University of Strathclyde

Strathclyde Institute of Pharmacy and Biomedical Sciences

Biophysical characterization of the processes that drive ligand association with the minor groove of DNA

Hasan Y. Alniss

A thesis presented in fulfilment of the requirements for the degree of Doctor of Philosophy

2010

Declaration

The copyright of this thesis belongs to the author under the terms of the United Kingdom Copyright Acts as qualified by University of Strathclyde Regulation 3.50. Due acknowledgement must always be made of the use of any material contained in, or derived from, this thesis.

This thesis is the result of the author′s original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Signed: **Date:**

Acknowledgements

I am particularly grateful to my supervisor Prof. Simon MacKay for giving me the opportunity to carry out this research and for his guidance and patience. I respect his dedication to science and I appreciate his positive attitude and words of encouragement and support that motivated me during this research.

Special thanks to Dr. John Parkinson for teaching me macromolecular NMR and being supportive throughout our fruitful collaboration. His advice gave more inspiration for the thesis presented here. I look to Dr. John Parkinson with great admiration and respect as a scientist and I am grateful to him for passing me on a lot of knowledege in NMR field.

I want to thank Dr. Nahoum Anthony for teaching me the use of the ITC instrument and the modelling programs used in this study and for his constant help.

I would like to thank Prof. Colin Suckling for welcoming me into his lab to carry out the synthetic work of the MGBs. I also thank Dr Abdawn Khalaf for providing the MGBs and for his insightful discussions on my synthetic work. I thank Dr. David Breen for his advice and help in the synthesis of alkene linked MGBs.

I owe a debt of gratitude to the Strathclyde Institute of Pharmacy and Biomedical Sciences, academic staff and technicians for the training I received and the invaluable time I spent at the University of Strathclyde, ʺthe place of useful learningʺ.

Thanks to all the PhD students on level three in SIPBS and level four in chemistry and to anyone whom I have forgotten and who deserves my thanks.

Finally, I want to thank my wonderful wife, Rana Barham, for her endless support and devotion and for making life easier during the course of my PhD study. To my parents, to whom this thesis is dedicated, goes my deepest respect and appreciation for their support, encouragement and countless self-sacrifices and for teaching me the value of learning from the very beginning.

Table of Contents

List of Tables

List of Figures

Abbreviations

A	Adenine
C	Cytosine
CD	Circular dichroism
$CDCl_3$	Deuteriarated chloroform
CDI	Carbonyldiimidazole
CG	Conjugated gradient minimisation algorithm
EDTA	Ethylenediaminetetraacetic acid
ESI-MS	Electrospray ionization mass spectrometry
DCM	Dichloromethane
DMF	*N*,*N*-Dimethylformamide
DMAP	Dimethylaminopropylamine
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
Dp	Dimethylaminopropylamine
FID	Free induction decay
fs	Femtosecond
G	Guanine
HBTU	*O*-Benzotriazole-*N*,*N*,*N*′,*N*′-tetramethyluroniumhexafluorophosphate
HCl	Hydrochloric acid
Hp	Hydroxy pyrrole
Hz	Hertz
Im	Imidazole
I.R.	Infrared
ITC	Isothermal titration calorimetry
kcal	kilocalory
M	Molar concentration
Me	Methyl
mg	Milligram
MHz	Megahertz

MGB	Minor groove binder
ml	Millilitre
mol	Mole
m.p.	Melting point
MS	Mass spectrometry
NMR	Nuclear magnetic resonance
ODNs	Oligodeoxynucleotides
PDB	Protein data bank
ppm	Part per million
ps	Picosecond
Pyr	*N*-Methylpyrrole
SD	Steepest descents
Spec.	Spectroscopy
T	Thymine
TMEDA	Tetramethylethylenediamine
T_3P	Propylphosphonic anhydride
TFA	Trifluoroacetic acid
THF	Tetrahydrofuran
TLC	Thin layer chromatography

NMR data

s = SINGLET
d = DOUBLET
t = TRIPLET
q = QUARTET
Q = QUINTET
S = SEPTET
m = MULTIPLET

List of publications

Ranking ligand affinity for the DNA minor groove by experiment and simulation

Wittayanarakul K, Anthony NG, Treesuwan W, Hannongbua S, Alniss H, Khalaf AI, Suckling CJ, Parkinson JA and MacKay SP (2010) Med. Chem. Lett., 1, 367-380.

A detailed binding free energy study of 2:1 ligand-DNA complex formation by experiment and simulation

Treesuwan W, Wittayanarakul K, Anthony NG, Huchet G, Alniss H, Hannongbua S, Khalaf AI, Suckling CJ, Parkinson JA and Mackay SP (2009) Phys Chem Chem Phys, 11, 10682-10693.

Structural and energetic experimental evidence for subtle sequence selection by thiazotropsin A in the DNA minor groove

Hasan Alniss, Nahoum G. Anthony, Abedawn I. Khalaf, Simon P. Mackay, Colin J. Suckling, Roger D. Waigh, Nial J. Wheate andJohn A. Parkinson (2010) Chemical Science, (In preparation).

Project Abstract

The cationic lexitropsins, which bind non-covalently to the minor groove of DNA, have shown therapeutic potential in the treatment of cancer, viral and bacterial diseases. Understanding the factors that drive ligand-DNA associations, particularly the structural features, molecular forces and the energetics that dictate the overall binding process is of fundamental scientific interest as well as a prerequisite for the rational design and development of novel drugs. In this study, a holistic approach was followed to tackle this issue by combining thermodynamic and structural studies to gain insight into the factors that drive lexitropsin-DNA interactions and thus to allow potential ligands to be developed based on a rational approach.

The interaction of lexitropsins with different ODN sequences were studied using isothermal titration calorimetry (ITC), circular dichroism (CD), NMR spectroscopy and molecular modeling. ITC and CD experiments were used to obtain a complete thermodynamic profile for lexitropsin-DNA complexes and that included the determination of the binding affinity (K), stoichiometry (N), enthalpy (ΔH), entropy (ΔS), heat capacity (ΔC_p) and free energy of binding (ΔG) for the interaction. ITC studies showed that the lexitropsin-DNA interactions are mainly enthalpically driven via hydrogen bonding and van der Waals forces, and the unfavourable entropies associated with these interactions are indicative of “induced fit” binding. The distinct thermodynamic signature of these interactions allowed the differentiation between the molecular forces responsible for the binding. Furthermore, the thermodynamic binding characteristics of closely related ligand structures to a specific binding site helped to establish how modifications in the structure influence binding affinity.

NMR spectroscopy and restrained molecular dynamics simulations were used to obtain structural details for the interaction of two lexitropsin ligands, thiazotropsin B and AIK18-51, with the decamers d(5′-CGACTAGTCG-3′)$_2$ and d(5′-CGACGCGTCG-3′)$_2$, respectively. The NMR NOE derived inter-proton distances were used to generate three dimensional structures for these complexes. The location

of the binding site was determined by measuring the changes in chemical shifts of DNA protons upon ligand binding. NOE connectivities observed between the protons on the ligand and the protons on the DNA also provided more detailed information on the location of the binding site and allowed resolving the structural conformation of the ligand dimers. NMR and modeling results revealed that the two lexitropsin molecules bind to the minor groove as dimers in an anti-parallel side-by-side fashion and induced large perturbations in the DNA grooves.

Finally, in order to investigate the role of the traditional amide links of these ligands in the binding to the DNA minor grooves, and to evaluate their importance in forming hydrogen bonds with DNA bases, an analogue with an alkene in place of an amide was synthesized. Thermodynamic evaluation by ITC showed that the alkene-linked ligand did not bind to DNA.

1 CHAPTER 1: Introduction

1.1 DNA structure and function

Deoxyribonucleic acid (DNA) is the genetic material of all living cells. The secondary structure of DNA was solved by Watson and Crick in 1953,[4] and was one of the most important events in the history of science. The double helix of DNA is composed of two complementary, antiparallel polynucleotide strands paired together by specific hydrogen bonding interactions between nucleotide bases. Chemically, DNA is a polymer of nucleotides which contains a pentose sugar (deoxyribose), a phosphate residue, and one of the heterocyclic bases, adenine (A), guanine (G), thymine (T), or cytocine (C). The bases are attached to the sugar rings by a glycosidic bond while the phosphate and sugar moieties are joined by ester bonds to make the backbone of DNA, which forms the helical grooves (minor and major grooves) within which the edges of DNA bases are exposed[1] (Fig.1.1.b). The two strands are held together by interstrand hydrogen bonding; A pairs with T (2 H-bonds), and G pairs with C (3 H-bonds) (Figure 1.1 C). The dominant form of DNA in solution (B-DNA) exists as a right-handed helix and is characterized by a shallow, wide major groove and a deep, narrow minor groove.[5] The chemical and structural properties of both the minor and major grooves are characteristic of any DNA sequence, which forms the basis of DNA molecular recognition by small molecules and proteins.[5]

Three main types of conformations are known in DNA: A-form, B-form and Z-form (Figure1.2). The B-form was the basis for the model proposed by Watson and Crick. The A-form is right handed and characterized by a shallow, wide minor groove and narrow, deep major groove. The Z-form is a left handed double helix and has a zigzag backbone shape. Contrary to A- and B-DNA, the major and minor grooves of Z-DNA show little difference in width. Supercoiling in DNA can lead to the formation of tertiary structures as was observed in *Escherichia coli* (*E. coli.*), whose circular DNA contains about 4 million base pairs. In many instances, palindromic sequences in DNA and RNA lead to the formation of other secondary and tertiary structures such

