMR

studies were diluted using 100 mM ammonium acetate (pH 7.0) to give a final concentration of the complex of 10 pmol/μL and 5 mM NaCl/0.5 mM NaH_2PO_4. An electrospray probe tip potential of ~2 kV and a source temperature of 40 °C were used. Electrospray analysis was performed in negative ion mode using a Micromass Q-TOF 2. Other conditions were similar to those described recently for protein-DNA complexes. (Kapur *et al.*, 2001)

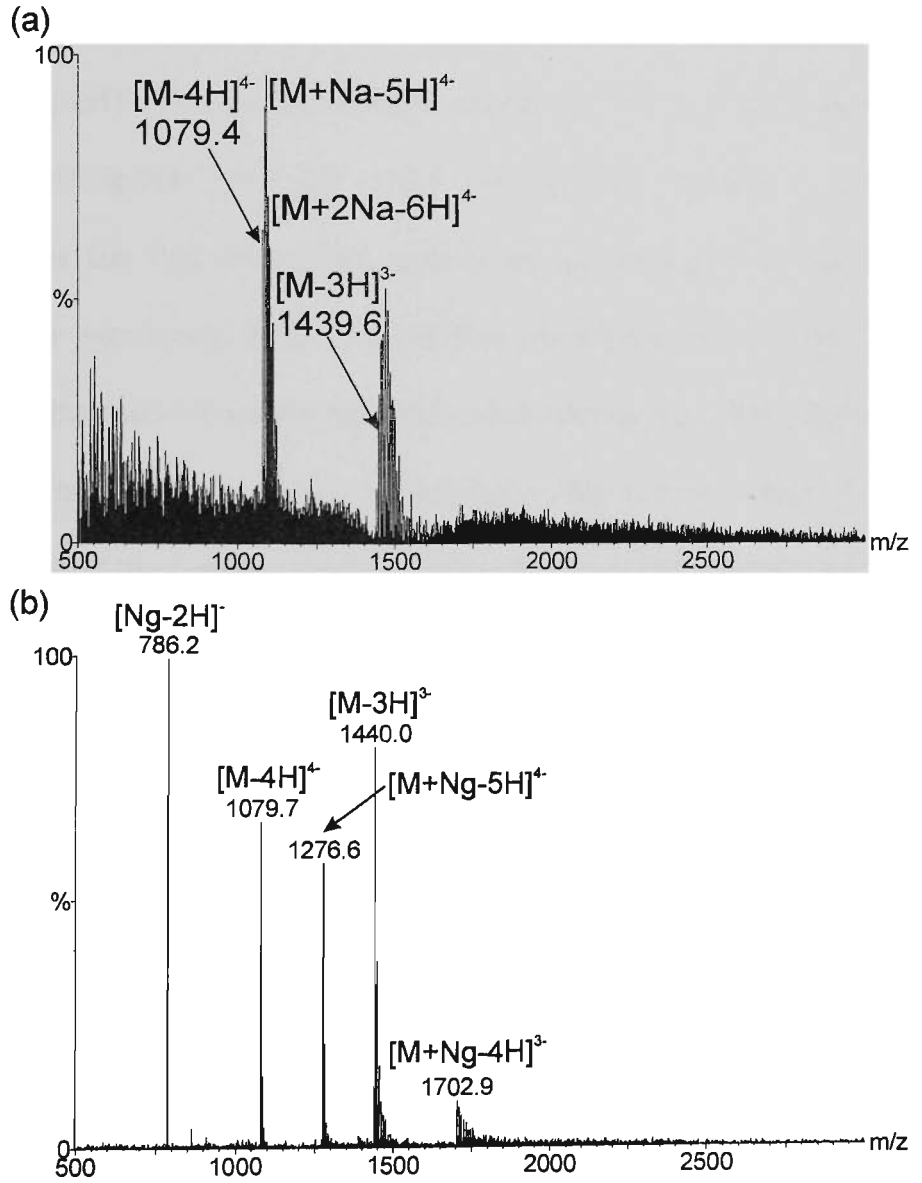


Figure 4.24: ESI-MS spectra of (a) the free oligonucleotide 5′-d(GCGAACGACGAAGT)-3′ and (b) the 5′-d(GCGAACGACGAAGT)-3′-Ng complex

The ESI mass spectrum of the unbound oligonucleotide (5′-d(GCGAACGACGAAGT)-3′) is shown in figure 4.24(a). The peaks at *m/z* 1079.4 and 1439.6 represent the $[M-4H]^{4-}$ and $[M-3H]^{3-}$ ions of the free oligonucleotide. There were a large number of peaks owing to multiple sodium adducts making interpretation of the spectrum difficult. In contrast, in the ESI mass spectrum of the complex formed with nogalamycin (Ng-5′-d(GCGAACGACGAAGT)-3′) (figure 4.24b), peaks from sodium adducts were fewer in number and of lower intensity. The peak at *m/z* 786.2 is owing to the free form of the ligand $[Ng-2H]^{-}$. The spectrum shows abundant ions from the drug-DNA complex at *m/z* 1276.6, $[M+Ng-5H]^{4-}$, and *m/z* 1702.9, $[M+Ng-4H]^{3-}$, confirming a binding stoichiometry of 1:1. The fact that the sodium adducts are suppressed when the complex is formed is particularly interesting. It is possible that the stabilisation of the complex by the drug precludes the need for stabilising alkali metal counter ions along the polyanionic backbone. We are currently exploring this phenomenon with a wider range of DNA structures. The observation of the complex by ESI-MS in itself was somewhat unexpected as intercalation at the site of the single strand break was expected to introduce additional flexibility and reduced binding affinity, however, the conditions were sufficiently gentle to allow the detection of this relatively weak interaction. The presence of unbound oligonucleotide suggests that, even under the gentle ionisation conditions employed here, some dissociation of the complex occurred during ionisation/detection.

4.16 Conclusions

NMR and UV spectroscopy were utilised for the complete characterisation of the interaction between nogalamycin and this unusual oligonucleotide, which contained a break in the DNA backbone. The hairpin 5′-d(ACGAAGT)-3′ was first examined and the resonances assigned by NMR spectroscopy. These assignments facilitated the complete assignment of the double hairpin (loop sequence) 5′-d(GCGAAGCACGAAGT)-3′. The NOESY, TOCSY and DQF-COSY spectra, as well as the 1D-Watergate spectra, provided evidence of the formation of the loop and stem regions. Anomalous chemical shift values for the sugar H4', H5' and H5" protons were characteristic of the adenine involved in the loop. The presence of imino protons in the 1D NMR spectra measured in water provided evidence for the formation of the double stranded stem as these were only observed if they were protected from exchange through hydrogen bonding interactions between base pairs.

The reaction between nogalamycin and 5′-d(GCGAAGCACGAAGT)-3′ was monitored by 1D NMR experiments. The appearance of new peaks in the spectra owing to the complexed form and the attenuation of peaks owing to the free oligonucleotide were evidence of the formation of a 1:1 complex. Nogalamycin was found to bind at the TpG step with a large number of NOEs observed between the ligand and the T_{14} and G_1 bases in particular. Upon ligand binding, changes in chemical shift values for the DNA resonances were observed, with the largest differences noted for the nucleotides in the intercalation site, *i.e.* T_{14}, G_1, C_7 and A_8.

The 3′-hairpin, 5′-d(ACGAAGT)-3′, was observed to be far less stable than that of the hairpin studied previously, 5′-d(GCGAAGC)-3′ with T_m values of 48.5 °C and 76 °C respectively. (Yoshizawa *et al.*, 1997) The loop sequence was found to have a T_m of 65.3 °C. It is proposed that the co-axial stacking of the hairpins at the nicked site provides additional stability to the overall structure. The width of the melting curve, however, suggested that there was an overlap of the two transitions of the individual hairpins. Thus, the hairpins appear to behave independently with flexibility at the site of the nick. Nogalamycin had a stabilising effect on this unusual DNA structure with base stacking interactions between the aglycone chromophore and the base pairs accounting for an elevated melting temperature for the complex. Where previously the free loop sequence had been observed to undergo two melting events, in which the individual hairpins melted separately, the complex was observed to melt in one event with simultaneous attenuation of all peaks in the 1D spectra. The intercalation of nogalamycin lead to a significant sharpening of the melting curve.

Electrospray ionisation mass spectrometric analysis was undertaken using very gentle ionisation conditions and the intact Ng-5′-d(GCGAAGCACGAAGT)-3′ complex was detected. The applicability of mass spectrometry to the detection of non-covalent complexes direct from NMR solutions was demonstrated.

Bibliography

Altona, C. and Sundaralingam, M. (1972) *J. Am. Chem. Soc.* **94**, 8205-8212

Altona, C. and Sundaralingam, M. (1973) *J. Am. Chem. Soc.* **95**, 2333-2344

Beck, J.L.; Colgrave, M.L.; Ralph, S.R. and M. M. Sheil (2001) *Mass Spec. Rev.* **20(2)**, 61-87

Fox, K.R. and Waring, M.J. (1984) *Biochim. Biophys. Acta* **802**, 162-168

Gale, D.C. and Sondhi, S.M. (1995) *J. Am. Soc. Mass Spectrom.* **6**, 1154

Gao, Q.; Cheng, X.; Smith, R.D.; Yang, C.F. and Goldberg, I.H. (1996) *J. Mass Spectrom.* **31**, 31-36

Gao, X. and Patel, D.J. (1988) *J. Am. Chem. Soc.* **110**, 5178-5182

Greig, M. and Griffey, R.H. (1995) *Rapid Commun. Mass Spectrom.* **9**, 97-102

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

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

Hosur, R. V.; Ravikumar, M.; Chary, K. V.; Sheth, A.; Govil, G.; Zu-Kun, T. and Miles, H. T. (1985) *FEBS Letters* **205**, 71-76

Kan, L.S.; Chandrasegaran, S.; Pulford, S.M. and Miller, P.S. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 4263-4265

Kapur, A.; Beck, J.L. and Sheil, M.M. (1999) *Rapid Commun. Mass Spectrom.* **13**, 2489-2497

Kapur, A.; Beck, J.L.; Brown, S.E.; Dixon, N.E. and Sheil, M.M. (2002) *Protein Science* **11(1)**, 147-157

Mazerska, Z.; Dziegielewski, J. and Konopa, J. (2001) *Biochem. Pharmacol.* **61**, 685-694

Morris, H.R.; Paxton, T.; Dell, A.; Langhorne, J.; Berg, M.; Bordoli, R.S. Hoyes, J. and Bateman, R.H. (1996) *Rapid Commun. Mass Spectrom.* **10**, 889-896

Patel, D.J.; Kozlowski, S.A.; Ikuta, S. and Itakura, K. (1984) *Biochemistry* **23**, 3207-3217

Prive, G.G.; Heinemann, U.; Chandrasegaran, S.; Kan, L.S.; Kopka, M.L. and Dickerson, R.E. (1987) *Science (Washington D.C.)* **238**, 498-504

Searle, M.S.; Hall, J.G.; Denny, W.A. and Wakelin, L.P. (1988) *Biochemistry* **27**, 4340-4349

Searle, M.S. and Wakelin, L.P. (1990) *Biochem. J.* **269**, 341-346

Williams, H.E.L. and Searle, M.S. (1999) *J. Mol. Biol.* **290**, 699-716

Yoshizawa, S.; Kawai, G.; Watanabe, K.; Miura, K. and Hirao, I. (1997) *Biochemistry* **36**, 4761-4767

Zheng, P.; Liu, C.L.; Xi, Z.; Smith, R.D. and Goldberg, I. H. (1998) *Biochemistry* **37**, 1706-1713

CHAPTER FIVE: *Bulge Sequences, a variation on the Hairpin*

5.0 Introduction

Bulge structures in DNA have been shown to act as potential targets for therapeutic drugs. They are simple and common, non-helical, features of RNA and DNA and may vary from single nucleotide bulges to multiple base bulges. In DNA, bulges play a role in frame shift mutagenesis in sequences with repeating base pairs. (Okada *et al.*, 1972) A more detailed description of the structural aspects of bulged sequences may be found in section 1.4. There have been relatively few detailed reports on the structural characterisation of the interactions of bulged DNA with ligands.

The importance of bulge sites as recognition elements has been established by a number of studies. The intercalator, ethidium was found to bind selectively to cytosine bulges in preference to duplex DNA by an order of magnitude greater in strength. (Nelson & Tinoco, 1985; White & Draper, 1987) Another example of the heightened affinity of drugs for bulge sites in DNA is 9-aminoacridine, which preferentially binds DNA molecules that contain an extrahelical base. (Woodson & Crothers, 1988) The cleavage of duplex DNA by the neocarzinostatin chromophore, at the 3′-side of cytosine and thymine bulges, has been investigated extensively. (Kappen & Goldberg, 1993a; Kappen & Goldberg, 1993b; Stassinopolous & Goldberg, 1995; Kappen & Goldberg, 1997; Xi *et al.*, 1999; Gu *et al.* 2000) The neocarzinostatin chromophore has been demonstrated to have up to 20 times higher selectivity for bulged adenine sites over control sites. (Gu *et al.*, 2000) Saito and co-workers recently showed that an altromycin analogue selectively alkylates guanines

opposite a bulge. (Nakatani *et al.*, 1999) The syntheses of drugs designed to bind specifically to duplex DNA containing a single guanine bulge was carried out by Nakatani *et al.* (Nakatani *et al.*, 2000) These naphthyridine derivatives acted as recognition molecules that differentiated between the types of bulged bases. Caceres-Cortes and Wang used nogalamycin as a probe for the binding of intercalators to bulged sites in DNA, but their results appear to be strongly influenced by end-effects (see discussion below).

Unpaired bases arise either as a result of recombination between strands that are not completely homologous, or from errors during replication. Duplex DNA with bulged bases present have been proposed as intermediates in frame-shift mutagenesis, (Streisinger *et al.*, 1966) as well as having implications in site-specific protein recognition. (Wu & Uhlenbeck, 1987) The examination of the structure and stability of these bulged sequences may aid the understanding of their biological roles. DNA repair mechanisms are governed by the conformation and stability of the unpaired base at the helix interruption site in duplex DNA. (Hare *et al.*, 1986) Unpaired, or bulged sites, in DNA are repaired by specific repair enzymes, which act to repair errors made during DNA replication or as a result of some types of chemical damage to DNA. These enzymes have also been implicated in base excision repair in the repair of physical and/or chemical damage to DNA and as a part of a cell-cycle check-point control system in which they recognise certain types of DNA damage and trigger cell-cycle arrest. (Kolodner, 1995) Mismatch repair is crucial in biochemical processing as shown by its role in the cancer repair systems, whereby defects in components of mismatch repair have been implicated in hereditary and sporadic tumours. (Kolodner, 1996) MutS proteins carry out the recognition of mispaired or unpaired bases during mismatch repair. (Malkov *et al.*, 1997)

5.1 The Bulge Sequences

Two hairpin oligonucleotides incorporating a bulged thymine (T_b) within the stem region were designed to investigate further the binding of nogalamycin to DNA with distorted backbones. Nogalamycin shows sequence selectivity to 5′-TpG (CpA) sites. The hairpin oligonucleotides were designed to incorporate this high affinity site. The oligonucleotides, 5′-d(GTGCGAAGCT$_b$AC)-3′ and 5′-d(GCT$_b$ACGAAGTGC)-3′, were purified as described previously (section 4.2) and analysed by NMR prior to titration with nogalamycin and the subsequent assignment of the resulting complex.

The possible modes of folding for these sequences are shown in figure 5.1. Nogalamycin (Ng) requires a TpG step as described earlier. It is thought that the G-C base pair is also required for Ng recognition. To satisfy this requirement, upon Ng binding to the oligonucleotide sequence, 5′-d(GCTACGAAGTGC)-3′, the thymine at T_3 must adopt a bulged conformation as opposed to a frame-shift event to maintain a G-C base pair at the intercalation site (figure 5.1b). The sequence 5′-d(GTGCGAAGCTAC)-3′ may form a thymine bulge at T_{10}. Alternatively, frame shift may occur resulting in the formation of T-T and G-A "wobble pairs" (figure 5.1a). Hence the cross in figure 5.1 is indicative of an unfavourable frame shift event for 5′-d(GCTACGAAGTGC)-3′. In this work, the question of the relative stability of the bulged hairpin structures is addressed. Further, we have examined the way in which the Ng sequence specificity influences the propensity for bulged or frame shift events to occur upon binding of the ligand.