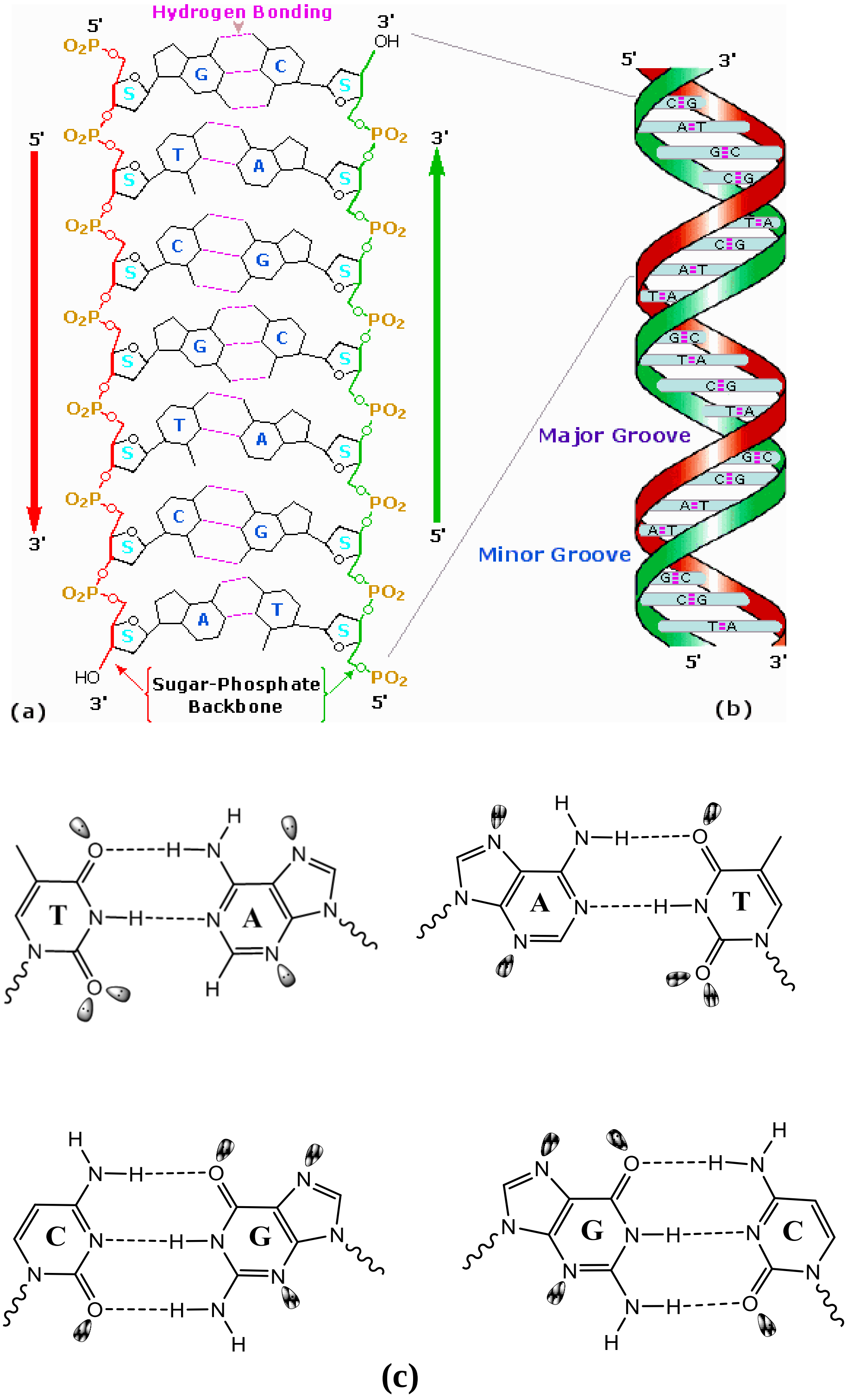


Figure 1.1 The double helix structure for DNA.[6] **(A)** Two strands of DNA are aligned anti-parallel to each other (opposite 3′ and 5′ ends); **(B)** the minor and major grooves of DNA; **(C)** the four arrangements of Watson-Crick base pairings.

as cruciforms, multistrands, hairpins, bulges and loops,[1] some of which are illustrated in Figure 1.3.

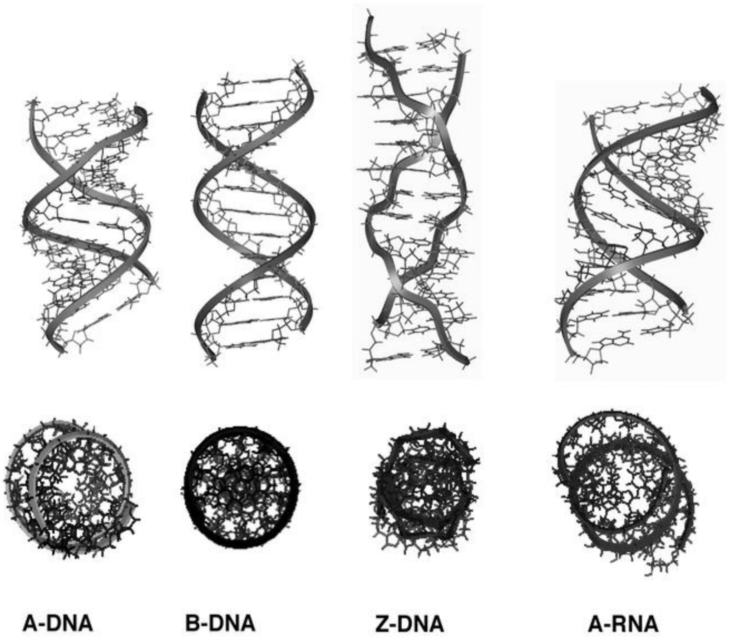


Figure 1.2 Molecular modeling representations of the major nucleic acid duplex conformations.The sugar/phosphate backbone is represented by a ribbon. Bottom views: orthogonal representations.[1]

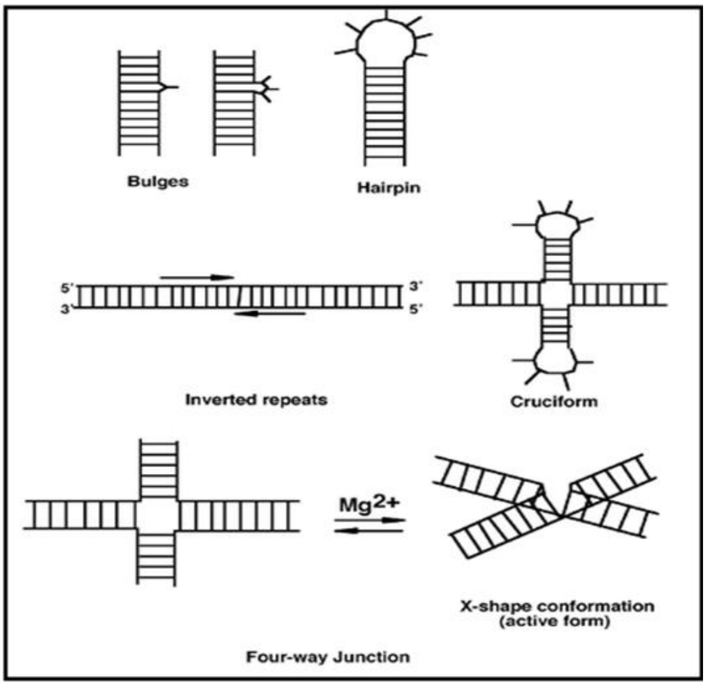


Figure 1.3 Simplified representation of DNA and RNA main secondary structures.[1]

In a eukaryotic cell, DNA is 2 m in length and tightly packaged into chromatin by hierarchical folding and compaction. Nucleosomes not only serve to compact DNA, but also affect accessibility of specific sequences by having interactions with the molecular machinery in the nucleus. The structure of the DNA-histone complex is dynamic, and the ability of DNA recognition in chromatin is affected by the structure of nucleosomal DNA.[7] To access the sites that are blocked by the histone octamer, DNA binding molecules need to have high binding affinity.

The sequence of bases on one of the strands is the genetic code. DNA carries the genetic information for proteins and via RNA (ribonucleic acid) converts this information via transcription and translation during protein synthesis. Proteins are the direct products of genes and also carry out nearly all cellular functions. Proteins interact with other proteins and biomolecules such as RNA, DNA, polysaccharides, phospholipids and different ligands to conduct the vital functions of life such as growth, development, repair and reproduction.[8] Changes in a protein (either post-translational modifications or induced upon such interactions) may alter its function.[9] Dysfunction of the protein may lead to problems in macromolecular recognition, cell cycle regulation, and lead to diseases such as cancer, cell invasion and amyloidoses.[10]

1.2 New approaches to drug design

The different functions of nucleic acids and proteins involved in many biological processes make them potential targets in modern drug design. A lack of detailed understanding of cell biology in the past meant that early approaches to drug discovery and design targeted the pathogenic cells as a whole, with only a few accidental discoveries of naturally-occurring drugs like penicillin by Alexander Fleming in 1929 that provided some insight into the mechanism of antibiotics.[11] It was not until advances in biochemistry and biochemical techniques in the mid-20th century that more directed targeting of enzymes and receptors became more common e.g. sulpha drugs/sulphonamides.[12]

It was not long after the double helical structure of DNA was solved sixty years ago that the design of drugs targeting DNA was initiated to combat diseases linked with its replication such as cancer and viral infections. Different types of antitumour agents have been developed that can interact with DNA in different ways such as non-covalent interaction by minor groove binders (e.g. distamycin A and Hoechst 22358) or by intercalation (e.g. daunomycin and ethidium bromide); by covalent bonding such as mitomycin and cisplatin; and by DNA backbone cleavage (e.g. duocarmycin, pepleomycin).[13] Many of these compounds show sequence selectivity, which makes them potentially useful as targeting agents for DNA, for instance, many minor groove binders prefer to bind to AT-rich sequences while intercalators prefer GC-rich sequences.[13-14] Although drug candidates can be developed to target RNA and proteins, DNA is considered more tractable as only one molecule of the drug target is required. Despite this advantage, because of the fundamental role that DNA plays in cell function and growth, DNA binding drugs in general have a major limitation associated with non-specificity, which can result in broad cytotoxic effects.

The rational approach to drug discovery and design is based on the understanding and examination of the three-dimensional receptor structure to reveal potential binding sites for drug molecules. Structural information and modeling data can be used to design or refine a ligand that fits within the binding site, like a key in a lock. Such information about the receptor structure can significantly improve the chances of obtaining a successful hit, and can help eliminate unlikely structures at an early stage of the drug discovery process. In the 1980s, modern drug discovery was heavily influenced by the concept of rational drug design.[15]

Combinatorial chemistry is a relatively recent development (in the early 1990s) that involves automated, high volume synthesis and characterisation of potential new hits for the drug discovery pipeline. In this technique, many molecules are synthesised and screened using high throughput screening (HTS) to find a small number of viable hits that interfere with a disease-related biological pathway in cells by targeting either proteins or DNA, usually in an *in vitro* assay.[12]

In the mid-1990s, novel approaches to discovering hits using 3D database searching were developed, which involved the use of a *virtual library* of computational descriptions of small molecules that any given combinatorial reaction scheme can produce. The standard techniques used for 3D structure determination of the biomolecule targets are X-ray crystallography and NMR spectroscopy. The 3D structure of the drug binding site can be used to select those molecules that best satisfy the restraints of the pharmacophore, and these molecules can be tested *in vitro* and submitted for biological assay, quite often in HTS format.

To effectively target biomolecules, it is necessary to understand the role they play in a disease related pathway. Most nucleic acids and proteins function in complicated biological interlinked processes with other molecules, often as part of bigger complexes. Genomes of pathogens (bacteria and viruses) and most recently, the human genome, have been sequenced using high-throughput DNA sequencing methods.[16] This increase in genome sequence data has generated an opportunity to study the functional units of related cellular processes, in a field often referred to as functional genomics. Four levels of analysis are possible: genes (the genome), messenger RNA (the transcriptome), proteins (the proteome) and the metabolites (the metabolome).[17]

The total number of genes is relatively constant between species. For example, the total number of genes in humans is similar to the number of genes in the nematode worm *Caenohabditis elegans.*[18] This suggests that complexity in an organism arises largely from differences in the gene products, their expression and post-tranlational modifications. This in itself is dependent upon the physiological and developmental state of the species and thus cannot be predicted from the gene sequence alone.

The proteome is the entire complement of expressed proteins from a cell at a given time in a given environment, and proteomics is the characterisation of these proteins. Proteomics is more complicated compared with other ″omes″ because proteins may undergo different modifications and behave differently in different conditions, acting by themselves or in a complex of proteins and other biomolecules. Proteomics requires both new, more powerful analytical techniques and more work in well-defined ″model organisms″.

The accumulation and consolidation of these approaches will enable new targets for modern drug design to be identified, and enzymes or receptors that function abnormally in diseased states will be the main targets. Gene therapy offers an alternative approach, where a malfunctioning receptor or enzyme (or gene) can be replaced by incorporating agents into the body that would allow the host to make its own functional protein. The use of these ″biotech″ drugs is, currently, an expanding world market.[19]

1.3 DNA binding molecules as potential therapeutics and tools for molecular biology

The ability of certain molecules to interact with DNA is of considerable interest, not only because of their potential as tools to control gene expression in molecular biology, but also for their therapeutic use, for example in cancer treatment.[20-21] Genetic information is stored in DNA, transcribed to RNA and then translated to protein products (Figure 1.4). These protein products such as enzymes and receptors control cellular function. According to the ″central dogma of molecular biology″, nucleic acids alone can specify the sequence of protein products, while proteins cannot specify a particular nucleic acid or protein sequence.[22] Interference at the DNA level is therefore the highest level of target control and can provide efficient ways of switching on/off the synthesis of protein products of interest. The number of genes present in a cell is far fewer than the corresponding mRNAs or protein products, and therefore theoretically requires much smaller quantities of compound to target the DNA. For these reasons, DNA is considered a primary target for binding or chemical modification by several classes of molecules.

Regulatory proteins repress or stimulate the flow of genetic information through DNA or RNA[23] and control gene expression throughout the lifetime of a cell. Small molecules can also alter the activities of nucleic acids and regulate gene expression by disrupting these regulatory protein-DNA interactions *in vivo*.[24-25] These molecules can be isolated from natural sources or synthetically prepared. The structural simplicity of such molecules, combined with facile synthesis, has accelerated work directed toward understanding the relationships between molecular structure and DNA recognition.[25-26] By investigating novel synthetic DNA binding molecules, we can expand our ability to probe DNA and develop new approaches for DNA-targeting chemotherapeutics.

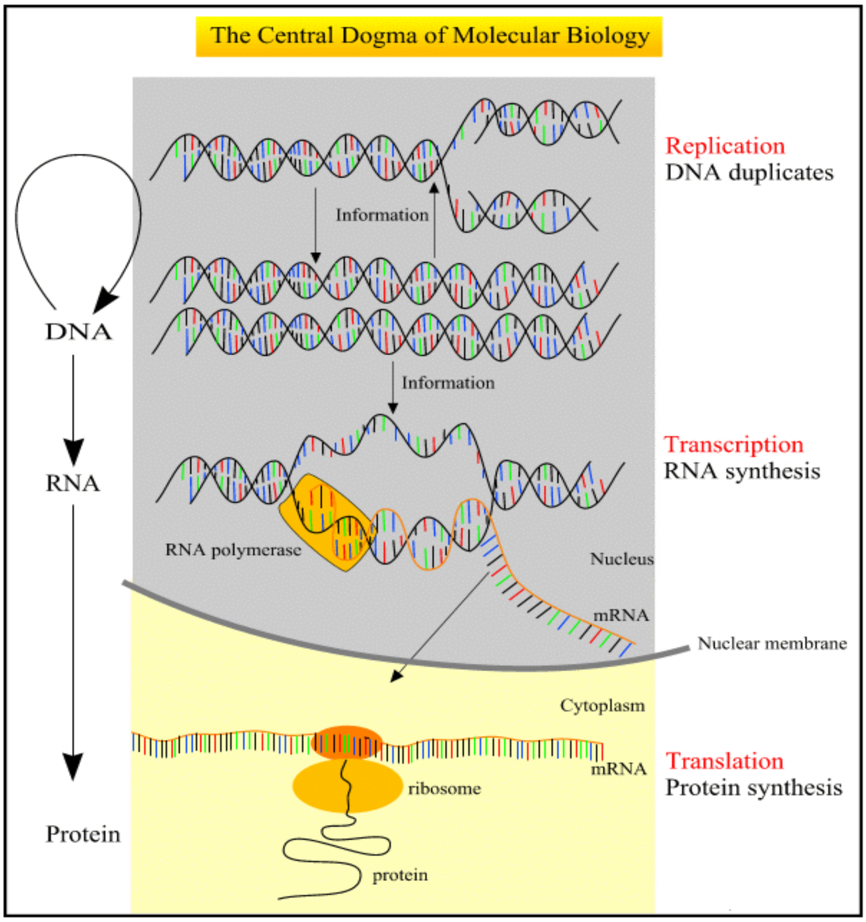


Figure 1.4 Illustration of the "Central Dogma of Molecular Biology".[27]

1.4 Drug- DNA Complexes

Classical DNA binding drugs specifically involve duplex DNA and involve one of three modes of DNA recognition; (1) binding in the DNA grooves; (2) direct bond formation with the DNA bases; and (3) intercalation between DNA base pairs. More recently, new modes of targeting non-canonical DNA structures have been identified that involve DNA tertiary structures such as G-quadruplexes and three-way junctions,[28] which can be stabilised by small molecules. The ideal anticancer agent would discriminate between normal and cancer cells, and this is not easily achieved with DNA-interactive drugs. Nevertheless, DNA is still considered a viable molecular target, and all the above modes of targeting DNA are still being pursued in order to develop anticancer agents with increased specificity and reduced cytotoxicity.

1.4.1 Alkylating agents

The first kind of DNA interacting compounds that were identified were the alkylating agents, which form direct bonds with the DNA bases (Figure 1.5A). Monoalkylating drugs such as temozolomide interact with a single DNA strand,[29] whrereas other agents can crosslink DNA strands. Examples of these are the clinically used drugs cyclophosphomide and mephalan, which crosslink the two complementary DNA strands,[30] and the platinum drugs, which result in intrastrand crosslinking.

1.4.2 DNA intercalators

Another common mode of DNA-drug binding is intercalation (Figure 1.5B). Intercalators contain polycyclic aromatic rings, which yield favorable stacking and hydrophobic interactions with the DNA base pairs.[31] Intercalation typically results in DNA structural distortion that eventually leads to cell death. Intercalating molecules generally interact with 2-4 base pairs and it is unlikely that such interactions will have enough sequence specificity to be selectively toxic in tumor cells. Attempts have been made to overcome these limitations by combining intercalation with other modes of DNA recognition. For example, the combined effect of a polyaromatic (intercalator) and cyclic peptide (minor groove recognition) moieties imparts higher sequence specificity to actinomycin D. Other DNA intercalators such as doxorubicin disrupt specific protein-DNA complexes such as topoisomerase enzymes that target a transient kinetic intermediate during DNA replication. Both topoisomerase I and II are involved in maintaining DNA structure and superhelicity via the formation of transient breaks in DNA. The most common inhibitors of these enzymes intercalate with DNA within the covalent binary protein-DNA complexes, eventually resulting in programmed cell death. This class of DNA-interacting agents work only when the macromolecular complex is assembled,[32] and sequence specificity of the topoisomerase enzymes facilitates the sequence specificity of the drug. This type of interaction presages a novel approach to the development of sequence-specific DNA binding agents that can repress gene transcription at selected genomic loci, and has much potential in the design of agents that target the unusual DNA structures that are a common occurrence in cancerous cells.

1.4.3 Code-reading molecules

This class of DNA-interacting drugs recognises the DNA grooves and generally has more sequence selectivity than alkylating and intercalating agents (Figure 1.5C). The discovery of minor groove recognition molecules such as distamycin and netrospin (Figure 1.6) paved the way for the design and synthesis of more specific code-reading molecules, and currently compounds that specifically target either the minor groove or the major groove of DNA are being developed.[33] Important candidates within this class are the pyrrole-imidazole polyamides that bind in the DNA minor groove in a sequence-specific manner, and can distinguish all four Watson-Crick base pairs. Such groove binding results in inhibition of protein binding to DNA, and blocks several different classes of eukaryotic transcription factors from binding to their cognate DNA sites.[34-36] These polyamides have recently been further modified to increase their specificity by pairing with amino acid moieties such as an Arg-Pro-Arg positive patch, which competes with protein side chains for DNA backbone phosphate contacts.[37] Another class of code-reading molecules under investagations are oligonucleotides, which bind in the major groove and form sequence-specific triple helices with the target DNA. The main drawback with these groove-binding agents to date, despite offering great potential as sequence-specific DNA targeting molecules, is their large size and the limitations this imposes on their druggability and delivery.

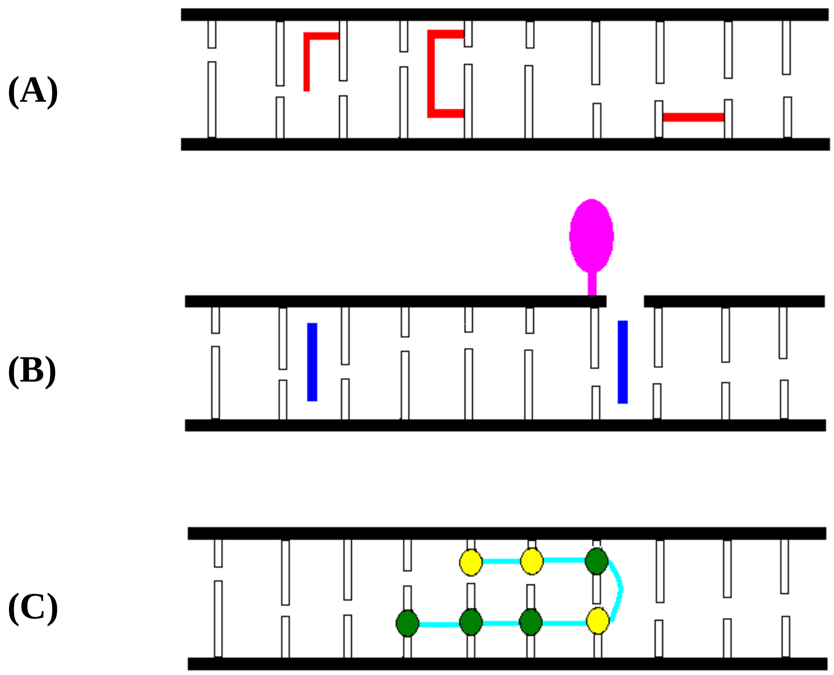


Figure 1.5 DNA-interaction modes of different DNA-interactive agents. **(A)** Alkylating agents **(B)** DNA intercalators with and without DNA interacting proteins and **(C)** code-reading molecules.