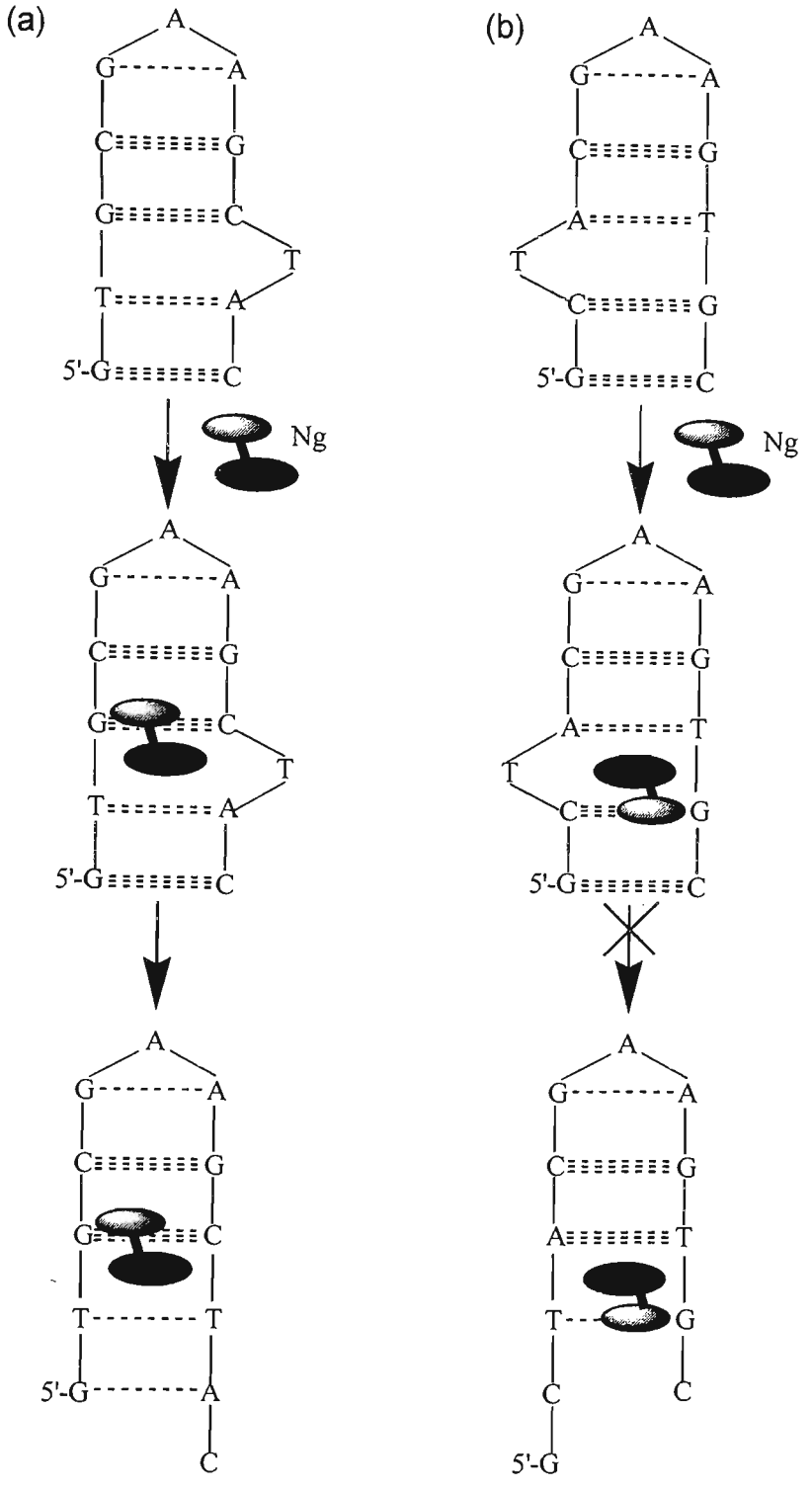


Figure 5.1: Possible modes of folding for 5′-d(GTGCGAAGCTAC)-3′ and 5′-d(GCTACGAAGTGC)-3′ upon binding to nogalamycin.

5.2 Evidence for Loop Formation

The investigations of these unusual structures were undertaken by first characterising the free oligonucleotides followed by the examination of the complexes formed upon reaction with nogalamycin. Figure 5.2 shows the H4'-H2'/H2" region of the 300 ms mixing time NOESY spectrum of 5'-d(GCTACGAAGTGC)-3'. The NOEs between the H2'-H2", H4'-H2' and H4'-H2" resonances are indicated. The loops were determined to be formed through the observation of the up-field shifted H4' and H5'/H5" sugar resonances as shown in figure 5.2. For the bulge sequence 5'-d(GTGCGAAGCTAC)-3', the A_6 H4' and H5'/H5" protons (loop region) were observed at 2.094 and 2.299 ppm whilst the A_{11} H4' and H5'/H5" protons (stem region) were observed at 4.163 and 4.073 ppm respectively. This shows a difference in chemical shift between the stem adenine and loop adenine H4' and H5'/H5" sugar protons of greater than 2 ppm. The H5' and H5" sugar proton resonances were overlapped for both and hence an average value is given. The NMR spectrum of bulge sequence 5'-d(GCTACGAAGTGC)-3' (data not shown) showed similar trends with the A_7 sugar H4', H5' and H5" protons observed at 2.116, 2.241 and 2.361 ppm respectively. These were again greater than 2 ppm up-field of the stem region A_4 sugar H4', H5' and H5" protons at 4.414, 4.024 and 4.122 ppm each. The A_7 sugar H2' and H2" protons were also noted to be down-field shifted by 0.35 and 0.56 ppm respectively for the adenines involved in the loop with the largest shift observed for A_6 for 5'-d(GTGCGAAGCTAC)-3' and A_7 for 5'-d(GCTACGAAGTGC)-3'.

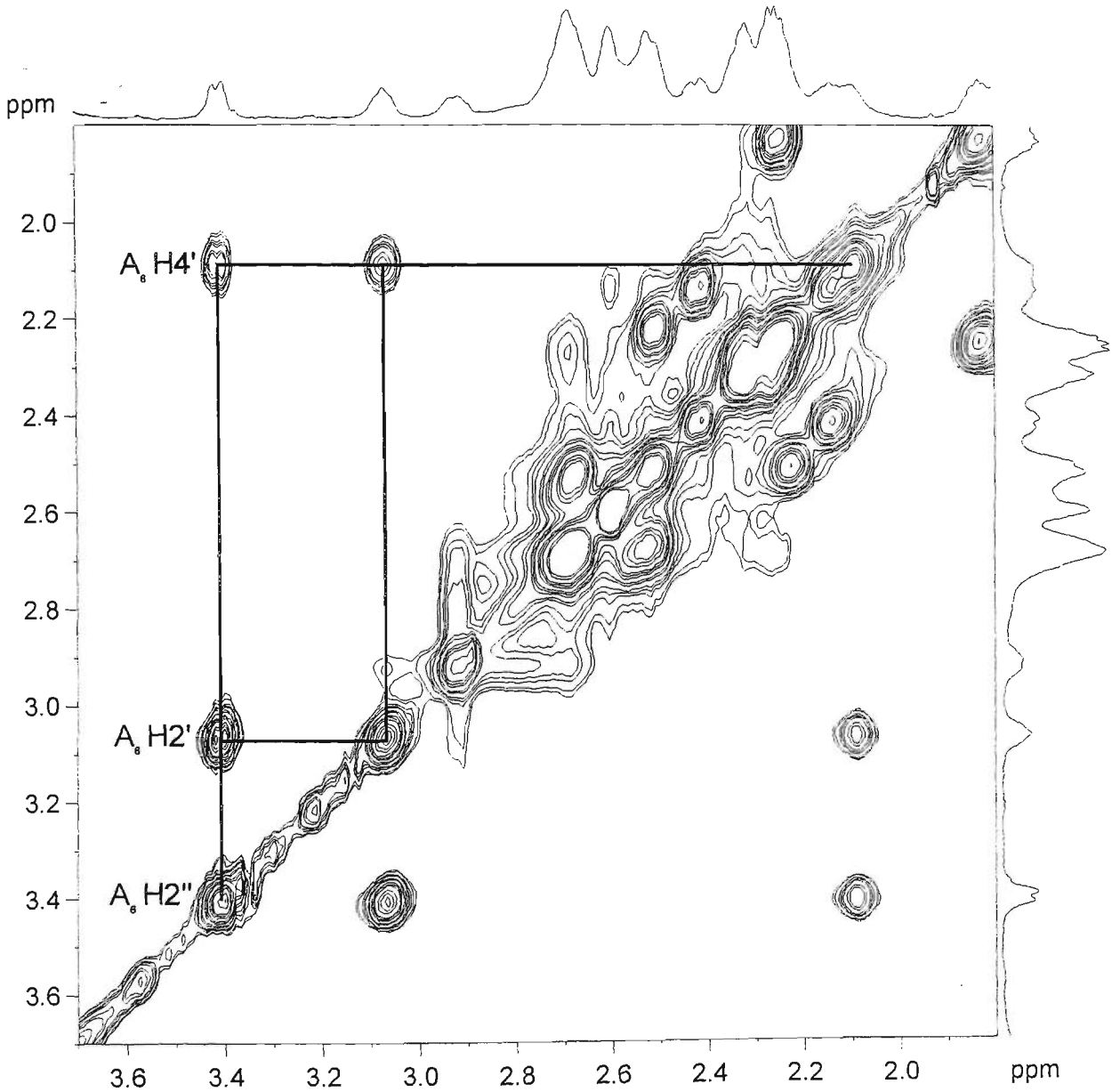


Figure 5.2: Section of the 300 ms NOESY spectrum showing the H4'-H2'/H2" resonances of A_6 in the sequence 5'-d(GTGCGA_6AGCTAC)-3' (3 mM) in D_2O containing 0.1 M NaCl/100 mM NaH_2PO_4 (pH 7.0).

5.3 Evidence for Stem Formation

Figure 5.3 shows the 1D NMR spectra recorded when each of the oligonucleotides were dissolved in water (10% D_2O), together with the corresponding NMR spectra of the complexes with nogalamycin. Figure 5.3(a) shows two sharp peaks corresponding to the

imino protons of the G_3 and G_8 bases (13.128 and 13.060 ppm respectively) for 5′-d(GTGCGAAGCTAC)-3′. One sharp and one broader signal were observed at 12.784 and 13.860 ppm corresponding to the imino protons of G_9 and T_{10} for the sequence 5′-d(GCTACGAAGTGC)-3′ (figure 5.3c). The identities of these imino protons were confirmed through the 2D-Watergate-NOESY experiment showing the connectivity between the imino protons and the aromatic base protons. Figure 5.3(b) and (d) show the 1D spectra of the complexed form of these oligonucleotides with Ng and will be discussed in more depth later.

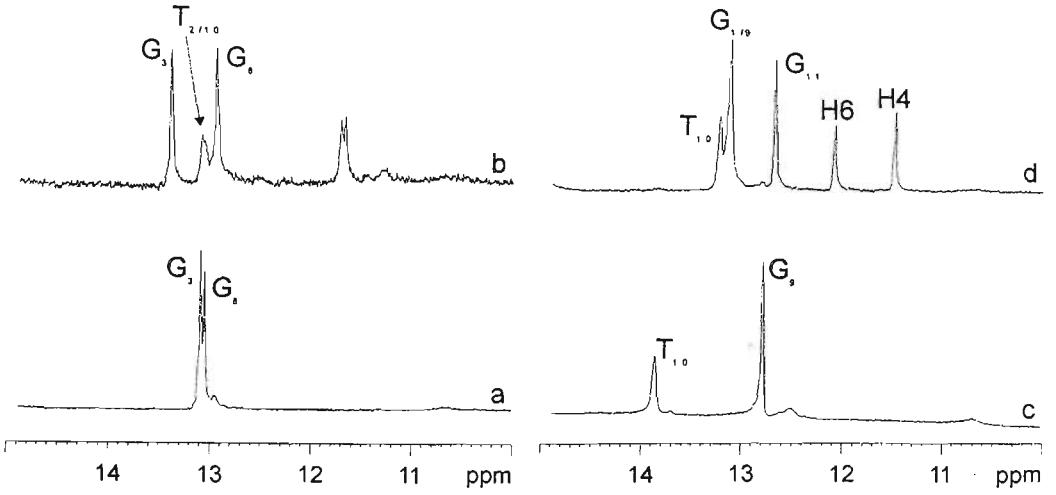


Figure 5.3: 1D spectra showing the imino protons of the bulge sequences: (a) 5′-d(GTGCGAAGCTAC)-3′ and (c) 5′-d(GCTACGAAGTGC)-3′ and the complexes with nogalamycin: (b) 5′-d(GTGCGAAGCTAC)-3′-Ng and (d) 5′-d(GCTACGAAGTGC)-3′-Ng recorded at 15 °C.

The G_3-C_9 and C_4-G_8 Watson-Crick base pairs of 5′-d(GTGCGAAGCTAC)-3′ are formed in the free oligonucleotide as evidenced by the presence of their imino peaks in the 1D spectra. The G_1 and T_2, T_{10} imino peaks are absent, however, suggesting that the terminal

section of the stem, *i.e.* after the mispaired thymine, is not formed. Hence, it was observed that the unpaired thymine has a destabilising effect on the terminal base pairs. It is interesting to note, however, that although the terminal base pairs are not formed, the individual strands still exist to a significant extent as B-form DNA as is evidenced by the sequential connectivity observed along the backbone. If the strands were adopting a random conformation, there would be no discernable pattern of backbone NOEs to follow, which is not the case as shown in figure 5.4. Many of the sequential NOEs are, however, quite weak suggesting some degree of conformational averaging.

Figure 5.4 shows the H8/H6-H2'/H2" and H8/H6-H1' sequential connectivity along the DNA backbone for (a) 5'-d(GTGCGAAGCTAC)-3' and (b) 5'-d(GCTACGAAGTGC)-3'. The solid lines show the observed NOEs and connectivity, while the dashed lines show the loss of sequential connectivity caused by the 5'-GAA loop. The squares indicate where NOEs between the nucleotides in the 5'-GAA loop would have been observed if the oligonucleotide was adopting a B-type conformation. A loss of sequential connectivity is also observed between C_4-G_5 in the H1' region of the NOESY spectrum for 5'-d(GTGCGAAGCTAC)-3'. Many NOEs are weak suggesting that the structure may be dynamic, *i.e.* very weak NOEs are observed between the C_9 and T_{10} and between T_{10} and A_{11}, presumably because of dynamic changes between a bulged T conformation as well as a frame-shifted conformation. In the NOESY spectrum of 5'-d(GCTACGAAGTGC)-3', (figure 5.4b) a complete loss of sequential connectivity is observed through the T_3 nucleotide, *i.e.* the unpaired thymine, (dashed lines) suggesting that this thymine is not stacking strongly with its adjacent bases. Again, as was the case for the oligonucleotide, 5'-

d(GTGCGAAGCTAC)-3′, loss of sequential connectivity through the 5′-GAA loop of 5′-d(GCTACGAAGTGC)-3′ as well as between C_5-G_6 is observed suggesting a *high-anti* conformation.