In Strathclyde, a library of pyrrole-imidazole/thiazole polyamide analogues of distamycin has been synthesised. This approach has aimed to design new, sequence-specific DNA binding agents with improved physical properties by enhancing their lipophilicity via the introduction of alkyl groups at different positions and replacing imidazole rings with the more lipophilic thiazole to enhance their chances of becoming commercial therapeutic drugs. These compounds show altered patterns of binding selectivity when compared with distamycin; for example, thiazotropsin A binds in the DNA minor groove at the sequence 5´-ACTAGT-3´ in a 2:1 ratio.[38]

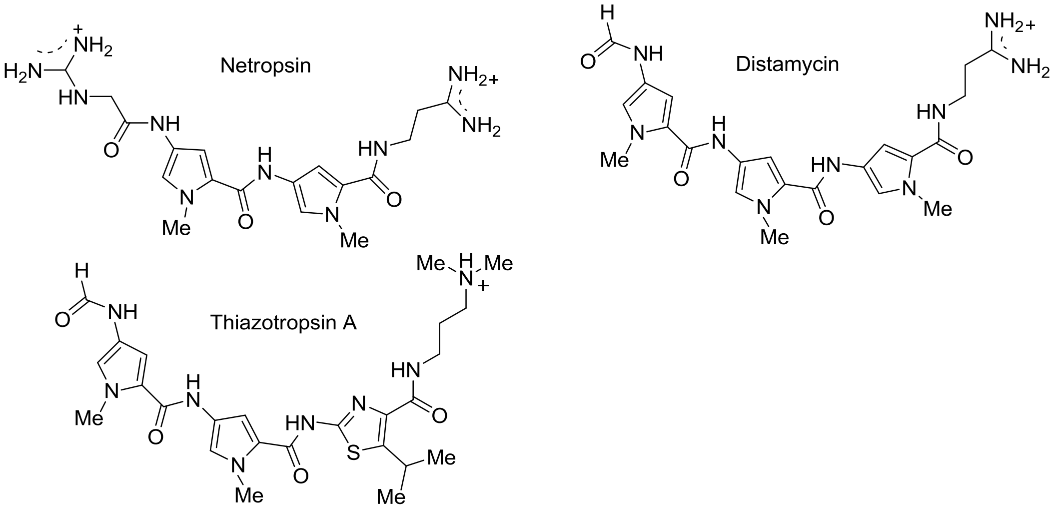


Figure 1.6 Structure of netropsin, distamycin and thiazotropsin A.

1.4.3.1 Structure of distamycin and it analogues

The naturally occurring heterocyclic polyamides distamycin and netropsin (Figure 1.6) were isolated from *Streptomyces netropsis* and *Streptomyces distallicus*, respectively. These compounds are toxic and bind to the minor groove of AT-rich sequences of double stranded DNA.[7,39] The general structure of these compounds (Figure 1.6) is composed of a head group, a series of heterocycles with amide links between each and a tail group such as an amidine or a trialkylamine[40] also linked by amides. One important feature in the structure of these molecules is their natural isohelical curvature which helps them to match that of the minor groove.

The selectivity of these molecules to AT-rich sequences is dictated by a series of hydrogen bonds formed between the NH′s of the amide links and the terminal amidinium groups in distamycin and netropsin which are pointed toward the N-3 of adenines and the O-2 of thymines in A•T on the floor of the minor groove. The backbone of DNA makes it a polyanionic molecule, and the positively charged head and/or tail groups on netropsin and distamycin will complement these negative charges. The distribution of electrostatic potential within the DNA grooves shows that A•T regions have a greater negative potential than the G•C regions,[39] which promotes distamycin and netropsin binding in these regions. Hydrogen bonding is also favoured in A•T regions as the topology of the groove can be considered a smooth curve. In regions containing G•C base pairs, the exocyclic N-2 protrudes into the groove and leads to a steric clash with the C-3 protons of the pyrrole rings in these agents, and lowers affinity by preventing hydrogen bonds from being formed effectively.

Several studies have been carried out[7,41-42] to solve the structure of netropsin complexed to DNA and all have shown essentially the same pattern of bifurcated hydrogen bonds (Figure 1.7). Netropsin binds exclusively in a 1:1 complex with DNA,[7,39] almost certainly due to the repulsive force which would occur by having two positively charged groups side by side. Distamycin however, whilst binding in a 1:1 complex in a similar way to netropsin, also has the ability to bind in a 2:1 complex with DNA as an anti-parallel dimer, which is only possible because the positively charged tails are well separated (Figure 1.8). This has been shown not to significantly distort the overall DNA structure but widens the groove width by 2 Å at the binding site.[39]

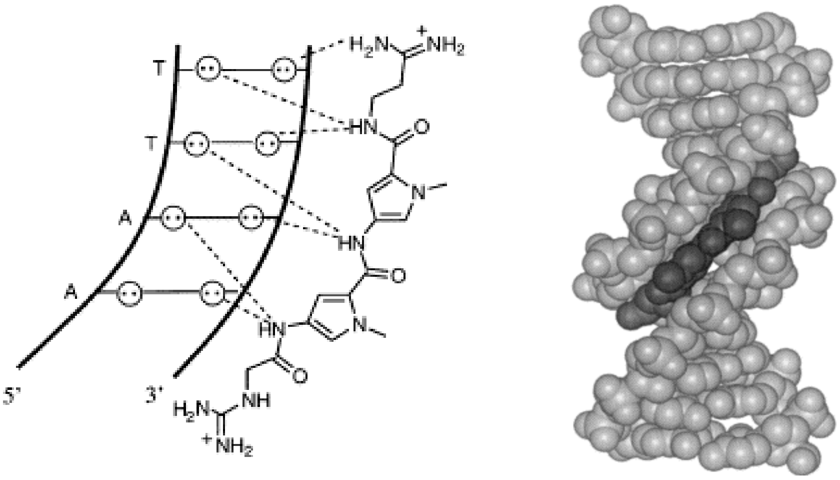


Figure 1.7 [7] Netropsin bound in the minor groove 1:1.

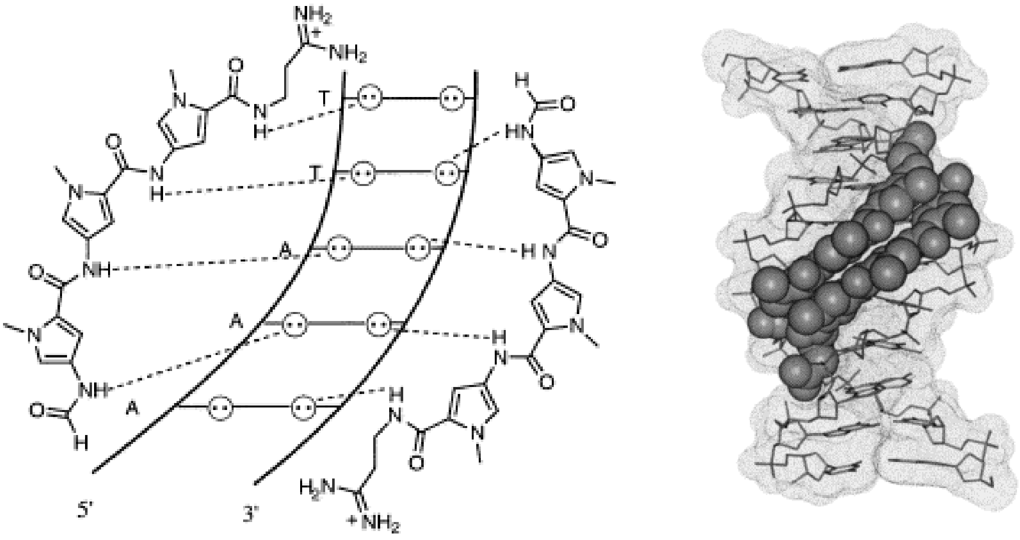


Figure 1.8 [7] Distamycin bound in the minor groove 2:1.

The ability of some MGBs such as distamycin to form 2:1 complexes (ligand to DNA) with each molecule reading a single strand of DNA, allows for more subtle discrimination between the DNA bases and the possibility of distinguishing G•C from C•G and A•T from T•A base pairs. This work, largely carried out by Dervan *et al*, led to a set of pairing rules[7,26,39] (Table 1.1) for all four base pairs by compounds binding in a 2:1 motif. Such discrimination between different base pairs required the introduction of new heterocycles: the recognition of G•C was achieved by introducing *N*-methylimidazole to remove the steric clash with the N-2 of guanine whilst simultaneously introducing a new hydrogen bond between the two moieties[7] (Figure 1.9).

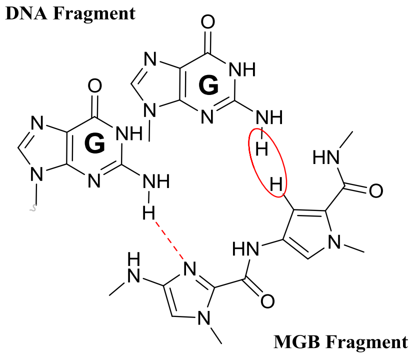


Figure 1.9 Illustration of imidazole binding to a G of a G•C base pair, and the steric clash which results with pyrrole.

Distinguishing T•A from A•T base pairs was achieved by introducing 3-hydroxy-*N*-methylpyrrole, which forms a specific hydrogen bond between the 3-OH group and the O-2 of thymine[26] (See Table 1.1 and Figure 1.11).

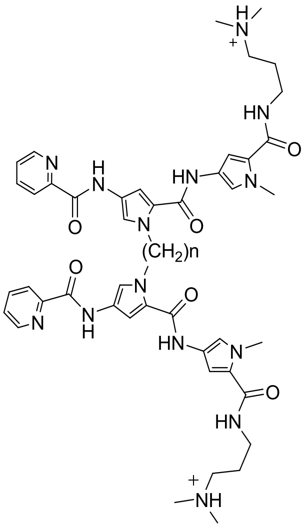
The application of these pairing rules and the 2:1 motif has led to increased specificity and binding affinity of these molecules. More specific code-reading molecules have been achieved by: (1) adding more heterocycles to extend the length, and therefore increase the reading frame; (2) using internal β-alanine residues, which are intrinsically more flexible than aromatic rings and relaxes the curvature of these ligands to maintain isohelicity with longer DNA sequences; (3) using cross-linked dimers, hairpin-linked dimers and cyclic dimers to reduce the number of possible conformations compared to an unlinked 2:1 motif, in order to increase the DNA binding affinity and specificity. Some structures of these compounds are shown in Figure 1.10.

Table 1.1[7] Heterocycles used for specific base pair recognition; X indicates which heterocyclic pairs bind to which base pairs. Im - *N*-methylImidazole, Py - *N*-methylpyrrole, Hp - 3-hydroxy – *N*-methylpyrrole.

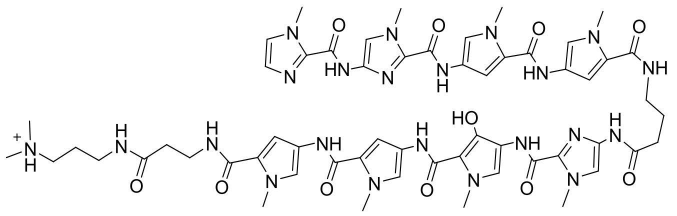
Pair	G•C	C•G	T•A	A•T
Im/Py	X			
Py/Im		X		
Hp/Py			X	
Py/Hp				X

These ligands exploit the same hydrogen bonding and van der Waals interactions to recognise the minor groove of DNA and exhibit the same orientational preference as the unlinked dimers with respect to the 5′→3′ direction of the adjacent DNA strand (Figure 1.11).