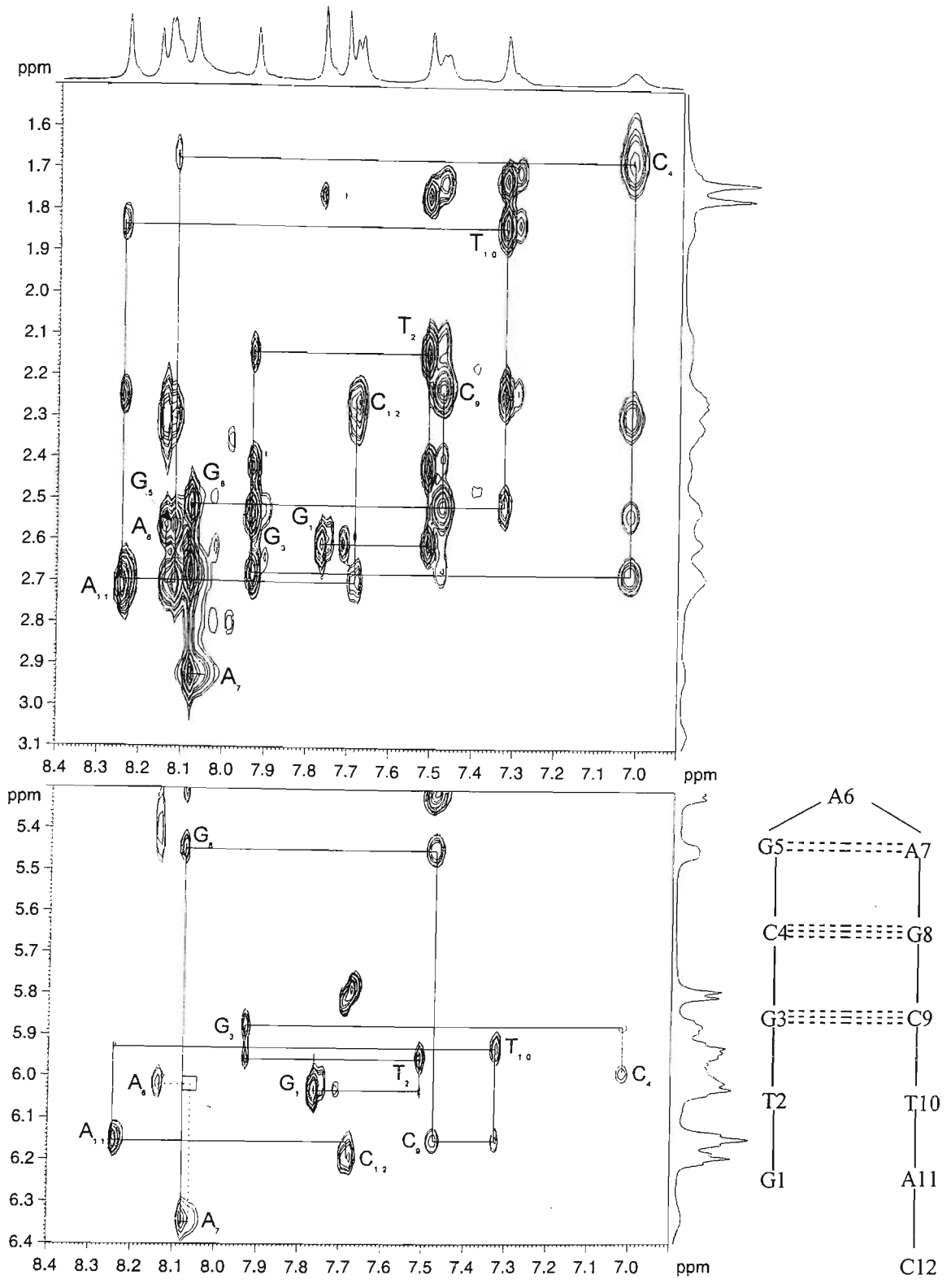


Figure 5.4(a): Section of the 300 ms NOESY spectrum showing sequential connectivity along the oligonucleotide chain for 5′-d(GTGCGAAGCTAC)-3′ (3 mM) in D_2O containing 0.1 M NaCl/100 mM NaH_2PO_4 (pH 7.0): H8/H6-H2'/H2" (top) and H8/H6-H1' (bottom).

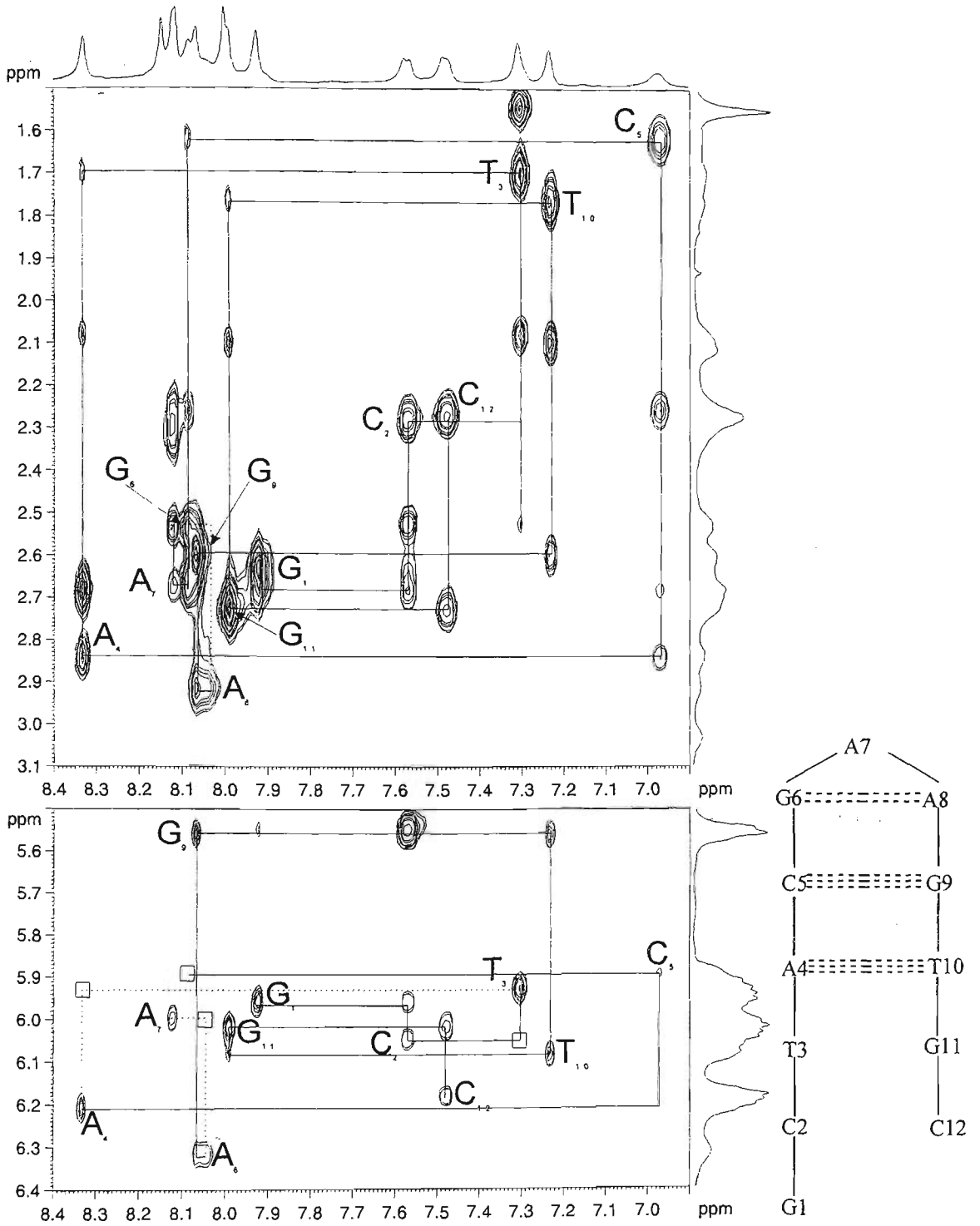


Figure 5.4(b): Section of the 300 ms NOESY spectrum showing sequential connectivity along the oligonucleotide chain for 5′-d(GCTACGAAGTGC)-3′ (3 mM) in D_2O containing 0.1 M NaCl/100 mM NaH_2PO_4 (pH 7.0): H8/H6-H2'/H2" (top) and H8/H6-H1' (bottom).

The C_5-G_9 and A_4-T_{10} Watson-Crick base pairs of 5′-d(GCTACGAAGTGC)-3′ are formed in the free oligonucleotide, however, there was no evidence for the formation of the C_2-G_{11} or G_1-C_{12} base pairs indicating fraying at the end of the chain as was the case for the previous oligonucleotide sequence (figure 5.5). Although, the absence of the imino protons is indicative of the fraying of the stem, the oligonucleotide still appears to be in the B-conformation as evidenced by internucleotide stacking along each strand. NOEs allowing the sequential connectivity along the DNA strand were observed.

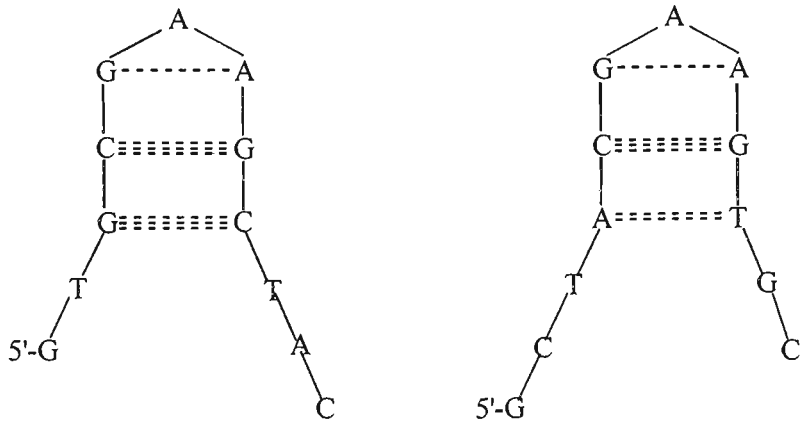


Figure 5.5: Schematic of the proposed structures of the bulge sequences showing fraying at end of stem.

In general, it has been noted that the inclusion of an unpaired base in the hairpin sequence had a significant destabilising effect on the duplex region. (Dodgson & Wells, 1977; Wallace *et al.*, 1979) The work presented here shows that the oligonucleotide 5′-d(GTGCGAAGCTAC)-3′ was in dynamic equilibrium between bulged and frame-shifted conformations. Destabilisation and end-fraying were observed for both sequences studied, however, the single strands remained in a B-type conformation.

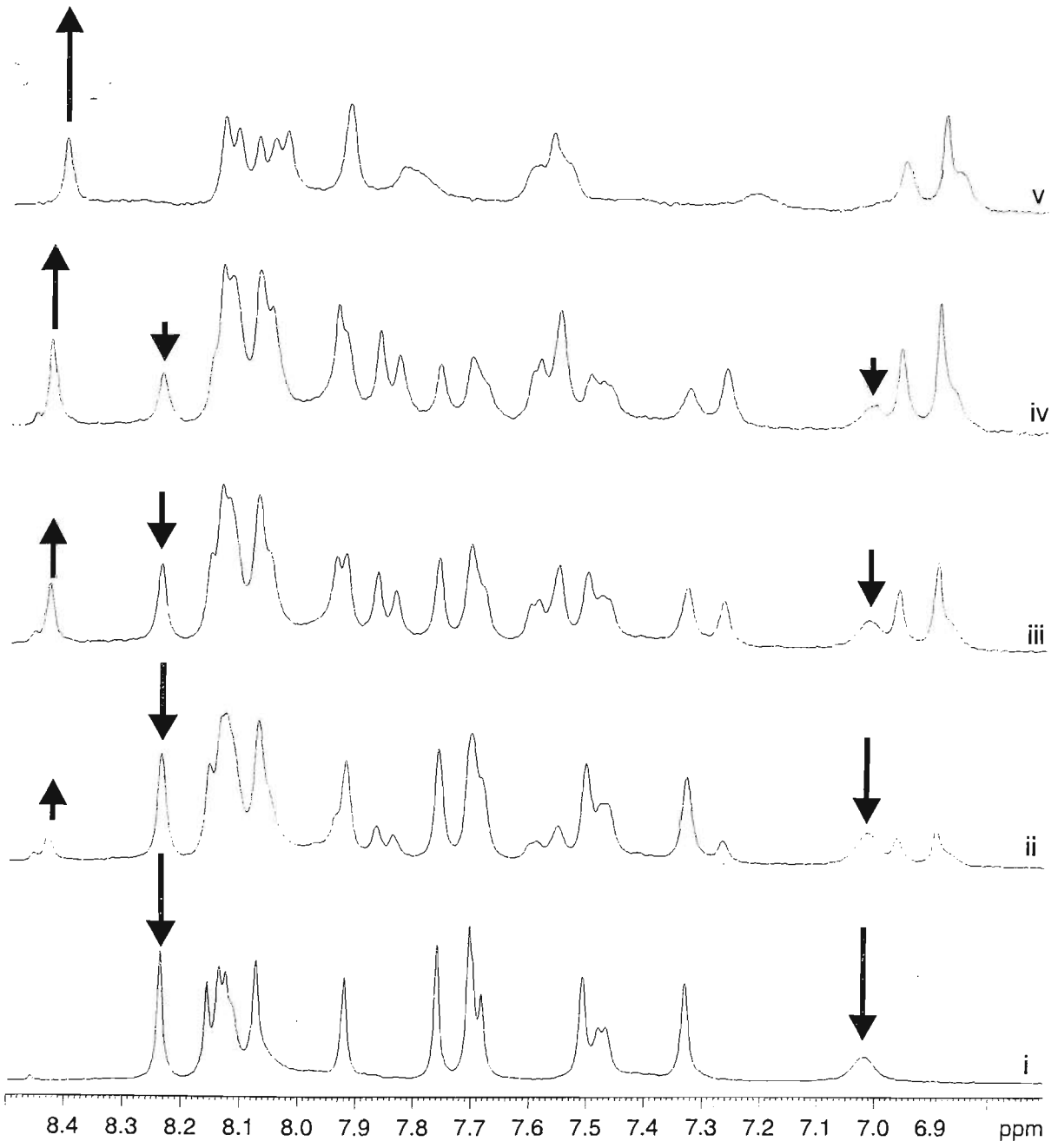


Figure 5.6: 1D NMR spectra acquired after the addition of 5 mM nogalamycin: (i) 0:1 (ii) 0.25:1 (iii) 0.5:1 (iv) 0.75:1 (v) 1:1 complex. The arrows highlight a number of clearly defined signals, owing to free DNA that decrease upon addition of the nogalamycin. Arrows also indicate new signals that increase owing to complex formation.

5.4 Complex Formation with Nogalamycin

The following section describes the results of structural studies of the complexes of nogalamycin with oligonucleotides that adopt the novel bulge conformations. The nomenclature scheme for nogalamycin is shown in figure 4.17. NMR spectroscopy was utilised for the elucidation of the intercalation site, orientation of the drug molecule within this site and overall structure of the complexes. Melting temperature studies were also undertaken to complement the results gained from the NMR studies.

5′-d(GTGCGAAGCTAC)-3′

Figure 5.6 shows 1D NMR spectra acquired after the addition of aliquots of nogalamycin to the free oligonucleotide, 5′-d(GTGCGAAGCTAC)-3′. These 1D-NMR titration studies showed that Ng formed a clean 1:1 complex with 5′-d(GTGCGAAGCTAC)-3′, with resonances from the free DNA replaced by those of a complex containing a single-bound nogalamycin. For example, the resonance at 8.24 ppm corresponding to the A_{11}H8 of the free oligonucleotide decreases in intensity upon addition of nogalamycin, whilst a resonance appears at 8.43 ppm corresponding to the A_{11}H8 of the bound species. The observed down-field shift results from nogalamycin intercalating between the base pairs 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. (Williams & Searle, 1999)

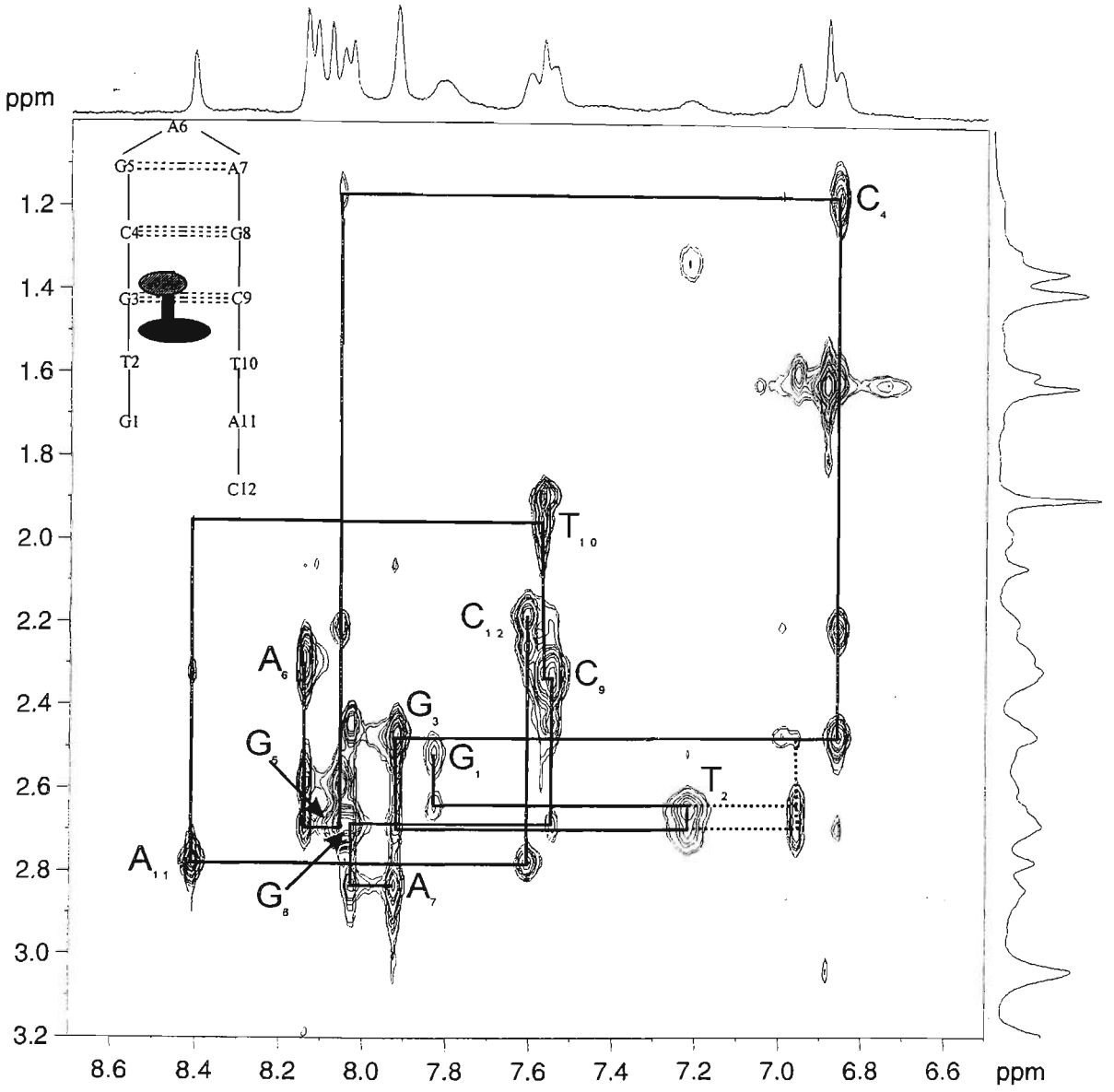


Figure 5.7: Section of the 300 ms NOESY spectrum of the 5′-d(GTGCGAAGCTAC)-3′-Ng complex (3 mM) in D_2O containing 0.1 M NaCl/100 mM NaH_2PO_4 (pH 7.0), showing H8/H6-H2'/H2" sequential connectivity along the DNA backbone.

Figure 5.7 shows a portion of the NOESY spectrum of the 5′-d(GTGCGAAGCTAC)-3′-Ng complex indicating the H8/H6-H2'/H2" sequential connectivity along the DNA backbone. The solid lines show the sequential connectivity along the DNA backbone including the loop region, i.e. $G_1T_2G_3C_4G_5A_6A_7G_8C_9$ and $A_{11}C_{12}$. A loss of sequential connectivity

(dashed lines) and weak NOEs observed between C_9 and T_{10}, give further weight to the suggestion that the oligonucleotide has a dynamic structure such that there is an equilibrium between different conformations. NOEs were also observed between the Ng-H11 and T_2H2'/H2" as shown by the dotted lines in figure 5.7. It is interesting to note that upon ligand binding, a number of resonances have been further shifted. For example, the A_{11}H8 resonance is down-field shifted by 0.192 ppm, while the T_2H6 has been shifted up-field by 0.284 ppm. This is typical of an adenine involved in a 5'-(CpA) intercalation site. (Williams & Searle, 1999)

Upon binding of the ligand to the oligonucleotides, the most pronounced changes in chemical shift occurred to the nucleotides within the intercalation site, *i.e.* T_2, G_3, C_9, T_{10} and A_{11}. The change in chemical shift upon binding for the aromatic base protons (H8/H6) and the sugar H1' protons are given in table 5.1.

Table 5.1: Change in chemical shift upon nogalamycin binding: (a) H8/H6 and (b) H1'.

(a) **H8/H6**

1G	2T	3G	4C	5G	
-0.059	0.284	0.013	0.163	0.066	**6A**
12C	**11A**	**9C**	**8G**	**7A**	0.003
0.078	-0.165	-0.067	0.051	0.124	
	10T				
	-0.242				

(B) **H1'**

1G	2T	3G	4C	5G	
0.025	0.050	0.321	0.102	---	**6A**
12C	**11A**	**9C**	**8G**	**7A**	0.031
0.062	-0.167	-0.175	0.092	0.026	
	10T				
	-0.249				

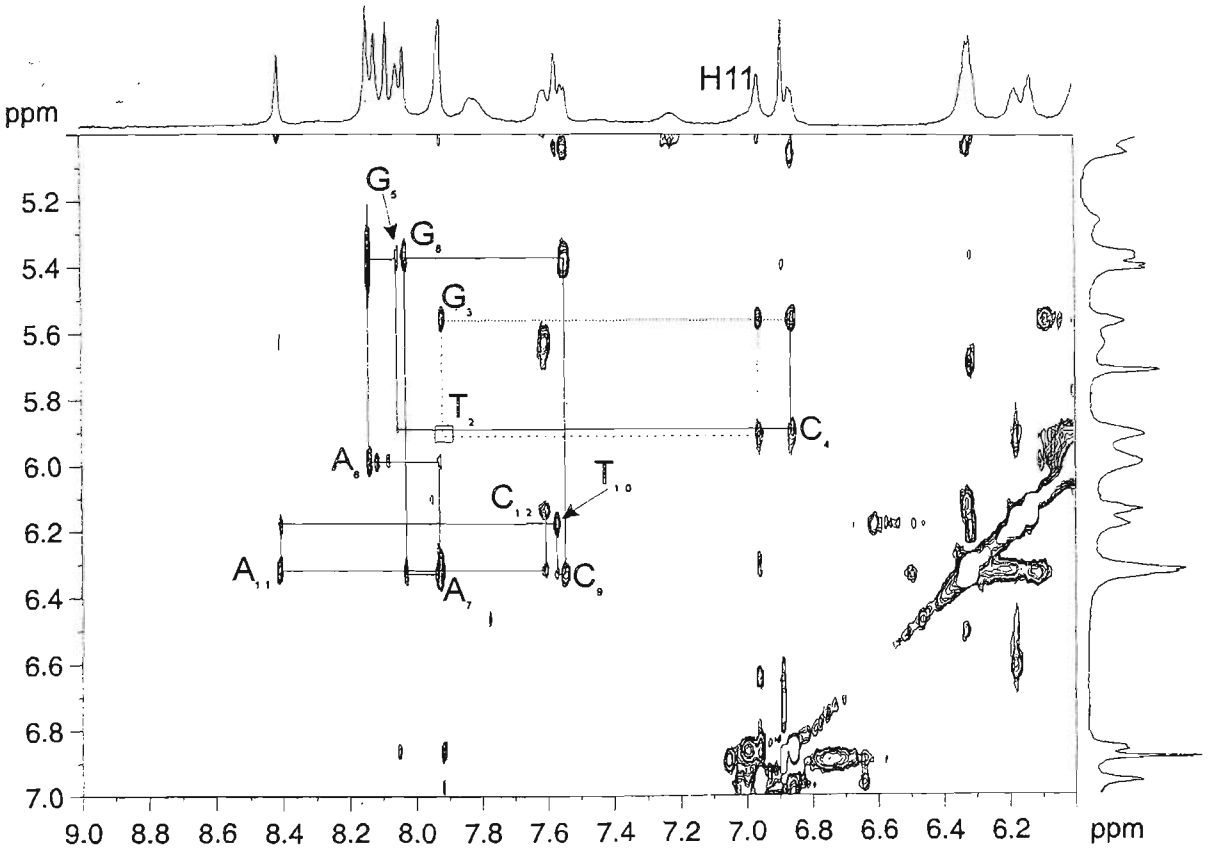


Figure 5.8: 300 ms NOESY spectrum of 5′-d(GTGCGAAGCTAC)-3′ complexed with nogalamycin (3 mM) in D_2O containing 0.1 M NaCl/100 mM NaH_2PO_4 (pH 7.0), showing the resonances that define the intercalation site. H8/H6-H1' region highlighting the drug-DNA NOEs.

Figure 5.8 highlights the NOEs that define the intercalation site. The sequential connectivity can be followed along the DNA backbone through the intercalation site with the ligand bridging the gap between the T_2 and G_3 resonances. Connectivity was observed for T_2 through to C_4. The connectivity, however, did not follow from T_2 to G_3, as was the case in the free oligonucleotide (dashed lines). Instead it stepped from T_2H1' to the H11 of nogalamycin to G_3H1' (dotted lines). These data clearly define the intercalation site to be the TpG step as was expected. The backbone could then be followed from C_4 to C_{12},

however, NOE crosspeaks between A_6, A_7 and G_8 were significantly weaker than in the stem region of the oligonucleotide. NOEs were also observed between C_9-T_{10}, C_9-A_{11} and T_{10}-A_{11}. A number of other drug-DNA NOEs were observed in the NOESY spectrum. These are given in table 5.2 and indicate that nogalamycin is oriented so that the nogalose sugar moiety points into the minor groove, whilst the aminoglucose sugar moiety points into the major groove. A large number of NOEs were observed between the nogalose sugar and the protons of the T_{10} and C_9 bases in DNA, *i.e.* NgH2'-T_{10}H2"/H4'/H5', NgH2'-C_9H1'/H6 and NgH3'-G_3H5', whilst only a single NOE was observed for the bicycloaminoglucose sugar (NgH1"-T_2Me). This is consistent with earlier studies of nogalamycin binding to duplex DNA (Searle *et al.*, 1988; Zhang & Patel, 1990; Searle & Bicknell, 1992; Searle & Lane, 1992; Williams & Searle, 1998, 1999).

Table 5.2: Drug-DNA NOEs observed in the 300 ms mixing time of the NOESY spectrum of the 5'-d(GTGCGAAGCTAC)-3'-nogalamycin complex.

NgH1"-2TMe	NgH7-3GH8	NgH11-2TH1'	NgH14-1GH3'	NgH2'-10TH2"
NgH5"-12CH3'	NgH9-3GH1'	NgH11-2TH2'	NgH14-1GH5'	NgH2'-10TH4'
NgH3-12CH3'	NgH9-3GH2'	NgH11-2TH2"	NgH14-1GH8	NgH2'-10TH5'
NgH7-2TH4'	NgH9-3GH5'	NgH11-2TH3'	NgH14-2TH1'	NgH3'-3GH5'
NgH7-3GH1'	NgH10-3GH1'	NgH11-2TH4'	NgH14-2TH2'	NgH3'-4CH1'
NgH7-3GH2'	NgH10-3GH2'	NgH11-3GH1'	NgH2'-9CH1'	NgH3'-4CH5'
NgH7-3GH4'	NgH10-3GH8	NgH11-3GH4'	NgH2'-9CH6	

The assignment of this section of the spectrum recorded at 15 °C (288 K) was made considerably more difficult by the overlap of the C_9H6 and T_{10}H6 resonances. The NOESY experiment was run at an elevated temperature of 25 °C (298 K) in order to resolve the two crosspeaks and confirm the NOE connectivity through C_9-T_{10}-A_{11} as was the case for the free oligonucleotide. The NOE connectivities seen here point towards the observation of

more than one form of the complex, in which an equilibrium between a frame-shifted conformation and a bulged conformation exists in solution. NOEs were observed between T_{10}H2'/H2"-A_{11}H8 and C_9H2'/H2"-A_{11}H8 as well as T_{10}H1'-A_{11}H8 and C_9H1'-A_{11}H8. The sequential NOEs, *i.e.* T_{10}-A_{11} were less intense, suggesting that the equilibrium slightly favours the bulged conformation. The sugar H2'/H2" protons were observed at similar chemical shifts to those in the free oligonucleotide, whereas in the bulged oligonucleotide, 5'-d(GCTACGAAGTGC)-3', these protons would be observed considerably down-field shifted. The methyl proton of T_{10}, however, was observed to be considerably down-field shifted (by 0.5 ppm).

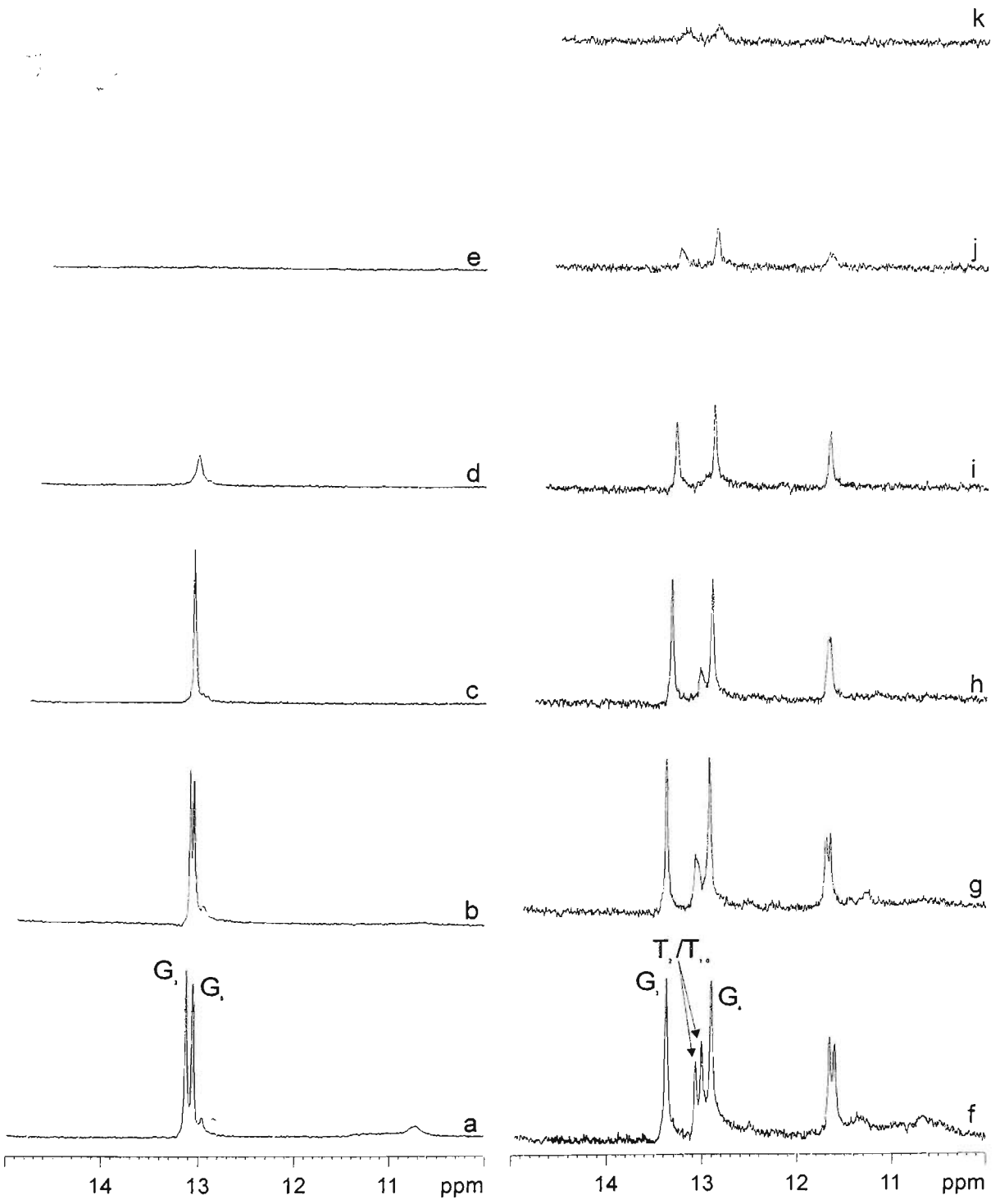


Figure 5.9: 1D NMR of the 10-15 ppm range of 5′-d(GTGCGAAGCTAC)-3′ (a-e) and the nogalamycin complex (f-k). The initial 1D spectra were recorded at 5 °C and the temperature was increased in 10 °C increments until the melting transition occurred as evidenced by the attenuation of the imino peaks.

Figure 5.9 shows the 1D NMR spectra of the 5′-d(GTGCGAAGCTAC)-3′-Ng complex recorded in water, allowing examination of the imino protons present in the complex. The NMR spectra showed an additional four peaks in the 12-15 ppm region of the spectrum (figure 5.9). Two of these peaks were from the phenolic protons of the nogalamycin at 11.6 and 11.7 ppm respectively, whilst the other two peaks are of great interest. The peaks occurred between 13.0-13.1 ppm and were only resolved at a temperature of 5 °C (278 K). Above this temperature they merged to appear as one broad peak. The peaks were of equal, but lower intensity than those of the G_3 and G_8 imino protons, and as such it was postulated that they may be from a T-T "wobble-pair" arising as a result of frame shift induced by the binding of the antibiotic. Alternatively, the peaks observed may be from the T_2 and G_1 imino protons of the Watson-Crick base pairs that form as a result of the complexed oligonucleotide adopting a bulged conformation. As is shown in figure 5.9(f-h), these imino protons were only stable at low temperature and were only present up to 25 °C. The G_3 and G_8 imino protons were stable up to 45 °C after which the attenuation of these peaks was observed. Owing to the temperature instability of these peaks in the 1D NMR spectra, it is proposed that these imino protons were from the formation of the T-T "wobble pair" rather than Watson-Crick base pairs. Above 25 °C, the ends of the oligonucleotide are thought to fray, with the nogalamycin stabilising the oligonucleotide above the intercalation site.