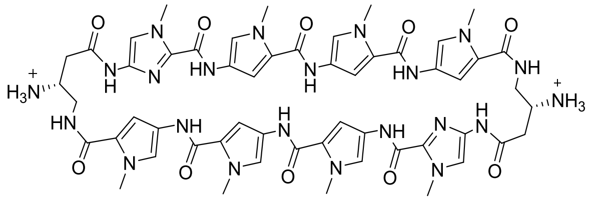
Although such molecules have been shown to interfere with gene expression *in vitro*,[43] these ligands suffer from poor cellular uptake as their molecular weights are often high (>1,000) breaking at least one of Lipinski′s ″rule of five″ for drug-like character, and suggests these compounds would suffer from poor membrane permeability.



Cross-linked dimer[39]



Hairpin dimer[7]



Cyclic dimer[44]

Figure 1.10 Structures of cross-linked, hairpin and cyclic dimers.

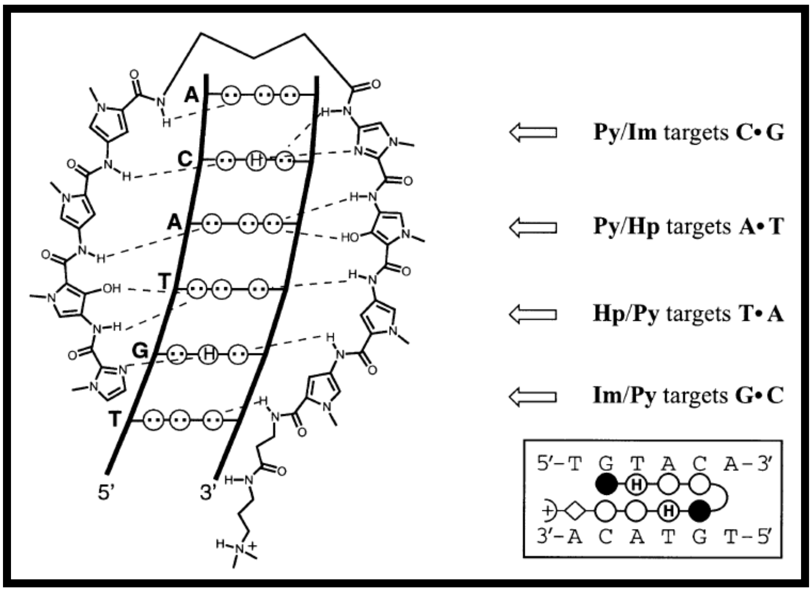


Figure 1.11[7] Minor groove hydrogen bonding patterns of Watson–Crick base pairs. Circles with dots represent long pairs of N(3) and O(2) of pyrimidines, and circles containing an H represent the 2-amino group of guanine. Hydrogen bonds are shown as dashed lines.

1.5 Molecular forces involved in ligand-DNA interactions

Understanding the molecular forces that drive ligand-DNA interactions is of prime importance for the rational design and development of novel drugs that can bind to nucleic acids specifically. Moreover, identifying the forces involved in such interactions is fundamental to solving what drives molecular recognition in general and ligand-DNA binding in particular. The forces that are known to contribute to ligand-DNA associations are hydrogen bonding, the hydrophobic effect, electrostatic and van der Waals/packing interactions. The binding process is also often associated with structural changes in the DNA and the ligand in order for the two to accommodate each other.

1.5.1 Hydrogen bonding

Hydrogen bonding is a directional attractive force between the hydrogen attached to an electronegative atom (hydrogen bond donor) with another electronegative atom (hydrogen bond acceptor).[45] Usually the electronegative atom is oxygen, nitrogen, or fluorine, which has a partial negative charge, while the hydrogen has a partial positive charge. Hydrogen bonding has a very important effect on the structure and chemistry of most biological molecules. Hydrogen bonds in water give it unique physical properties because it can act as both hydrogen bond donor and acceptor (Figure 1.12). Each water molecule is capable of donating two hydrogen atoms to form hydrogen bonds, while each oxygen atom is capable of accepting two hydrogen bonds.[46] Hydrogen bonding is also very important in nucleic acids and ligand–DNA complexes. The hydrogen bonding between the DNA base pairs holds the two strands of the double helix together and ensures that adenine pairs with thymine, and guanine with cytosine (Figure 1.1). The specific interactions between proteins and DNA are governed by hydrogen bonds between the base pairs and amino acid residues in the protein. The interaction of minor groove binders with the DNA duplex is stabilised by hydrogen bonding between the ligands and the base pair edges on the groove floors (Figure 1.7 and 1.8).[7] Hydrogen bonds via water molecules trapped in the binding interface are also commonly observed in ligand-macromolecule interactions.

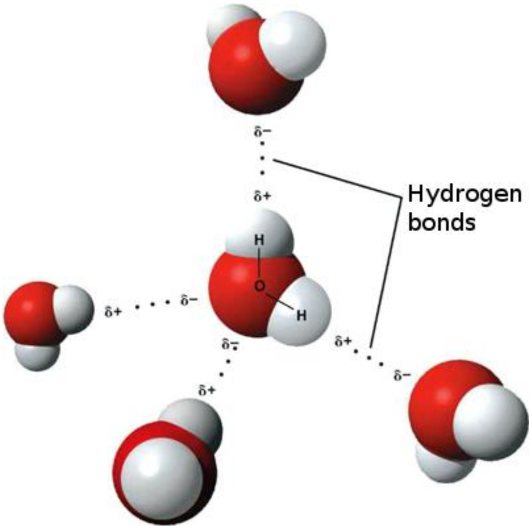


Figure 1.12 Hydrogen bonds between water molecules.[47]

Hydrogen bonds are directional, which means they are stronger when the hydrogen atom is aligned with the two electronegative atoms. Bent hydrogen bonds occur, but they usually have decreased stability. In addition to the bond angle, the strength of hydrogen bonds also depends on bond length, temperature, pressure and environment. The typical length of a hydrogen bond in water is 1.97 Å.[45] As the intermolecular distance between the hydrogen bond donor and acceptor increases, the strength of hydrogen bonding falls of rapidly; the strength of the interaction falls to negligible values at a distance of 5 Å.[45] Hydrogen bonds can vary in strength from very weak (1-2 kJ mol^{-1}) to extremely strong (>155 kJ mol^{-1}), as in the ion HF_2^-.

1.5.2 Hydrophobic interactions

The hydrophobic effect in ligand-DNA interactions can be defined as the removal of non-polar molecules from an aqueous environment into the binding site in order to reduce the unfavourable interactions between water and non-polar atoms. The hydrophobic effect happens mainly because non-polar atoms are unable to form hydrogen bonds with the polar solvent, and it is this inability to form hydrogen bonds with water, rather than the attractive forces between non-polar groups that is responsible for the hydrophobic effect.[48] Non-polar molecules aggregate together to minimise water-exposed accessible surface area, and water is forced to form structured cavities to accommodate these entities, which results in a loss of conformational entropy and an energetic penalty.[49] The hydrophobic effect is an entropy-driven process, which allows reducing the free energy of a system by minimising the surface interface between non-polar molecules and aqueous solvent. This process is energetically favourable because the entropic cost of separating water and non-polar molecules is smaller than the entropic cost of ordering water molecules around large hydrophobic-water interfaces, which can lead to mixed non-polar solutes with water.[49] However, if clusters do assemble, the hydrophobic molecules are held together by weak van der Waals interactions, especially London dispersion forces.[50] Overall, there is a drive to move hydrophobic molecules from an aqueous environment to a less polar binding site in a macromolecule.

1.5.3 Electrostatic interactions

Electrostatic effects have long been recognised as one of the most important non-covalent attractive forces between charged molecules. Electrostatic interactions between protonated ligands and the negatively charged DNA backbone play an important role in stabilising ligand-DNA complexes. Detailed analysis of the electrostatic contribution to the free energy of binding has been accomplished for various ligand-DNA,[51-53] protein-DNA,[54] and protein–protein[55] complexes. The basic equation for all electrostatic interactions is Coulomb′s law (Eq. 1.1), which states that the potential energy between two charges is directly proportional to the magnitude of the charges and inversely proportional to the distance separating the two charges:

$$V = -\frac{Kq_1q_2}{r^2}$$

Eq.1.1

where *K* is a proportionality constant known as the Coulomb′s law constant, q_1 and q_2 are the magnitude of the charges, and *r is* the distance between the two charges.

1.5.4 van der Waals interactions

van der Waals interactions are weak attractive forces between non-polar molecules or atoms that differ from electrostatic and hydrogen bonding because they are caused by correlations in the fluctuating dipoles of nearby molecules or atoms. The Lennard-Jones potential[56] is a simple mathematical model that can be used to describe the interaction between a pair of neutral atoms or molecules by the following equation (Eq 1.2):

$$V(r) = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$$

Eq.1.2

where ε is the depth of the potential well, σ is the finite distance at which the inter-particle potential is zero, and r is the distance between the particles. The r^{-12} term describes Pauli repulsions at short ranges due to overlapping electron orbitals and the r^{-6} term describes attraction at long ranges (van der Waals forces). These intermolecular attractive forces are much weaker than the covalent bonds within molecules. However, it is not possible to assign exact values as the strength of these forces is governed by both the size of the molecule and its shape. Non-polar molecules can exhibit London forces (induced dipole-induced dipole interactions) because the motion of electrons can result in a transient dipole moments and uneven distributions of electron density. When this uneven distribution occurs, a temporary multipole is generated which can interact with other nearby multipoles. Energetics of induced dipole-induced dipole interactions are similar to those of electrostatic interactions with the exception that the strength of the interaction falls off more rapidly with distance because they only involve transient partial charges.[57] Consequently, their effects are relevant to binding interactions over short distances of several Å.

1.6 Techniques used to study ligand-DNA associations

Four main methods have been used to study ligand-DNA complexes in this thesis. Isothermal titration calorimetry (ITC) and circular dichroism (CD) were used to obtain thermodynamic profiles for the binding interactions; NMR spectroscopy has been used to provide structural details of these complexes; and molecular dynamic simulation studies have been used to combine the thermodynamic profiles with the structural studies.

1.6.1 Isothermal titration calorimetry

Understanding the molecular basis of ligand–DNA-binding processes, particularly the structural features and forces that drive ligand-nucleic acid interaction, including sequence recognition, structural details and the thermodynamics of binding is a prerequisite for the rational design and developments of novel drugs. Thermodynamic evaluation, when used in conjunction with structure, sequence, and computational methods, can be used to design molecules to bind specific nucleic acid sequences and/or structures.

ITC is a well established biophysical technique that is used to directly determine the thermodynamics of intermolecular binding at constant temperature. It can provide a complete thermodynamic profile of biochemical equilibrium interactions and has become one of the fastest growing techniques used in biomedical research[58-60] since its emergence in 1989.[61] The principles and applications of ITC for studying the thermodynamics of biomolecular interactions have been comprehensively described in the literature.[3,62-63] The ITC instrument is composed of two identical cells surrounded by an adiabatic jacket (Figure 1.13). The sample cell is usually filled with a macromolecule solution (e.g. DNA, protein, etc) and a reference cell which is filled with buffer or water. Small volumes of ligand are titrated by a computer-controlled stirring syringe into the DNA solution in the sample cell. Alternatively, the ligand solution in the cell can be titrated with the solution of the macromolecule, which is recommended for poorly soluble compounds. If there is a binding interaction between the reactants, heat is either released or absorbed in direct proportion to the amount of binding that occurs. The instrument detects temperature differences (ΔT) between the reference and sample cell and maintains ΔT zero by increasing or decreasing the feedback power applied to the sample cell when the reaction is endothermic or exothermic, respectively.[64] When the macromolecule in the cell becomes saturated with added ligand, the heat signal diminishes until only the background heat of dilution is observed. Because the amount of free macromolecule available decreases after each successive injection, the intensity of the raw data peaks becomes progressively smaller until complete saturation is reached. Measurement of this heat change allows for the determination of binding constants and a thermodynamic profile of the reaction that includes the observed molar calorimetric enthalpy (ΔH_{obs}), entropy (ΔS_{obs}), and the change in free energy (ΔG).