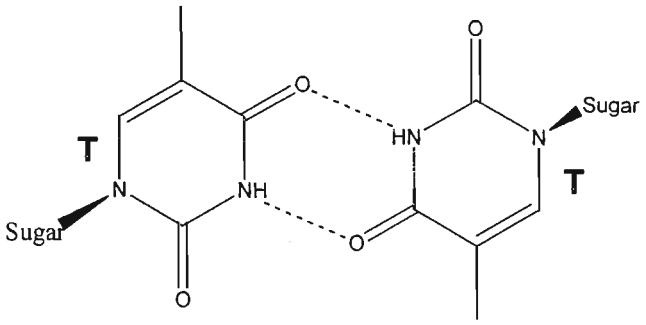


Figure 5.10: T-T "wobble pair" conformation

The latter possibility was further explored by examination of the Watergate-NOESY spectrum of the 5′-d(GTGCGAAGCTAC)-3′-Ng complex, as shown in figures 5.11 and 5.12. In the spectra, the imino protons show connectivities to the amino peaks, which in turn show connectivities to the aromatic base protons. Peaks are also observed showing the connectivity between the imino protons of one Watson-Crick base pair and the adjacent Watson-Crick base pair, *i.e.* between the T_2 imino proton and the G_3 imino proton. Figure 5.11 shows a section of the wg-NOESY spectrum in which the amino protons have been traced across from the amino-imino crosspeaks (not shown). There are two amino protons per adenine or guanine and hence, strong NOEs are observed between the amino protons of individual bases. The amino protons also show NOE crosspeaks to the H5 protons of the cytosine involved in the base pair. These connectivities are illustrated in figure 5.11. The blue lines show the connectivity observed for the C_4-G_8 base pair, while the pink lines represent the G_3-C_9 base pair. The NOEs seen for the amino protons of A_{11} to the NH_2 protons (green lines) could have been the result either of a T-T "wobble pair" or a T-A Watson-Crick base pair. The exact identity could not be confirmed by the method employed. No evidence of the terminal G_1-C_{12} base pair was observed.

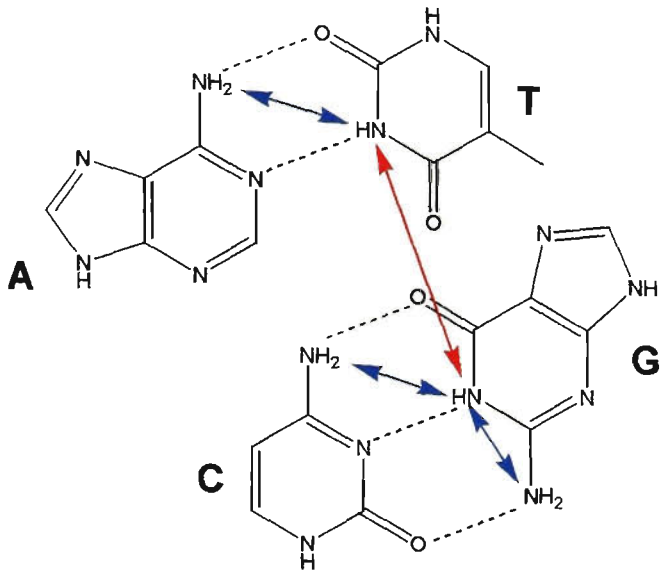


Figure 5.11: NOEs between imino protons and amino protons of base pairs and imino protons of adjacent base pairs in DNA.

Figure 5.11 is an illustration of the base pairing arrangement between the Watson-Crick base pairs in DNA. The blue arrows indicate the NOEs between the imino and amino protons of the individual base pairs, whilst the red arrow shows the NOE present between the adjacent base pair. Crosspeaks observed in the wg-NOESY spectrum corresponding to these types of NOEs allow the identification of the imino protons (as shown in figure 5.12).

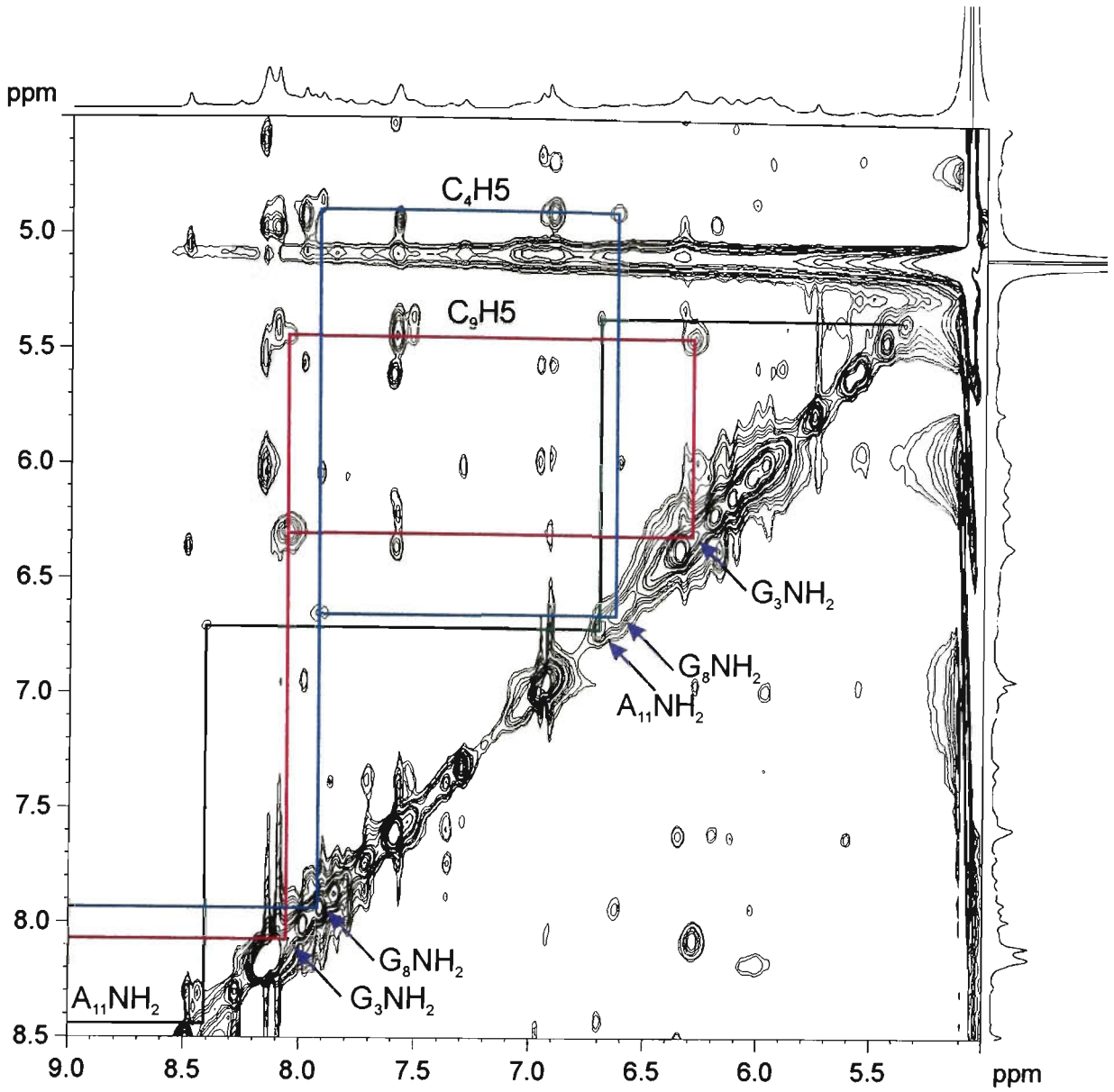


Figure 5.12: Section of the 300 ms wg-NOESY spectrum of the 5′-d(GTGCGAAGCTAC)-3′-Ng complex (3 mM) in H_2O containing 0.1 M NaCl/100 mM NaH_2PO_4 (pH 7.0), showing the connectivity of amino protons to amino protons and amino protons to aromatic base protons.

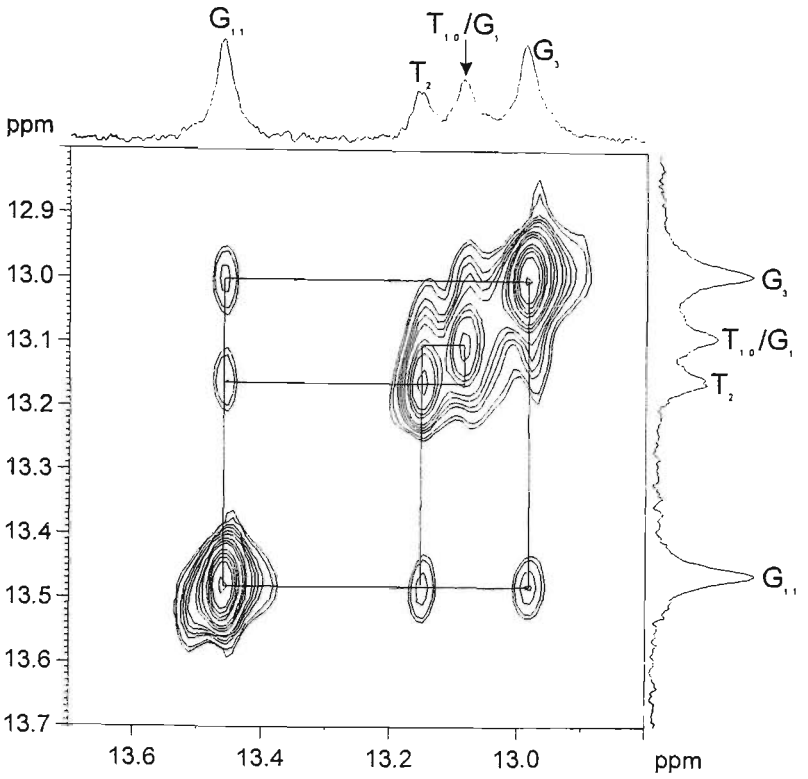


Figure 5.13: Portion of the 300 ms Watergate-NOESY spectrum of the 5′-d(GTGCGAAGCTAC)-3′-Ng complex (3 mM) in H_2O containing 0.1 M NaCl/100 mM NaH_2PO_4 (pH 7.0), showing NOE crosspeaks between the imino protons of adjacent base pairs.

Figure 5.13 shows the wg-NOESY spectrum of the 5′-d(GTGCGAAGCTAC)-3′-Ng complex, in which NOE crosspeaks are observed between the adjacent base pairs of the complex. There are two crosspeaks resolved in this spectrum. The first is between the G_8 imino proton of the C_4-G_8 base pair and the G_3 imino proton of the G_3-C_9 base pair. The second crosspeak corresponds to the NOE between the G_3 imino proton and the T_2 imino proton of either the T-T "wobble pair" or T-A base pair. The last peak in the 1D spectrum was from either the T_{10} or G_1 imino proton of the T_2-T_{10} pair or the G_1-C_{12} base pair. It was not possible to distinguish between the latter two possibilities.

There were extensive drug-DNA NOEs observed in the 5′-d(GTGCGAAGCTAC)-3′-Ng complex, between nogalamycin methyl resonances and the T_2 sugar as would be expected as the thymine was positioned within the intercalation site. NOEs were observed between NgH1"-T_2Me, NgH7-T_2H4', NgH11-T_2H1'/H2'/H2"/H3'/H4' and NgH14-T_2H1'/H2'. There was no evidence of any NOEs to the adenine (A_{11}), which if base paired with the T_2 would also have been positioned similarly. There were, however, NOEs noted for the nogalose sugar protons (H2", H4' and H5') to T_{10}, suggesting that this thymine was not in a bulged conformation and had taken the place of the A_{11} as would occur for a frame shift event.

5′-d(GCTACGAAGTGC)-3′

Figure 5.14 shows the 1D NMR spectra resulting from the addition of aliquots of nogalamycin to the free oligonucleotide, 5′-d(GCTACGAAGTGC)-3′. These 1D-NMR titration studies showed that nogalamycin (Ng) formed a clean 1:1 complex with 5′-d(GCTACGAAGTGC)-3′, with resonances from the free DNA replaced by those of a complex containing a single-bound ligand.

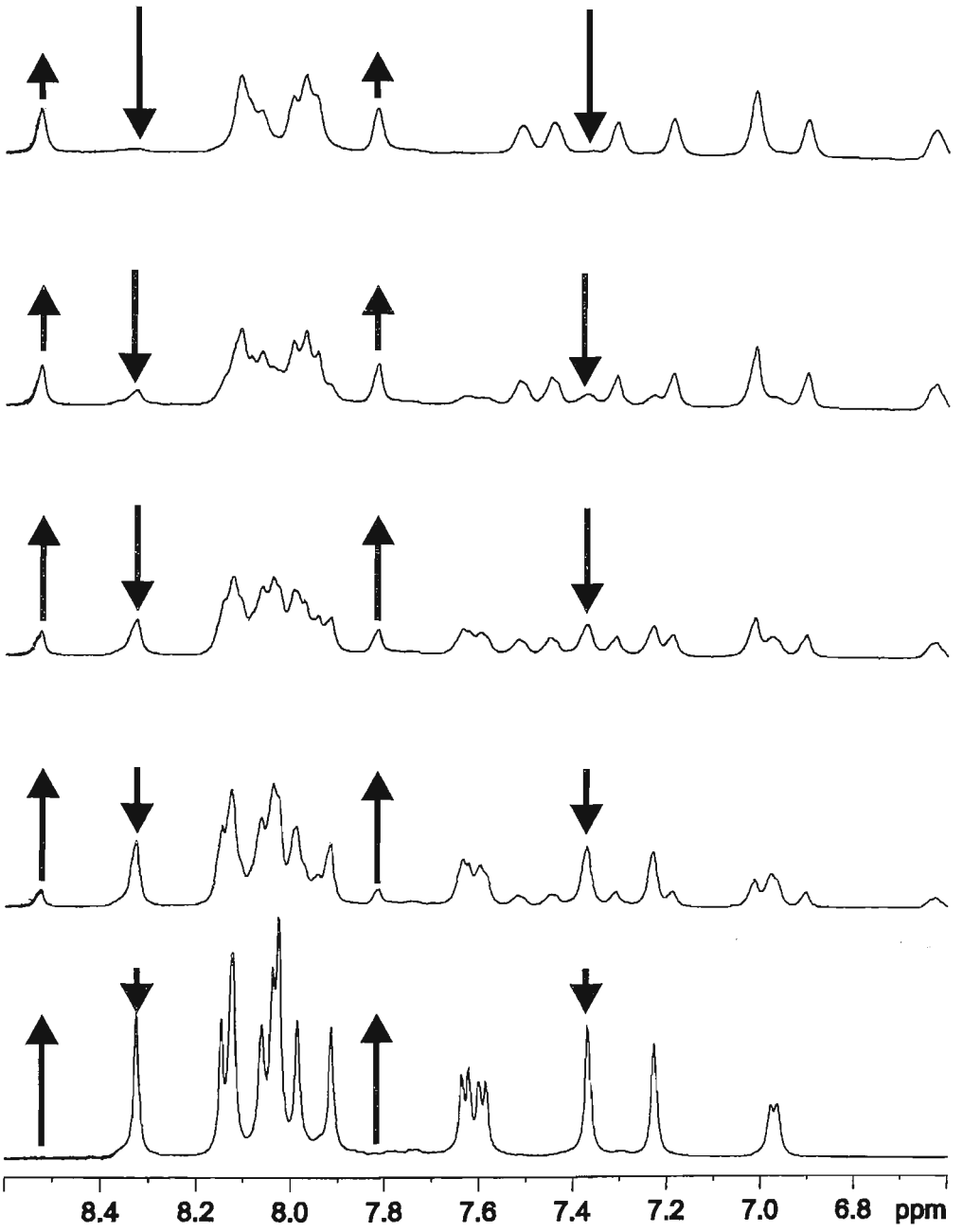


Figure 5.14: 1D NMR spectra acquired after the addition of 5 mM nogalamycin: (i) 0:1 (ii) 0.17:1 (iii) 0.35:1 (iv) 0.70:1 (v) 1:1 complex. The arrows highlight a number of clearly defined signals, owing to free DNA that decrease upon addition of nogalamycin. Arrows also indicate new signals that increase owing to complex formation.

As was seen with 5′-d(GTGCGAAGCTAC)-3′, the most pronounced changes to chemical shifts upon ligand binding to 5′-d(GCTACGAAGTGC)-3′occurred at the intercalation site, *i.e.* at the 5′-TpG step. In particular, the A_4H8 proton was shifted down-field to 8.545 ppm. Figure 5.15 shows the chemical shift perturbation upon ligand binding. The resonances of C_2, A_4, T_{10} and G_{11} are most dramatically affected, as are the T_3 resonances, which will be discussed separately. These data provide a clear indication of the position of the binding site within the oligonucleotide sequence.

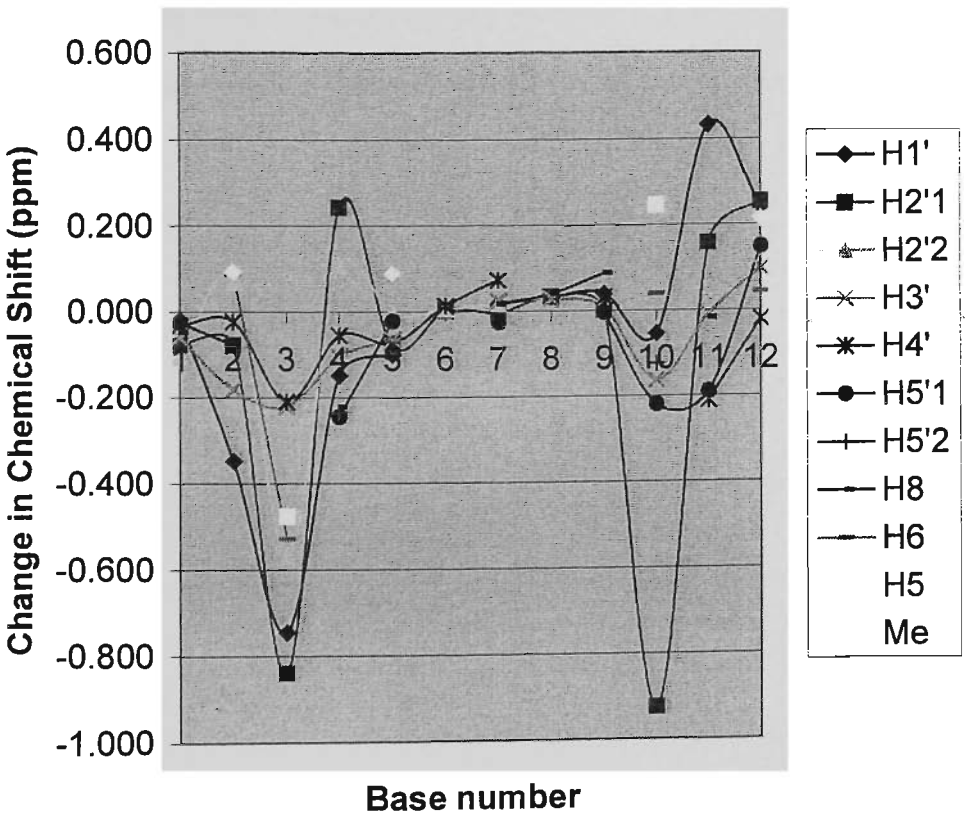


Figure 5.15: Chemical shift perturbation in the NMR spectra upon ligand binding to 5′-d(GCTACGAAGTGC)-3′.

Figure 5.16 shows a portion of the NOESY spectrum of 5′-d(GCTACGAAGTGC)-3′-Ng complex indicating the H8/H6-H2'/H2" connectivity along the DNA backbone. The connectivity can be followed from G_1 to C_2 to T_3, although the latter NOEs are significantly weaker than the afore-mentioned. Connectivity through $A_4G_5A_6A_7$ is also observed (solid lines). A loss of sequential connectivity is noted for the loop region, *i.e.* between A_7 and A_8. The connectivity along the 3′-strand can then be seen from A_8 to G_9, G_9 to T_{10} (solid lines), then from T_{10} to Ng-H11 and Ng-H11 to G_{11} (dotted lines) and from G_{11} to C_{12}. As was the case with 5′-d(GCTACGAAGTGC)-3′, the connectivity along the DNA backbone can be followed with the drug bridging the gap between T_{10} and G_{11}. These data clearly establish that the site of nogalamycin intercalation is between the expected TpG step.

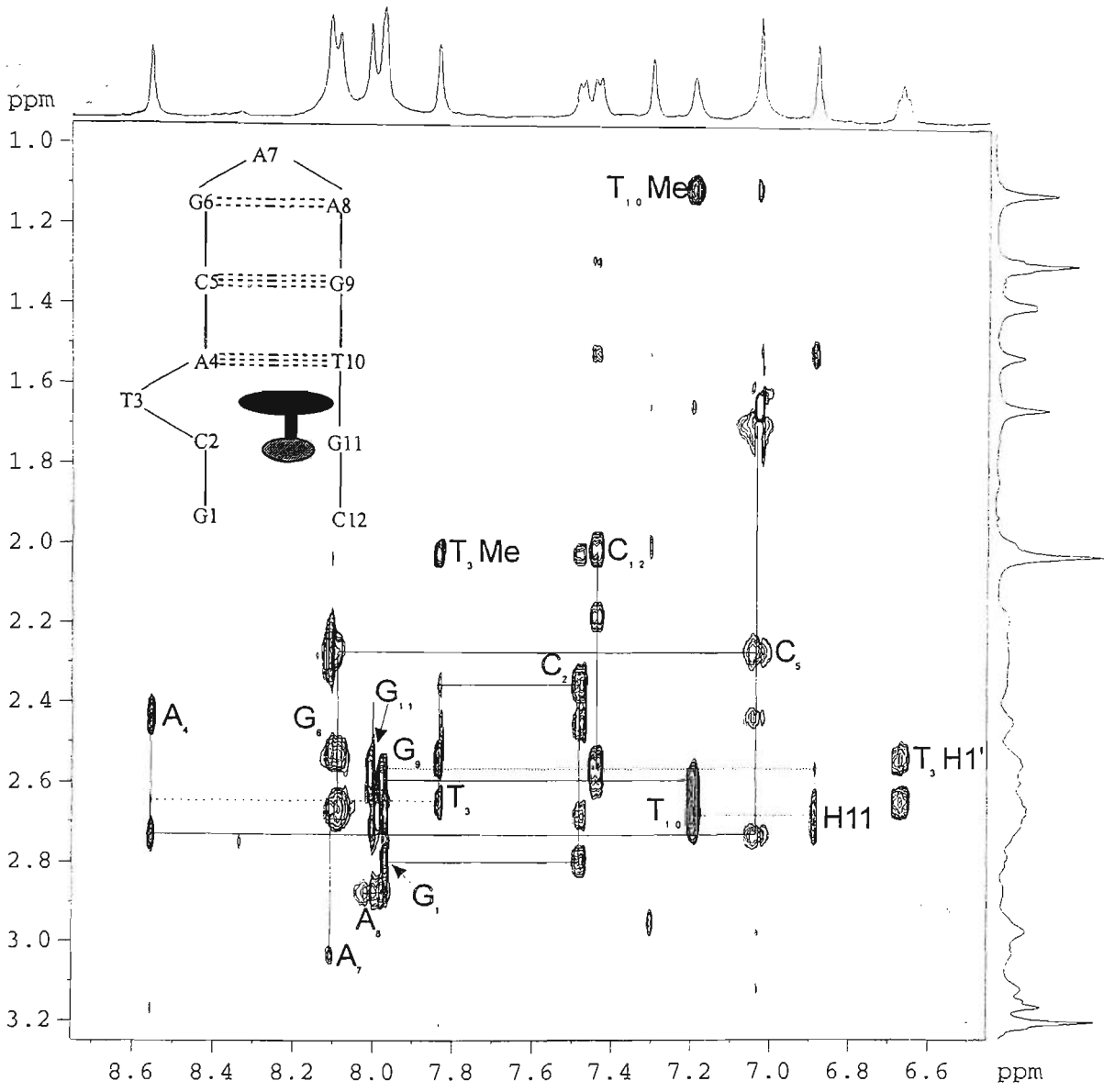


Figure 5.16: Section of the 300 ms NOESY spectrum of 5′-d(GCTACGAAGTGC)-3′-Ng complex (3 mM) in D_2O containing 0.1 M NaCl/100 mM NaH_2PO_4 (pH 7.0) showing the H8/H6-H2'/H2" connectivities. The solid lines show sequential connectivity, whilst the dashed lines show loss of connectivity. The dotted lines show connectivity observed through the nogalamycin at the intercalation site.

Table 5.2 shows the change in chemical shift upon ligand binding for the aromatic base protons (H8/H6) and the sugar H1' protons. It is notable that the largest chemical shift changes observed were for the T_3 resonances (more later).

Table 5.2: Change in chemical shift upon nogalamycin binding to 5′-d(GCTACGAAGTGC)-3′: (a) H8/H6 and (b) H1'

			3T			
			-0.529			
(a) **H8/H6**	**1G**	**2C**	**4A**	**5C**	**6G**	
	-0.048	0.088	-0.222	-0.066	---	**7A**
	12C	**11G**	**10T**	**9G**	**8A**	0.016
	0.042	-0.018	0.039	0.087	0.04	

			3T			
			0.745			
(B) **H1'**	**1G**	**2C**	**4A**	**5C**	**6G**	
	-0.019	-0.348	0.149	-0.102	---	**7A**
	12C	**11G**	**10T**	**9G**	**8A**	0.016
	0.241	0.434	-0.055	0.039	0.033	

A large number of drug-DNA contacts (32) were observed in the 300 ms NOESY spectrum of the complex and are given in table 5.4.

Table 5.4: Drug-DNA NOEs observed in the 300 ms mixing time of the NOESY spectrum of the 5′-d(GCTACGAAGTGC)-3′-nogalamycin complex

NgH1"-10TH6	NgH9-12CH5'	NgH11-11GH8	NgH3'(2)-12CH1'	NgH4'-3TMe
NgH1"-10TMe	NgH10-10TH6	NgH3'(1)-12CH1'	NgH4'-2CH1'	NgH4'-4AH4'
NgH5"-10TMe	NgH11-10TH1'	NgH3'(1)-12CH2"	NgH4'-2CH2'	NgH4'-4AH8
NgH3-10TMe	NgH11-10TH2'	NgH3'(1)-12CH5'	NgH4'-2CH2"	NgH5'-4AH4'
NgH9-11GH1'	NgH11-10TH3'	NgH3'(2)-2CH1'	NgH4'-2CH5'	
NgH9-11GH4'	NgH11-11GH1'	NgH3'(2)-2CH4'	NgH4'-3TH3'	
NgH9-12CH1'	NgH11-11GH4'	NgH3'(2)-3TMe	NgH4'-3TH6	

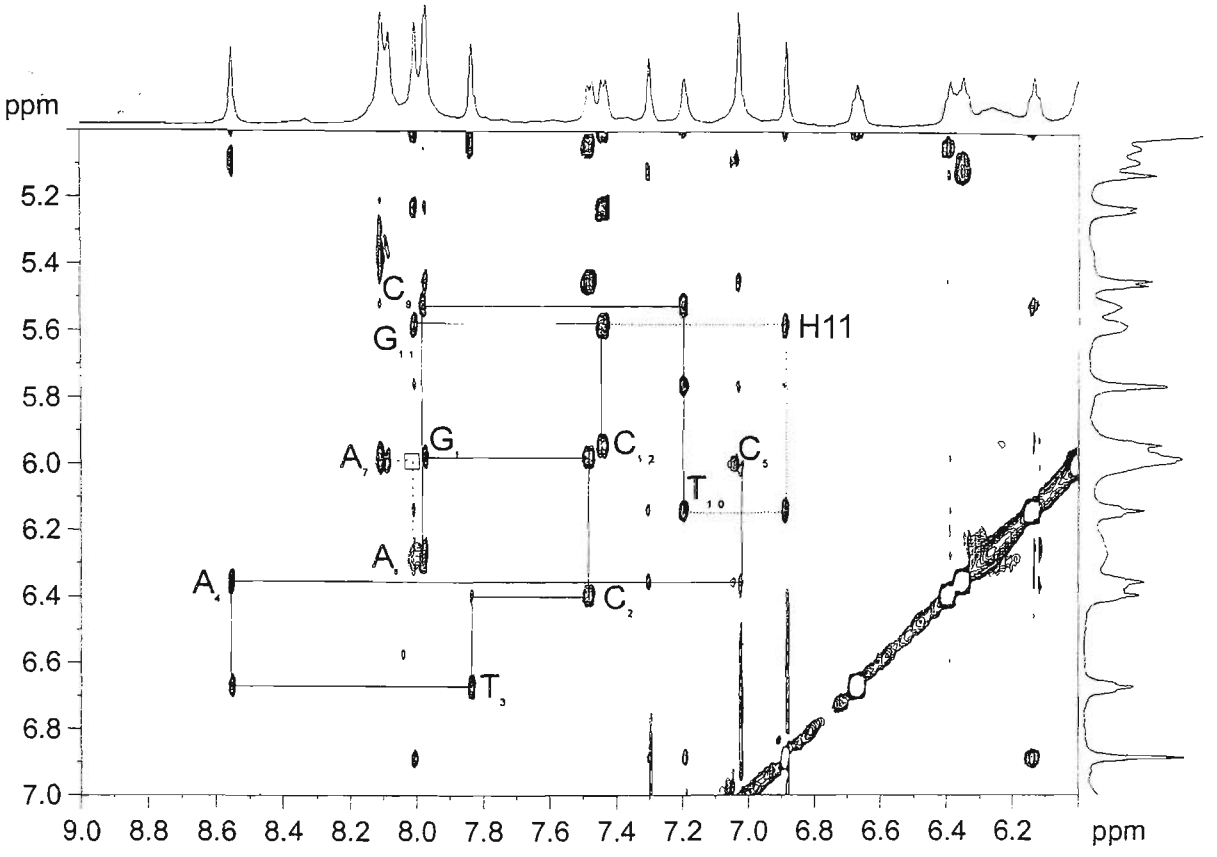


Figure 5.17: Section of the 300 ms NOESY spectrum of the 5′-d(GCTACGAAGTGC)-3′-Ng complex (3 mM) in D_2O containing 0.1 M NaCl/100 mM NaH_2PO_4 (pH 7.0) showing resonances that define the intercalation site. The H8/H6-H1' connectivities along the DNA backbone are shown by solid lines while the NOEs highlighting the drug-DNA NOEs are shown by dotted lines.

Figure 5.17 illustrates the sequential connectivity observed along the 3′-strand of the bulged hairpin in the 5′-d(GCTACGAAGTGC)-3′-Ng complex. The solid lines show the sequential connectivity along the DNA backbone while the dotted lines depict loss of connectivity at the intercalation site. NOEs are seen between the T_{10}H1' and H11 of nogalamycin and the NgH11 and G_{11}H1'. The square is indicative of further loss of sequential connectivity in the loop region (dashed lines). The 3′-strand can be followed

from G_1 to G_5, where a loss of connectivity is observed in the loop region, and from A_7 to C_{12} for the 5′-strand. NOE crosspeaks are observed between T_{10}H1' and Ng-H11 and Ng-H11 and G_{11}H1', showing clearly that nogalamycin is bound at the 5′-TpG step. It is important to note the NOEs between C_2 and T_3 are weaker relative to the other NOEs observed (figure 5.17). This suggests that the thymine is adopting an unusual conformation, such as a bulge, so that its protons are no longer sufficiently close in space to produce NOE crosspeaks with the adjacent bases (more later).