The heat change per injection (*dQ)* for the binding of a ligand (L) to a macromolecule (D) to form the non-covalent complex (DL) is expressed as:

$$dQ = d[DL] \times \Delta H \times V$$

Eq.1.3

where V is the volume of the cell.[61] This change in heat can be differentiated with respect to the total accumulated ligand concentration in the cell after each injection to yield a binding isotherm that can be fitted to the enthalpogram.

The ability of ITC to provide a reliable value for the binding constant K depends on the c-value:[65]

$$c = K \times [\mathrm{D}] \times n$$

Eq.1.4

where c is a unitless parameter that determines the shape of the binding isotherm, K is the binding constant, [D] is the concentration of macromolecules in the cell and n is the stoichiometry of binding. This value may range between 1 and 1000; for very low values, the titration curves become shallow and non-descriptive, while for very high values, the transition region between ligand binding and saturation is empty of data points. Usually, binding constants between 10^3 and10^8 can be measured accurately.[66]

The enthalpy change accompanying a binding interaction can be measured either directly or indirectly. Indirect methods involve calculating the thermodynamic parameters from theoretical relationships, such as van't Hoff analysis, where enthalpy changes upon binding are estimated from the temperature dependence of the binding constant. Both direct and indirect methods, when put into practice carefully, have been shown to agree within statistically significant margins.[67] However, several measurements in a wide temperature range have to be performed, which is often not experimentally attainable and leads to large errors in the parameters resulting from van′t Hoff analysis.[68-69] Since ITC is the only method that can be used to give direct characterisation of the heat involved in any process,[70] it is an ideal method to obtain a complete thermodynamic profile of a binding interaction. The advantage of ITC over other methods that involve kinetic techniques is that immobilization or modification of reactants is not required since heat of binding is a naturally occurring phenomenon.[58-60] If an ITC experiment is designed carefully, it can provide in a single titration both the enthalpy and the binding constant from which the entropy and the Gibbs free energy of binding can be obtained. Moreover, the temperature dependence of the enthalpy of binding allows the calculation of another important thermodynamic parameter; the heat capacity change (ΔCp) of binding. ΔCp is calculated from the slope of ΔH vs. temperature and can be positive (hydrophobic interactions are disrupted upon binding) or negative (hydrophobic interactions are formed).

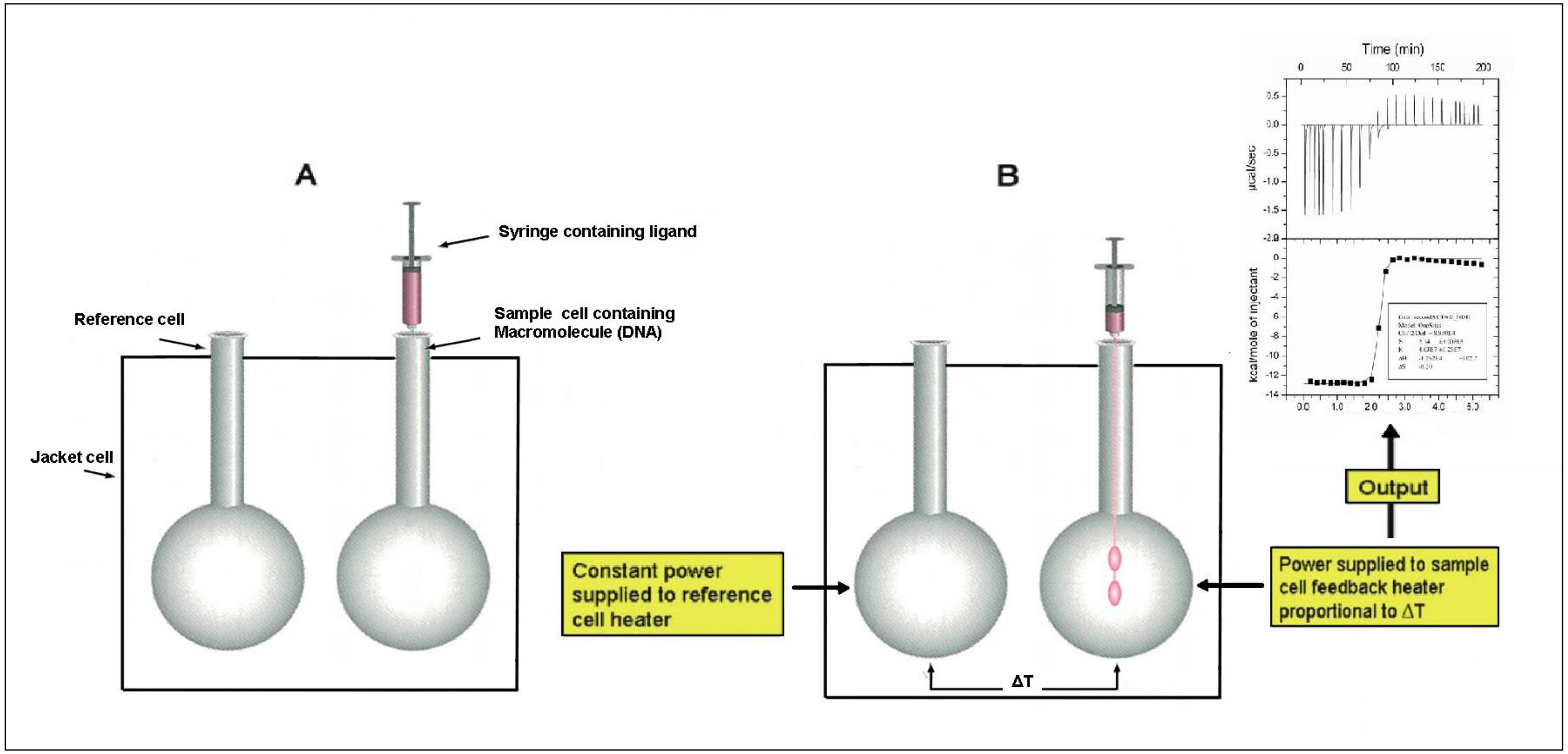


Figure 1.13 Schematic representations of isothermal titration calorimetry (ITC) instruments.[3] **(A)** An ITC instrument prior to performing a titration. The sample cell and the reference cell are kept at the same temperature, which is typically 25°C. The reference cell is always kept at the experimental temperature. The ligand is usually placed in the syringe and the macromolecule (DNA) in the sample cell. **(B)** An ITC instrument performing a titration. When an injection is made, heat is released or absorbed in direct proportion to the amount of binding (endothermic or exothermic) and this results in a change in temperature of the sample cell. A change in power (heat/s) is required to return the cells to identical temperatures (T) (i.e., $\Delta T = 0$). This change in power is recorded as a series of injections is made. In the raw data presented in the inset, each injection is accompanied by an interaction where heat is given out (exothermic). As more ligand is injected, the binding sites in the sample cell are gradually saturated, and the exothermic effect diminishes before new endothermic signals appears as a result of the heat of dilution of the ligand in the buffer.

1.6.1.1 The role of thermodynamics in ligand-DNA interactions

Ligand-DNA interactions are associated with a change in the free energy of the ligand-DNA complex compared with the free energy of both the ligand and the DNA existing freely in solution and can be represented by the following equation (1.5):

$$\boldsymbol{Ligand + DNA \overset{\Delta G}{\Leftrightarrow} \mathbf{Ligand \cdot DNA}}$$

Eq.1.5

The Gibbs free energy change (ΔG) for ligand-DNA interactions is related to the standard Gibbs free energy change (ΔG^{o}_{bind}), under standard conditions (at a temperature of 25 °C and a concentration of 1 M for both the DNA and ligand) and can be calculated using the following equation (1.6)*:*

$$\boldsymbol{\Delta G = \Delta G^{\circ}_{bind} - RTln\frac{[\mathbf{Ligand \cdot DNA}]}{[Ligand][DNA]}}$$

Eq.1.6

where R is the gas constant and T is the absolute temperature. At equilibrium and standard conditions, the ΔG^{o}_{bind} is zero, and thus equation 1.6 can be written as follows:

$$\boldsymbol{\Delta G = -RTlnK_b}$$

Eq.1.7

The difference in free energy of binding can also be described in terms of the enthalpy and entropy changes using equation 1.8:

$$\boldsymbol{\Delta G = \Delta H - T\Delta S}$$

Eq.1.8

where ΔH is the difference in enthalpy between the bound and unbound state and ΔS is the difference in entropy between the two states. These thermodynamic equations show that the tighter the binding of the ligand to its target, the greater is the difference in free energy between bound and unbound states with the bound state being the lower in free energy. Moreover, the differences in free energy of binding are governed by difference in both enthalpy (ΔH) and entropy (ΔS) between the bound and unbound states.

1.6.1.2 Enthalpy, entropy and drug design

Determination of the thermodynamic parameters that drive the binding of small organic molecules into the minor groove of DNA gives invaluable insight into the understanding of ligand–DNA interactions. The enthalpy changes (ΔH) relate directly to the heat of interaction between the ligand and its target in the bound and unbound states. This change in enthalpy reflects the total contribution from the formation or removal of non-covalent forces in the system upon binding.[71] In ligand-DNA interactions, hydrogen bonding and van der Waals forces are usually associated with an exothermic favourable (negative) enthalpy changes, while the hydrophobic and electrostatic interactions (polyelectrolyte effect), which are manifested entropically, are associated with either small positive or negative enthalpy and a favourable (positive) entropy.[72-73] For an exothermic binding interaction, the change in enthalpy is equal to the energy released upon binding, while for an endothermic reaction, the change in enthalpy is equal to the energy absorbed in the reaction and reflected by the sign of ΔH. If ΔH ($\Delta H = H_{product} - H_{reactant}$) is positive, the reaction is endothermic as heat is absorbed by the system. If ΔH is negative, the reaction is exothermic and the negative ΔH is due to the release of heat. A negative ΔH indicates enthalpically favoured binding.

The entropic term represents the change in the order of the system. This incorporates changes in conformational entropy (a gain/decrease in the conformational freedom of certain groups of the ligand or DNA) as well as desolvation entropy (the release of water and counter ions from the minor grooves upon ligand binding). A positive $T\Delta S$ results in entropically favoured binding- i.e. the system becomes more disordered (2nd law of thermodynamics). Favourable entropy changes are mainly caused by the hydrophobic effect, where an increase in solvent entropy is generated by the removal of non-polar hydrophobic groups from the aqueous environment and the release of water upon binding. The electrostatic interactions, for example those associated with the release of counterions from the backbone of DNA upon binding, also contribute to the positive $T\Delta S$.[72-73] The binding of a ligand to a macromolecule is usually associated with a decrease in conformational entropy as the degrees of freedom of both the ligand and the macromolecule are restricted upon binding. This decrease in entropy disfavours the interaction between the ligand and its target, but is often partially or completely offset by the release of water molecules from the hydrophobic surfaces of both the ligand and the binding site. Upon ligand binding to the macromolecule, both the ligand and the binding site are completely or partially desolvated. The hydrophobic effect is usually associated with a desolvation of the binding interface, whereas the hydrophilic interactions (i.e. hydrogen bonding), sometimes involve water molecules being trapped between the ligand and the binding site and help complex formation through a bridging effect.[74-77]

From the standpoint of rational drug design, the determination of these important thermodynamic parameters (ΔH & ΔS) for the binding of a ligand to its target can give a clearer understanding of the important attributes of binding. ITC is widely used nowadays in the field of drug design and optimisation to understand the principles of molecular recognition. The establishment of a link between structure and energy can lead to real progress in understanding the mechanisms of molecular associations and help reveal the important structural features that drive the binding event. This can be achieved by comparing the binding thermodynamic characteristics of closely related ligand structures to a specific binding site. Introducing a moiety with high enthalpic potential such as one capable of forming hydrogen bonds and van der Waals forces within the binding site is expected to improve binding. However, there are some drawbacks; enthalpy/entropy compensation is a ubiquitous phenomenon, which results from the nature of non-covalent interactions. Increases in bonding are often offset by an entropic penalty, reducing the magnitude of the change in affinity. For instance, the enthalpic gain from a hydrogen bond can be counteracted by a dehydration penalty via the burying of polar functional groups, or by conformational entropic losses from immobilisation of the constituent groups. Another problem is the flexibility of some binding sites especially protein pockets, which can become restricted on binding a ligand and result in a conformational entropic penalty.

An example that illustrates the obstacle which enthalpy/entropy compensation poses to the optimisation of binding affinity has been reported by Lafont *et al.*[78] The authors used the traditional method followed by medicinal chemists in the process of systematic optimisation of lead compounds. They tested how minor modifications in the structure of the HIV-1 protease inhibitors KNI-10033 and KNI-10075 (Figure 1.14) could affect the affinity for the target as a way to improve binding. The only difference between the two inhibitors was the replacement of the thioether group of KNI-10033 by a sulfonyl group in KNI-10075 with the aim of establishing a strong hydrogen bond with the enzyme. The crystal structure of the KNI-10075/protease complex proved the existence of the anticipated hydrogen bond with the amide of Asp 30, which improved the binding enthalpy by almost 4 $kcal\ mol^{-1}$. However, the gain in enthalpy was opposed by unfavorable entropy changes which negated an increase in binding affinity. Structural analysis suggested that the entropy loss was due to a decrease in both the desolvation and conformational entropy.

In the literature, there are many examples of using ITC to develop new drugs to maximise selectivity and specificity to the binding target by attempting to optimise enthalpy and entropy. ITC studies on HIV-1 protease inhibitors,[79-81] SARS drugs,[82] anti-malarial drugs,[83] and DNA gyrase inhibitors[84] have been used to obtain thermodynamic profiles of drug-target interactions. These data have given valuable information for the design of new drugs that bind to their targets with higher affinity and selectivity.

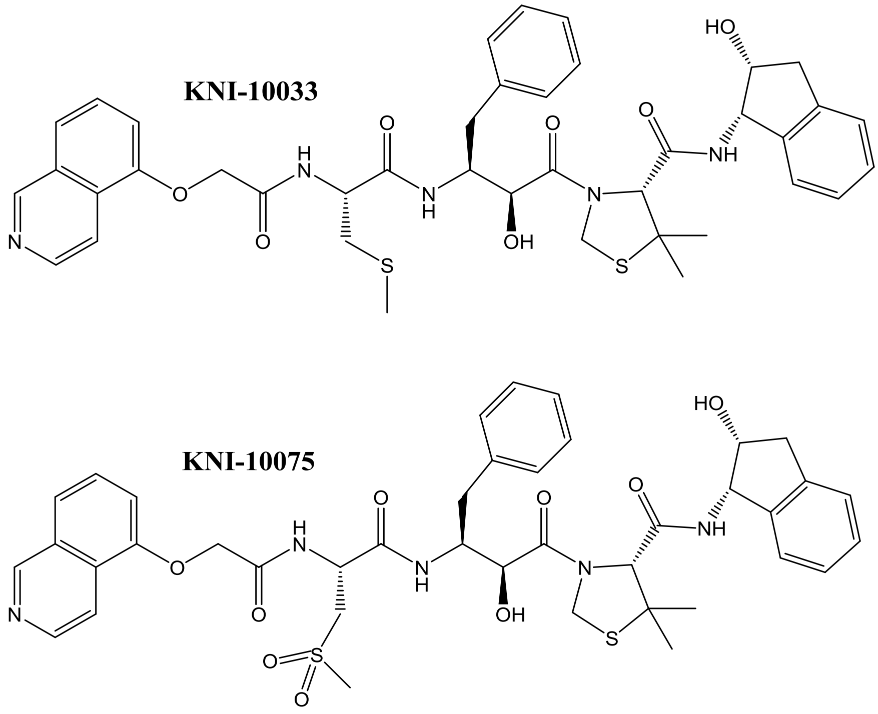


Figure 1.14 Structures of KNI-10033 and KNI-10075.[78]

1.6.1.3 Thermodynamics of ligands binding to the minor groove of DNA

The DNA duplex exhibits a fixed geometry and is much more rigid compared with protein-binding sites, which have arbitrary shapes and are flexible. The spatial position of hydrogen donor/acceptor functions, the negatively charged phosphate backbone of DNA and the overall van der Waals surface that can be deduced from sequence information are the main factors that drive minor groove recognition by small molecules. The thermodynamic characteristics of MGB binding have been the object of studies since the appearance of sensitive titration calorimeters. This requires not only the determination of the thermodynamic parameters (ΔG, ΔH, ΔS, K, ΔCp) of ligand–DNA interactions, but also linking these parameters to the various possible contributions from H-bonding, van der Waals forces, electrostatic interactions and hydration forces, etc.

MGBs have been extensively studied by several groups[72-73,85-87] using ITC. Table 1.2 summarises the thermodynamic profiles of the commonly studied MGBs. The published literature to date has revealed that minor groove recognition by small molecules can be driven by enthalpy, entropy, or by both. The thermodynamic signature of MGBs is highly dependent on the ligand structure, the binding site, and the molecular forces that drive the binding. MGBs are typically composed of several aromatic rings such as pyrrole, furan, benzene or benzimidazole that are connected by a bond with limited torsional freedom such as an amide or a conjugated biaryl bond (Figure 1.15). The amide linked MGBs (e.g. distamycin and netropsin) have a completely different thermodynamic profile compared with the biaryl- linked MGBs (e.g. Hoechst 33258).[73]

Amide linked minor groove binders (or ″lexitropsins″) can recognise DNA as a monomer (e.g. netropsin)[73] or a dimer (e.g. distamycin)[73,87] as previously described.

The monomeric association of the amide linked MGBs is mainly driven by favourable enthalpy and enhanced by positive entropy (e.g. netropsin and distamycin, 1:1 binding mode, Table1.2). This favourable enthalpy has been explained by the ability of the NH groups of amide links to form hydrogen bonds with the DNA bases on the groove floor in addition to van der Waals interactions between the aromatic rings of the ligand and the minor groove walls of DNA. The favourable entropy is likely to be due to the release of water from the minor grooves upon binding. There is also minor conformational disruption to the DNA helix with monomeric binding, which also has less of an impact on entropy. A similar thermodynamic signature was observed with the hairpin polyamide, ImPyPy-γ,-PyPyPy-β-Dp (where Im = imidazole, Py = pyrrole, γ = γ-aminobutyric acid, β= β-alanine, and Dp =

dimethylaminopropyl amide). Its binding into the minor groove is primarily enthalpically driven (73 %) with a small contribution from entropy.

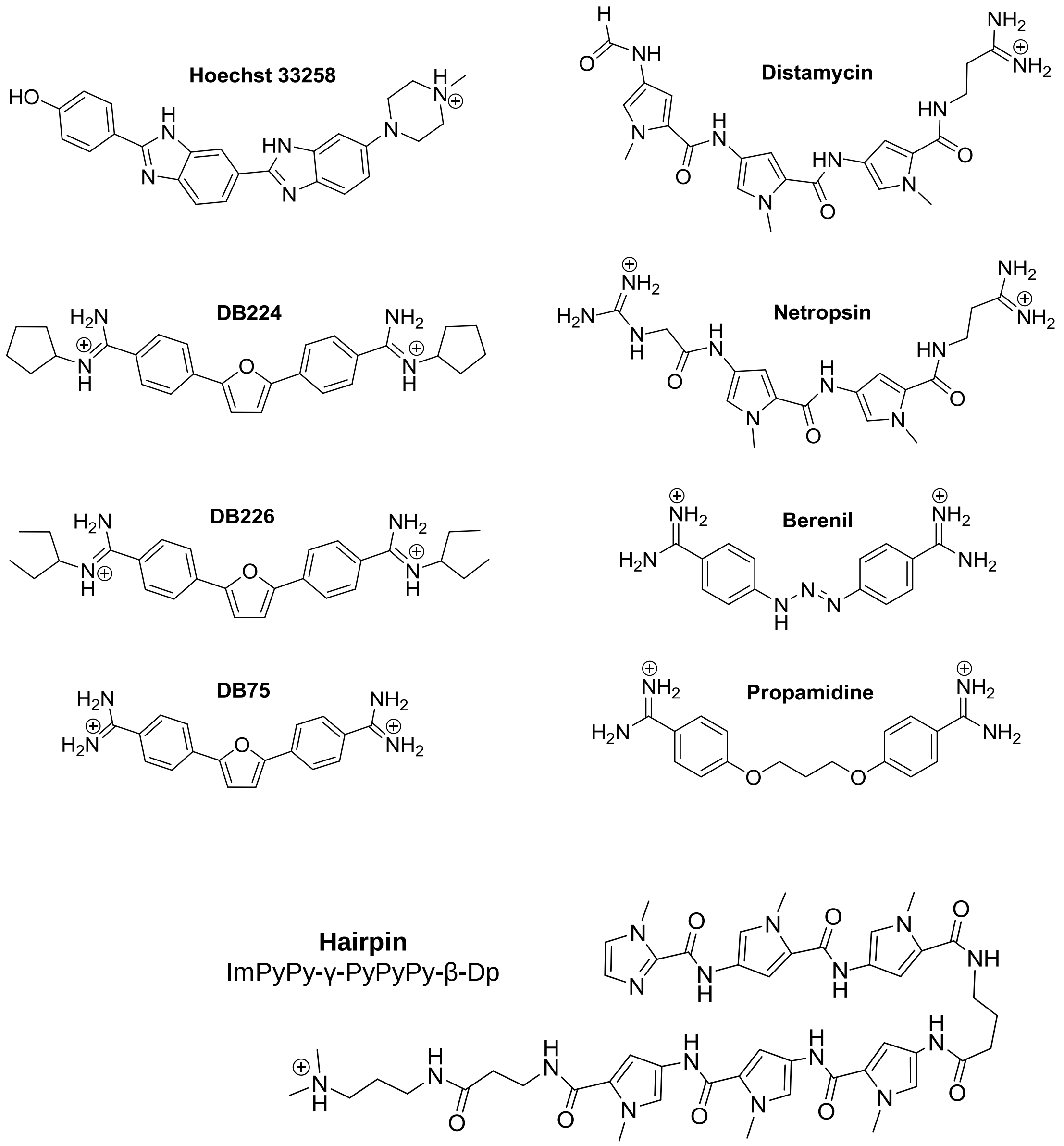


Figure 1.15 Structures of commonly studied minor groove binders.