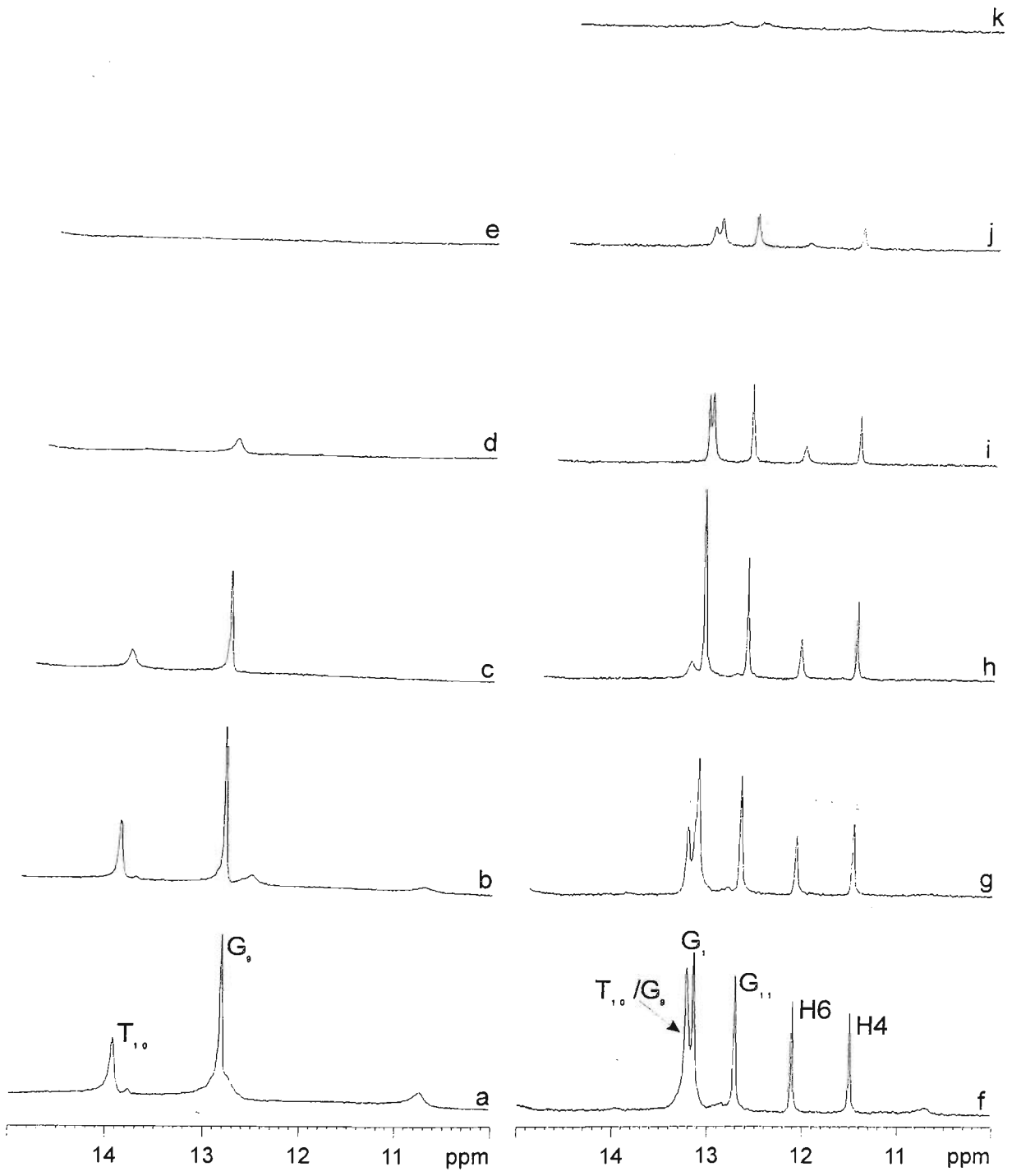


Figure 5.18: 1D NMR of the 10-15 ppm range of 5′-d(GCTACGAAGTGC)-3′ (a-e) and the nogalamycin complex (f-k). The initial 1D spectra were recorded at 5 °C and the temperature was increased in 10 °C increments until the melting transition occurred as evidenced by the attenuation of the imino peaks.

Figure 5.18 shows the 1D NMR spectra of 5′-d(GCTACGAAGTGC)-3′ and its complex with Ng recorded in water. Examination of the imino protons present in the complex showed five peaks in the 12-15 ppm region of the spectrum. Two of these peaks are again from the H4 and H6 phenolic protons of the nogalamycin at 11.5 and 12.1 ppm respectively. The peak observed at ~13.3 ppm was that owing to the T_{10} imino proton and was slightly up-field compared to the T_{10} in the free oligonucleotide. The G_9 imino proton was shifted up-field by ca. 0.5 ppm at ~13.3 ppm. This peak, however, had a much greater line width and was proposed to be the result of an overlap of the G_9 and T_{10} imino protons. These were resolved at elevated temperature (35 °C) as shown in figure 5.18(i). The peak at 13.2 ppm was from the G_1 imino proton. This peak was of greater intensity owing to the overlap occurring with the peak at 13.3 ppm. The peak observed at ~12.7 ppm was that owing to the G_{11} imino proton. These data indicate that the G_1-C_{12}, C_2-G_{11}, A_4-T_{10} and C_5-G_9 Watson-Crick base pairs are all formed and hence protected from hydrogen-deuterium exchange. The T_3 bulge was thus stabilised by the binding of the ligand.

Table 5.5 shows the chemical shift values for all of the observed T_3 resonances for both the free and complexed forms of 5′-d(GCTACGAAGTGC)-3′. A large number of differences between the free and complexed oligonucleotide were observed, with particular reference to the T_3 base. The T_3 methyl peak was observed to be down-field shifted considerably at 2.022 ppm, compared to 1.110 ppm for the T_{10} methyl peak of the complex, and 1.546 and 1.355 ppm respectively for the T_3 and T_{10} of the free oligonucleotide.

Table 5.5: Chemical shift values for the T_3 resonances upon nogalamycin binding to 5′-d(GCTACGAAGTGC)-3′.

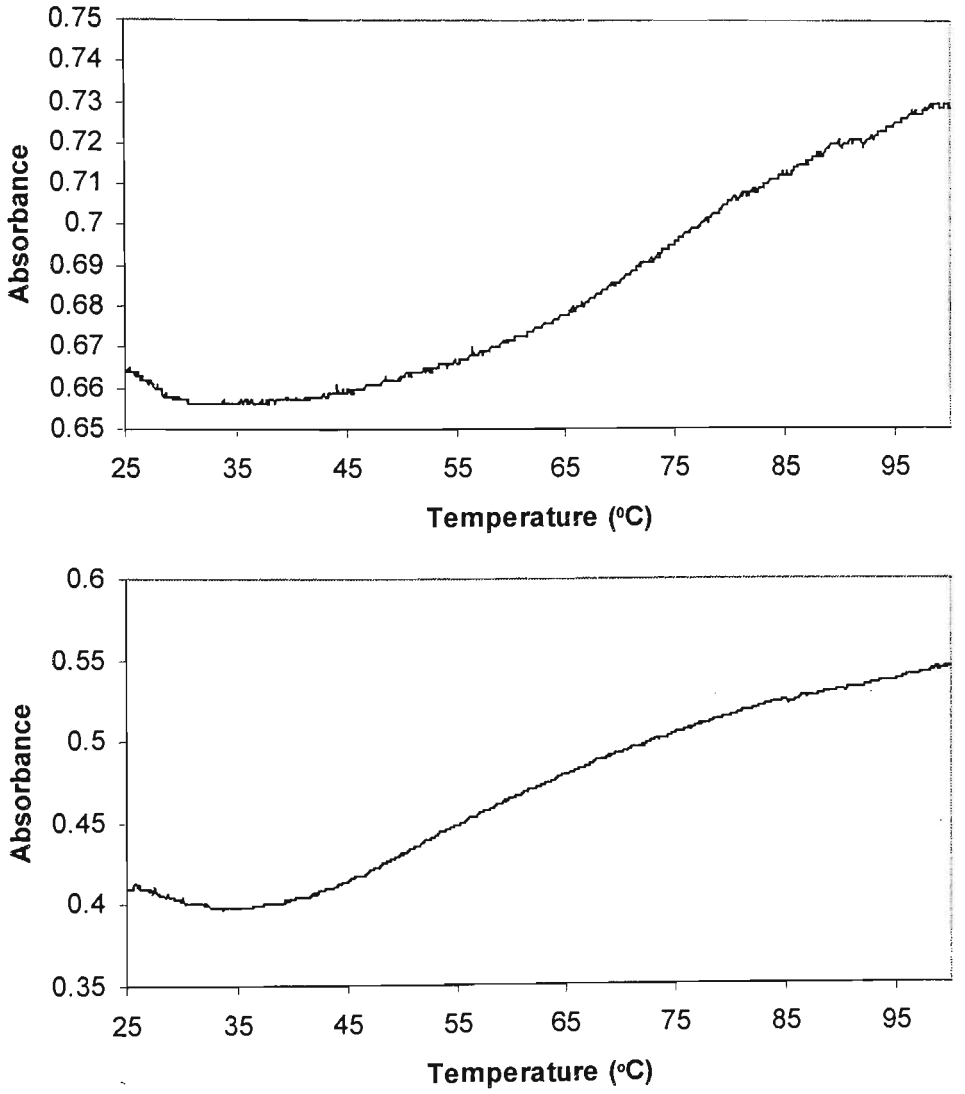
	H1'	H2'1	H2'2	H3'	H4'	H5'1	H6	Me
d(GCTACGAAGTGC)	5.920	1.699	2.077	4.766	4.239	4.102	7.299	1.546
d(GCTACGAAGTGC) +Ng	6.665	2.540	2.645	4.987	4.449	---	7.828	2.022

For the thymine (T_3) to adopt a bulged conformation, the torsion angles must be vastly different from those present in B-form DNA for this section of the backbone. The loss of sequential connectivity (figure 5.3b) through the T_3 arises as a result because the C_2H6 and T_3H1' are no longer close enough in space to observe NOEs between these resonances. Hence, these data compare to the chemical shift values of the free thymine nucleotide, which has been shown to have chemical shift values for TH6 between 8.0-8.5 ppm and TMe between 2.3-2.4 ppm. (Cantor & Schimmel, 1980) T_3H1' has a shift characteristic of single-stranded DNA where no base stacking or ring current effects are observed. This evidence further implies that the thymine is adopting a bulged conformation as the T is acting as a free nucleotide.

5.5 Melting Temperature Studies

Figure 5.19 shows the UV melting curves for (a) 5′-d(GTGCGAAGCTAC)-3′ and (b) the 5′-d(GTGCGAAGCTAC)-3′-Ng complex. The UV melting temperature studies gave melting transitions of 73.7 °C for the free bulge sequence, 5′-d(GTGCGAAGCTAC)-3′, and 63.6 °C for the complex with nogalamycin. The melting temperatures observed for both the free and complexed forms of 5′-d(GCTACGAAGTGC)-3′ were 65.6 °C and 64.3 °C respectively. The higher T_m noted for 5′-d(GTGCGAAGCTAC)-3′ compared to 5′-

d(GCTACGAAGTGC)-3′ can be explained in terms of the H-bonding interactions in the free oligonucleotides with two G-C base pairs for the former and one G-C and one A-T base pair for the latter. Three hydrogen bonds are formed in each G-C base pair compared



to two in each A-T base pair.

Figure 5.19: UV melting curves of (a) 5′-d(GTGCGAAGCTAC)-3′ (6 μM) and (b) the complex formed between 5′-d(GTGCGAAGCTAC)-3′ and nogalamycin (6 μM) in H_2O containing 300 mM NaCl/30 mM NaH_2PO_4 (pH 7.0).

The melting temperature data suggest that the complex formed between 5′-d(GTGCGAAGCTAC)-3′ and nogalamycin is less stable than the free oligonucleotide. The decrease in T_m observed is proposed to result from the destabilising effect of the formation of the T-T and G-A "wobble pairs" formed. Upon ligand binding, an equilibrium between a frame-shifted conformation containing "wobble pairs" and a bulged conformation exists in solution at low temperature. As the temperature is increased, fraying at the ends occurs owing to the lower stability of the terminal non-Watson-Crick base pairs. The melting curve is broad as a result of the dynamic nature of the oligonucleotide resulting from the conformational changes and end-fraying. Watson-Crick base pairs are formed between G_1-C_{12} and T_2-A_{11}, however, the additional stability imparted from their formation is less than the destabilisation arising as a result of bulge formation.

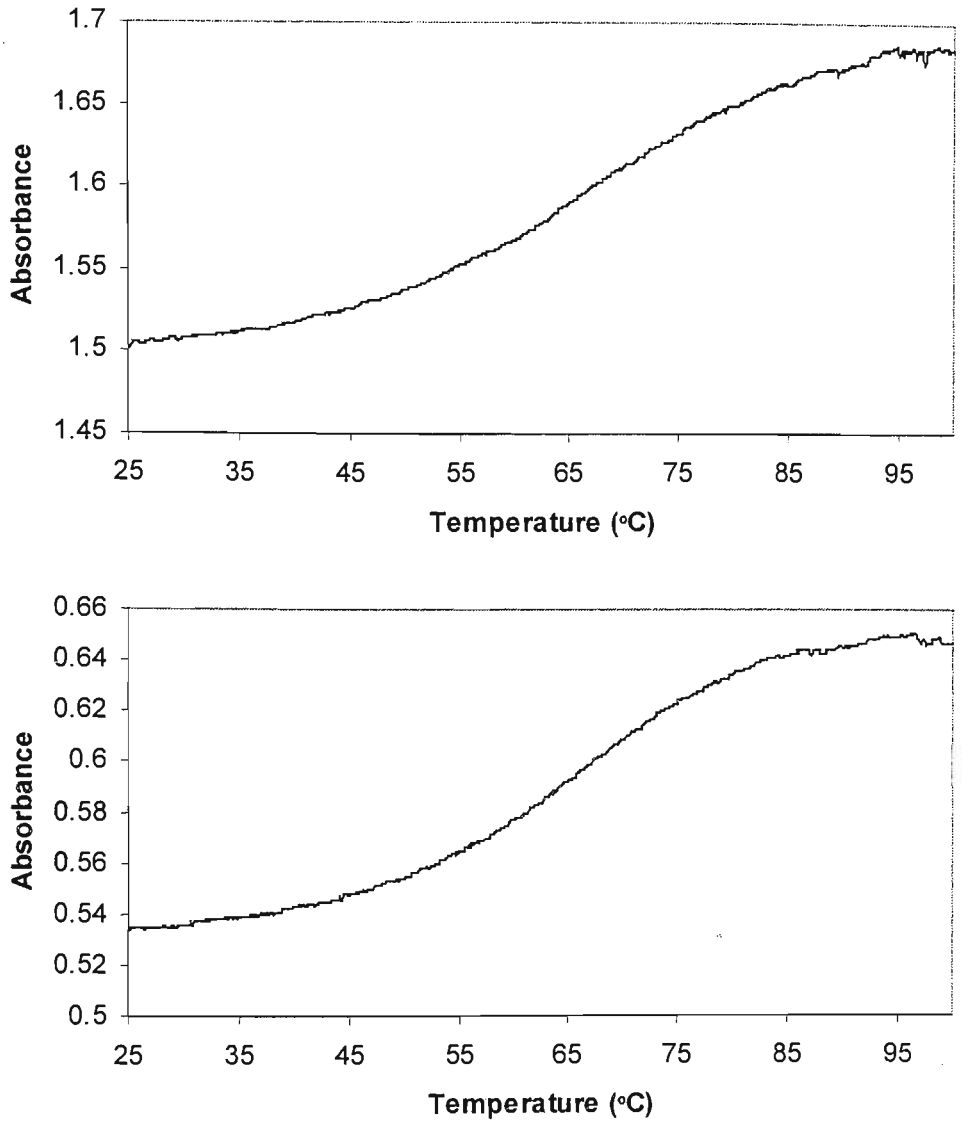


Figure 5.20: UV melting curves of (a) 5′-d(GCTACGAAGTGC)-3′ (6 μM) and (b) the complex formed between 5′-d(GCTACGAAGTGC)-3′ and nogalamycin (6 μM) in H_2O containing 300 mM NaCl/30 mM NaH_2PO_4 (pH 7.0).

Figure 5.20 shows the UV melting curves for (a) 5′-d(GCTACGAAGTGC)-3′ and (b) the 5′-d(GCTACGAAGTGC)-3′-Ng complex. The complex formed between 5′-d(GCTACGAAGTGC)-3′ and the ligand has approximately (within the range of error) the

same melting temperature. It was postulated that the degree of stabilisation brought about by ligand binding was approximately equal to the destabilisation of the oligonucleotide structure caused by inclusion of a single base bulge within the intercalation site. In contrast to the previous oligonucleotide, 5′-d(GTGCGAAGCTAC)-3′, the two Watson-Crick base pairs formed are both G-C base pairs and as such, impart further stability than would result from a single G-C and an A-T base pair. The melting curves for both the free bulge sequences and the complexed form of 5′-d(GTGCGAAGCTAC)-3′ were seen to be broad in comparison to the melting curve for the 5′-d(GCTACGAAGTGC)-3′-Ng complex. This can be explained in terms of the terminal fraying observed for this oligonucleotide and the frame shift event occurring for the Ng complex of 5′-d(GTGCGAAGCTAC)-3′. The destabilisation of the terminal base pairs results in the melting transition occurring over a greater temperature range accounting for the width of the curves. Four Watson-Crick base pairs are formed in the Ng complex of 5′-d(GCTACGAAGTGC)-3′ and as such a sharper melting transition is noted (figure 5.20). In a temperature dependent study of the imino proton line widths, it was also observed that all of the imino protons underwent simultaneous attenuation at ~45 °C (figure 5.18).

5.6 Electrospray Ionisation Mass Spectrometry (ESI-MS)

The complexes were also analysed by electrospray ionisation mass spectrometry. The procedures used were the same as described in section 4.16. Figure 5.21 shows the ESI mass spectrum of the T-bulge complex (Ng-5′-d(GCTACGAAGTGC)-3′). The peaks observed at *m/z* 1113.7 And 1485.3 represent the complexed forms of the oligonucleotide, *i.e.* the $[M+Ng-5H]^{4-}$ and $[M+Ng-4H]^{3-}$ ions respectively. The free form of the ligand, $[Ng-H]^{-}$, was observed at *m/z* 786.2. In this case, the ions from the drug-DNA complex were less abundant relative to the free oligonucleotide than was the case for the complex with the double hairpin under similar ionisation conditions (figure 4.24b). We examined the binding of nogalamycin to the bulged hairpin 5′-d(GTGCGAAGCTAC)-3′. In the ESI mass spectrum (not shown), the intact adduct was not detected, but rather only the unbound oligonucleotide was observed, indicating that this was the least stable complex. This is consistent with the NMR data for the Ng-5′-d(GTGCGAAGCTAC)-3′ complex that suggest that the complex is in equilibrium between a bulged and frame-shifted conformation, the latter resulting in the formation of a significant population of non-Watson-Crick base pairs (T-T and G-A) which are expected to have lower stability. Consequently, these unstable non-complementary base pairs, which the NMR imino proton exchange measurements suggest, are dynamic and involved in end-fraying, do not interact strongly with the drug on one side of the intercalation site.

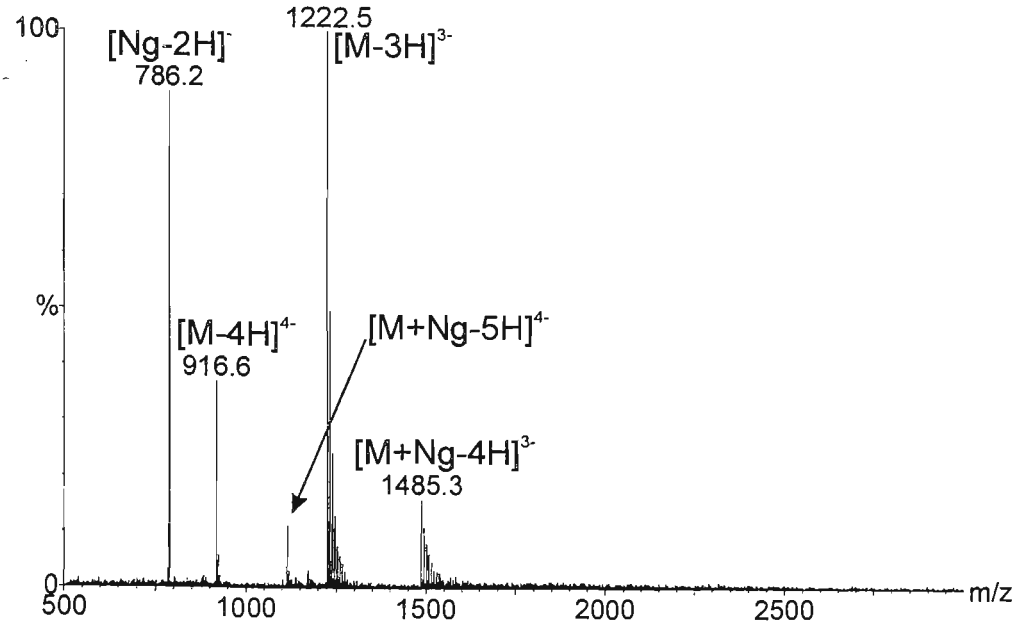


Figure 5.21: ESI-MS spectrum of the 5′-d(GTGCGAAGCTAC)-3′-Ng complex

5.7 Comparison with Previous Studies

The binding of nogalamycin to two DNA heptamers, 5′-CT$_b$GTACG-3′ and 5′-CGTACT$_b$G-3′, containing a bulged T nucleotide (T$_b$) has been examined previously by Caceres-Cortes and Wang (Caceres-Cortes & Wang, 1996). They found that both duplexes were significantly stabilised by the binding of the ligand and that the binding induced the extra T and opposing base to form "wobble-pairs" which stacked on the aglycone chromophore. This resulted in base mispairing and, rather than a bulged T, a frame-shift event. The data obtained in that study illustrate that this frame-shift resulted in "dangling" 3′-G and 5′-C nucleotides. The results gained in those studies, however, contrast with those gained here since nogalamycin binding to the oligonucleotides 5′-GTGCGAAGCTAC-3′ and 5′-GCTACGAAGTGC-3′ stabilised the bulged T nucleotide (figure 5.22). In the case of 5′-GTGCGAAGCTAC-3′, the complex was thought to be in dynamic equilibrium

between different conformations. The data obtained for the frame-shifted conformation are similar to the results described by Caceres-Cortes and Wang and can be explained in terms of the requirement for the G-C base pair for Ng recognition. As the G-C base pair is internal to the oligonucleotide, the ligand may still bind and the frame-shift event does not disturb this essential base pairing conformation. For the oligonucleotide, 5′-GCTACGAAGTGC-3′, however, the G-C base pair will only remain intact if the mispaired T adopts a bulged conformation.

The results gained by Caceres-Cortes and Wang were strongly influenced by end-effects as upon formation of a T bulge, only one Watson-Crick base pair would be formed. Two base pairs form upon T bulge formation for the oligonucleotides examined here giving rise to additional stability. It was postulated that if longer sequences of oligonucleotide in which the number of base pairs between the intercalation site and the termini were increased, the frame-shifted conformation observed for 5′-GTGCGAAGCTAC-3′ would no longer be observed. This is the subject of further investigation.

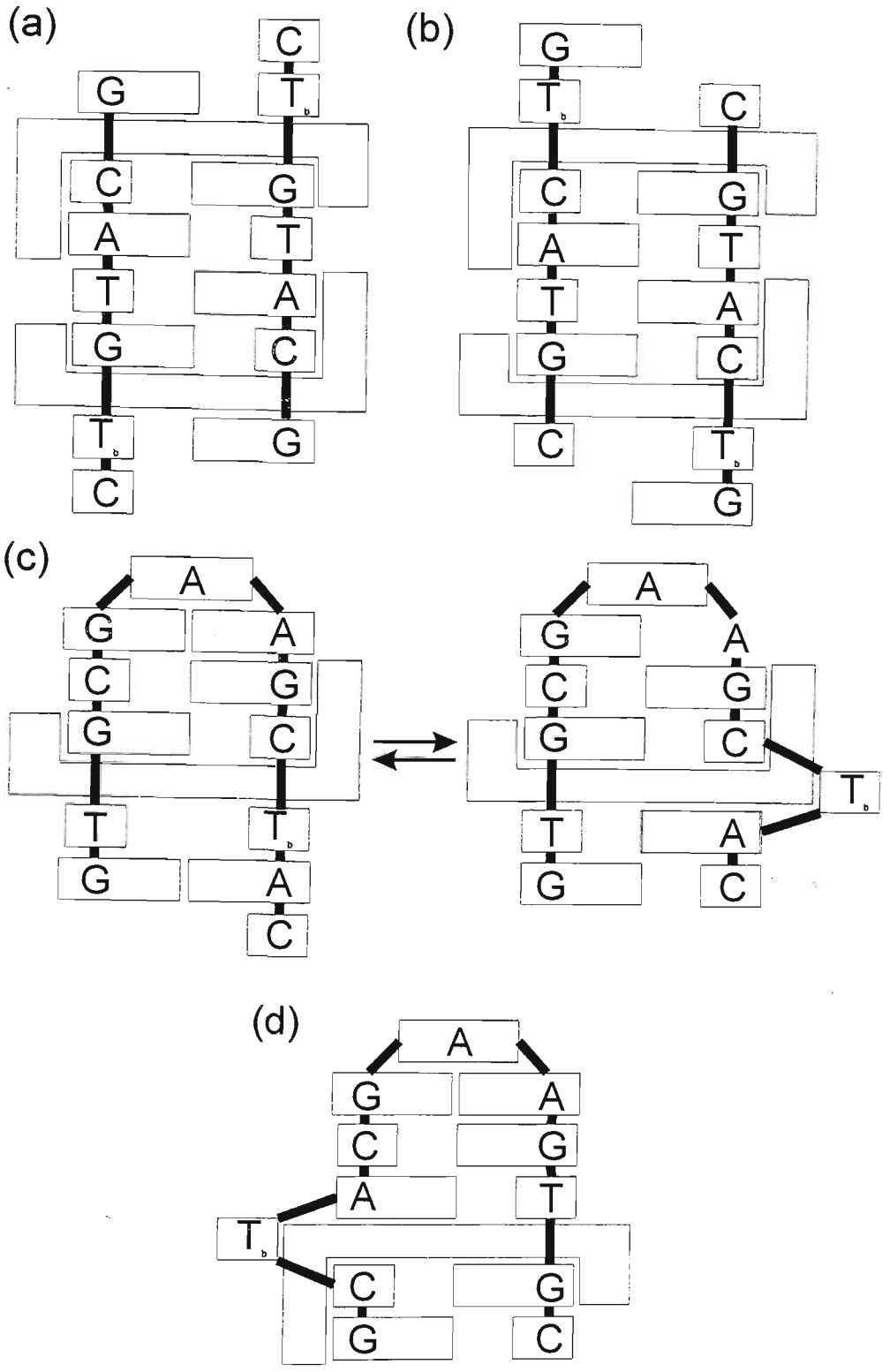


Figure 5.22: Schematic illustration of binding observed between nogalamycin and (a) 5′-CT_bGTACG-3′; (b) 5′-$CGTACT_bG$-3′; (Caceres-Cortes & Wang, 1996) (c) 5′-GTGCGAAGCTAC-3′; and (d) 5′-GCTACGAAGTGC-3′.

5.8 Conclusions

NMR spectroscopy was utilised for the complete characterisation of the bulged hairpins, 5′-d(GTGCGAAGCTAC)-3′ and 5′-d(GCTACGAAGTGC)-3′. For both of the bulge sequences studied, the unpaired base had a destabilising effect on the terminal base pairs causing fraying the end of the duplex. The strands maintained a B-form conformation, however, as was evidenced by the sequential connectivity observed along the DNA backbone. A loop region was formed as evidenced by the unusual chemical shift values observed for the sugar H4', H5' and H5" protons. The stem region above the bulge (unpaired thymine) was also formed with Watson-Crick base pairs between the G_3-C_9 and C_4-G_8 for 5′-d(GTGCGAAGCTAC)-3′ and between A_4-T_{10} and C_5-G_9 for 5′-d(GCTACGAAGTGC)-3′.

The reactions between nogalamycin and each of the bulge sequences were monitored by 1D NMR experiments. The appearance of new peaks in the spectra owing to the complexed form and the attenuation of peaks owing to the free oligonucleotide were testimony to formation of a 1:1 complex with both oligonucleotides. This was further confirmed by ESI-MS. Nogalamycin was found to bind at the 5′-TpG step with a large number of NOEs observed between the ligand and the T_2 and G_3 bases (5′-d(GTGCGAAGCTAC)-3′) and T_{10} and G_{11} bases (5′-d(GCTACGAAGTGC)-3′). Upon ligand binding, changes in chemical shift values for the DNA resonances were observed, with the largest differences noted for the nucleotides in the intercalation site, *i.e.* T_2, G_3, C_9, T_{10} and A_{11} for 5′-d(GTGCGAAGCTAC)-3′ (C_2, A_4, T_{10} and G_{11} for 5′-d(GCTACGAAGTGC)-3′).

The intercalation of nogalamycin gave rise to an equilibrium in solution between a frame shifted conformation and bulged conformation in 5′-d(GTGCGAAGCTAC)-3′ resulting in the 3′-C 'dangling' and the formation of a 'T-T' wobble pair for the former conformation. The equilibrium is thought to favour the bulged conformation as evidenced by stronger NOEs observed in the NOESY spectra for the C_9-A_{11} resonances over the T_{10}-A_{11} resonances. The 1D Watergate spectrum showed four peaks owing to imino protons present in the hydrogen bonds formed between G_3-C_9, C_4-G_8 and T_2-T_{10} (or T_2-A_{11} and G_1-C_{12}). The peaks owing to T_2 and T_{10} (or T_2 and G_1) were of equal but lower intensity compared to the peaks owing to the guanine imino protons. Formation of either the T-T "wobble pair", or the base pairs after the bulged site, was only seen at low temperatures and attenuation of these peaks was observed prior to melting of the complex. Therefore, the binding of the ligand apparently resulted in a dynamic equilibrium between a frame-shifted conformation and a bulged conformation. The melting temperature of the complex was noted to be lower than that of the free oligonucleotide and can be accounted for in terms of the destabilising effect of the frame shift event. The unusual breadth of the melting curve was a consequence of the dynamic conformational changes occurring.

The binding of nogalamycin to 5′-d(GCTACGAAGTGC)-3′ stabilised the bulge positioned at T_3. This was a consequence of the requirement of the G-C base pair below the intercalation site for nogalamycin recognition. In this case, four imino protons were observed in the 1D NMR spectrum of the complex corresponding to G_1-C_{12}, C_2-G_{11}, A_4-T_{10} and C_5-G_9. The melting curves were noted to be sharper than those of the previous oligonucleotide owing to the simultaneous melting of the entire structure. The T_m obtained

for the complex was approximately the same as the free oligonucleotide, suggesting that the stabilising effect of nogalamycin intercalation is approximately equal to the destabilising effect of the bulge.

The ESI-MS data indicate that the relative stability of the complexes formed with nogalamycin is in the order: Ng-5′-d(GCGAAGCACGAAGT)-3′ > Ng-5′-d(GCTACGAAGTGC)-3′ > Ng-5′-d(GTGCGAAGCTAC)-3′. These results were significant because they show that interactions in non-covalent drug-DNA complexes involving destabilising effects from single strand breaks and bulged-bases are much weaker than those in other studies, in which drug-DNA complexes have been examined by ESI-MS. (Beck *et al.*, 2001) We propose this is because the ionisation conditions can be controlled more carefully using salt-tolerant ESI source configurations. The relative order of binding affinities to this family of unusual DNA substrates could be qualitatively determined from the ESI-MS data, and drug binding stoichiometries of 1:1 in all cases has been confirmed, consistent with intercalation at the preferred 5′-TG step.

The results gained by Caceres-Cortes and Wang were thought to be strongly influenced by end-effects. Comparison of their results to those gained here, show that additional base pairs terminal to the intercalation and hence to the bulged site, stabilised the bulge.

Bibliography

Beck, J.L.; Colgrave, M.L. Ralph, S.R. and Sheil, M.M. (2001) *Mass Spectrometry Reviews*, **20(2)**, 61-87

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

Cantor, C.R. and Schimel, P.R. (1980) in *Biophysical Chemistry Part II: Techniques for the study of biological structure and function* W.H. Freeman & Co., NY, p523

Dodgson, J.B. and Wells, R.D. (1977) *Biochemistry* **16**, 2367-2374

Gu, F.; Xi, Z. and Goldberg, I.H. (2000) *Biochemistry* **39**, 4881-4891

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

Kappen, L.S. and Goldberg, I.H. (1993a) *Biochemistry* **32**, 13138-1345

Kappen, L.S. and Goldberg, I.H. (1993b) *Science* **261**, 1319-1321

Kappen, L.S. and Goldberg, I.H. (1997) *Biochemistry* **36**, 14861-14867

Kolodner, R.D. (1995) *Trends Biochem. Sci.* **20**, 397-401

Kolodner, R.D. (1996) *Genes Dev.* **10**, 1433-1442

Malkov, V.A.; Biswas, I.; Camerini-Otero, R.D. and Hsieh, P. (1997) *J. Biol. Chem.* **272**, 23811-23817

Nakatani, K.; Okamoto, A. and Saito, I. (1999) *Angew. Chem. Int. Ed.* **38**, 3378-3381

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

Stassinopoulos, A. and Goldberg, I.H. (1995) *Biochemistry* **34**, 5997-6002

Streisinger, G.; Okada, Y.; Emrich, J.N.; Tsugita, E. and Inouye, M. (1966) *Cold Spring Harbor Symp. Quant. Biol.* **31**, 77-84

Wallace, R.B.; Shaffer, J.; Murphy, R.F.; Bonner, J.; Hirose, T. and Itakura, K. (1979) *Nucleic Acids Res.* **6**, 3543-3557

White, S.A. and Draper, D.E. (1987) *Nucleic Acids Res.* **15**, 4049-4064

Woodson, S.A. and Crothers, D.M. (1988) *Biochemistry* **27**, 8904-8914

Wu, H.-N. and Uhlenbeck, O.C. (1987) *Biochemistry* **26**, 8221-8227

Xi, Z.; Mao, Q.K. and Goldberg, I.H. (1999) *Biochemistry* **38**, 4342-4354