Table 1.2 Thermodynamic data for commonly studied MGBs with DNA at 25 °**C.**

Compound	**Δ*G* (kcal mol^{-1})**	**Δ*H* (kcal mol^{-1})**	***T*Δ*S* (kcal mol^{-1})**	**Reference**
Distamycin 1:1	−10.5	−5.8	+4.7	73
Distamycin 2:1	-7.9	-15.7	-7.8	73
Distamycin 2:1	-10.5	-12.5	-2.0	87
Netropsin	−8.7	−5.8	+2.9	73
Hoechst 33258	−11.7	+4.3	+16.0	72
Hoechst 33258	−11.8	+10.0	+21.8	85
Berenil	−8.0	+0.6	+8.6	73
DB226	−8.5	−0.5	+8.0	86
DB244	−9.9	−2.3	+7.6	86
DB75	−9.0	−2.2	+6.8	86
Propamidine	−7.0	−1.1	+5.9	73
Hairpin	-9.2	-6.7	+2.5	88

The dimeric association of distamycin A with the DNA duplex is driven by enthalpy (ΔH -15.7[73]; -12.5[87] kcal/mol) and opposed by entropy ($-T\Delta S$ 7.8[73]; 2.0[87] kcal/mol). The favourable enthalpy is again due to the formation of hydrogen bonds and van der Waals interactions between the ligand and binding site. The unfavourable entropy means that the association induced relatively unfavourable conformational changes in either the ligand or DNA, and that the binding interface does not undergo desolvation upon binding. Binding into the minor groove as a dimer can potentially trap more water between the ligand and the binding site and facilitate complex formation by bridging through hydrogen bonding. This method of association would therefore not be accompanied by the release of water from the minor groove, which could explain the large favourable enthalpy and the unfavourable entropy.

The biaryl linked minor groove binders (e.g. Hoechst 33258, DB226, DB224 and DB75, Table 1.2) recognise the DNA as monomers and there are no examples of dimeric association per one site by this class of compounds to date. Their binding into the minor groove of DNA is driven by large favourable entropy changes with small contributions from enthalpy. The main forces that drive complex formation are the hydrophobic forces, in addition to the electrostatic interactions between the positively charged functional groups of the ligand and the negatively charged sugar-phosphate backbone of the DNA. Favourable entropy is due to the release of water from the minor grooves upon binding, and the release of counterions from the backbone. The hydrophobic effect removes the ligand from the water phase into the binding site. This close association enables van der Waals interactions between the aromatic rings of the ligand and groove walls that make small enthalpic contributions to the overall binding event.

1.6.1.4 Solvent and heat capacity

An important contribution to binding free energy is the interaction of a molecule with the solvent environment. Water is believed to play an important role in a number of biochemical interactions, including the formation of binding complexes between ligands and DNA, as well as complexes between ligands and proteins.[77,89-90] Water can contribute in the formation of ligand-DNA complexes in two different ways; by disrupting the solvent cage around a nucleic acid which can affect binding affinity through the entropically favorable disordering of the solvent cage;[73,91] or by individual water molecules forming H-bonded bridges between the nucleic acid and the ligand to assist complex formation.[74-77] Heat capacity changes (ΔCp) associated with intermolecular binding are believed to arise mainly from the influence of the solvent,[92-93] unlike enthalpy, entropy, and free energy, which have contributions from different sources. Therefore, it has been proposed that ΔC_p can be used to provide information about solvation decoupled from other effects. However, the heat capacity changes that accompany ligand-DNA binding interactions are often significant, and the molecular interpretation of these changes is very challenging. Although the magnitude of ΔC_p has been related to the solvent accessible surface area (SASA) of the molecule, other studies have shown that numerous potential factors can contribute to observed heat capacity changes. These factors include:

(1) The hydrophobic effect that results from the release of constrained water molecules accompanying the burial of non-polar surfaces has distinct effects on the ΔC_p. Specific ligand–DNA interactions have a characteristically large negative ΔC_p [-100 to -550 cal/mol.K] upon site specific binding,[63,72,86] which arises from the burial of groove floor walls and the ligand′s binding surface. However, this simple picture based on the burial of non-polar surfaces may be complicated by a number of additional factors, detailed below.

(2) The restriction of ″soft″ internal vibrational modes of polar groups and bound water molecules mediating the binding interaction.[74,93-94] On being trapped in a ligand–DNA interface as part of a defined network of hydrogen bonds, water molecules in addition to other polar groups at the binding site, can potentially suffer a reduction in their ″soft″ vibrational modes, reducing their heat capacity, which ultimately increases the magnitude of the negative ΔC_p term for complex formation.[93,95]

(3) Electrostatic interactions between the positively charged ligands and the negatively charged backbone of DNA play an important role in stabilizing ligand–DNA complexes and have an impact on observed ΔCp, which Sharp *et al* [96] have shown is positive in sign and small in magnitude (15–90 cal/mol.K). Ions in the solvent also contribute significantly to ΔC_p during ligand–DNA complex formation.[97] Both the DNA, and many DNA binding molecules are highly charged and undergo a substantial change in their local electrostatic field on binding, which in turn changes the number of counterions associated with the DNA. Consequently, ionic strength of the solution strongly affects binding affinity and heat capacity in almost all known ligand–DNA complexes.[98-99]

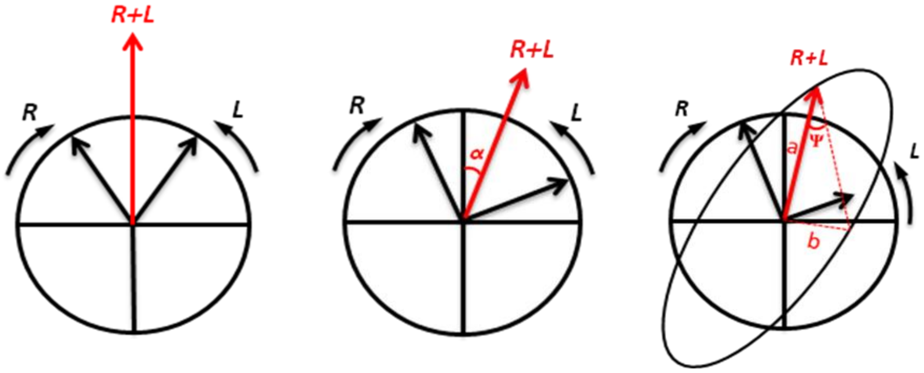
(4) The coupling or linkage to binding of any temperature dependent equilibria, such as protonation or conformational changes within the interacting macromolecules affects ΔCp.[93-94,100-101] In many cases, the formation of ligand-DNA complexes is not

a rigid body interaction, and the effects of the conformational changes in the structure of the ligand and/or DNA will be included in the overall observed ΔCp. These changes could affect the buried surface area of polar or apolar moieties; the burial of nonpolar surfaces results in a negative contribution to ΔCp, whereas the burial of polar surfaces produces a positive contribution to ΔCp. Proton release or uptake is often observed in biological systems when interactions take place,[102] and binding linked protonation during DNA –MGB complexation has been found to have an impact on enthalpy and heat capacity of binding.[103]

1.6.2 Circular dichroism (CD)

1.6.2.1 Principles of circular dichroism[104]

CD is the difference in the absorption of left (L-CPL) and right (R-CPL) circularly polarized light. This phenomenon is observed when a molecule that contains one or more chiral chromophore (light-absorbing group) absorbs left and right hand circular polarized light slightly differently. Linearly or plane polarized light is the superposition of two opposite circular polarized lights (R-CPL and L-CPL) of equal intensity and phase (Figure 1.16 A). The difference in the speed of L-CPL and R-CPL passing through an optically active material creates a phase-shift between the two circularly polarized components when they exit the sample. After the beam emerges from the medium, both components then resume their original speed but now the two components are out of phase. When the electric field vectors of the R-CPL and L-CPL are summed, an inclination (optical rotation) (α) of the plane of polarised light is evident (Figure 1.16 B).



(A) **(B)** **(C)**

Figure 1.16 A schematic representation of **(A)** plane polarised light: the superposition of two opposite circular polarized lights (R-CPL and L-CPL) of equal intensities and phase. **(B)** Optical rotation (α): the turning of the plane of linearly polarized light about the direction of motion as the light passes through an optically active medium. **(C)** Ellipticity (Ψ): the difference in absorption of the left- and right circularly polarized light leads to ellipticity (known also as circular dichroism) in addition to optical rotation.

The speed of light through a solution is represented by the refractive index (n), which is a measure of how fast light passes through a medium. The higher the value of n, the slower the light speed. At certain wavelengths, the optically active substance not only causes the L- and R-CPL to travel at different speeds, but if the R-CPL and L-CPL are absorbed to different extents, then the relative magnitudes of their electric vectors are also altered. Thus, not only an angle of rotation (α) is observed, the electric vector of the light follows an elliptical path, as shown in Figure 1.16 C, which corresponds to the elliptically polarised light. The major (a) and the minor (b) axes of the ellipse form a triangle and the angle opposite the minor axis is the ellipticity (Ψ). This angle is proportional to circular dichroism and mathematically expressed by the following equation:

$$\Psi = tan^{-1}\frac{a}{b}$$

Eq.1.9

where Ψ is the ellipticity in millidegrees, **a** and **b** are the length of the major and minor axes of the ellipse, respectively.

CD measurements are performed in the visible and ultra-violet region of the electro-magnetic spectrum. If the sample under study contains optically active chiral chromophores, a differential absorption of L-CPL and R-CPL will take place and a CD signal is observed as a result of this difference in absorption. Positive CD signals reflect a higher absorption of L-CPL compared to R-CPL, while negative CD signals reflect a higher absorption of R-CPL.

1.6.2.2 Instrumentation

CD spectra can be obtained by using a CD spectrometer (Figure 1.17), such as the Chirascan, which is a highly specialised derivative of a UV/visible spectrometer. The light source in the CD instrument is an extremely high intensity xenon (Xe) arc lamp (500 W). This is needed in conjunction with a high quality monochromator system because ΔA is usually small and because wavelengths as low as 185 nm are required. The polariser produces a plane-polarised light and the modulator converts the linear polarised light into L- and R-CPL, which are passed through the sample cell. In the absence of a CD active sample, the L- and R-CPL have equal light intensities. Consequently, there is a steady output from the light detector. A CD active sample will absorb either L- or R-CPL preferentially. Thus the detector generates a fluctuating signal depending on the different intensities of the L-CPL and R-CPL beams. This is amplified and translated to give a CD signal in units of millidegrees or ΔA. Plotting CD signal vs wavelength of the light generates a CD spectrum.

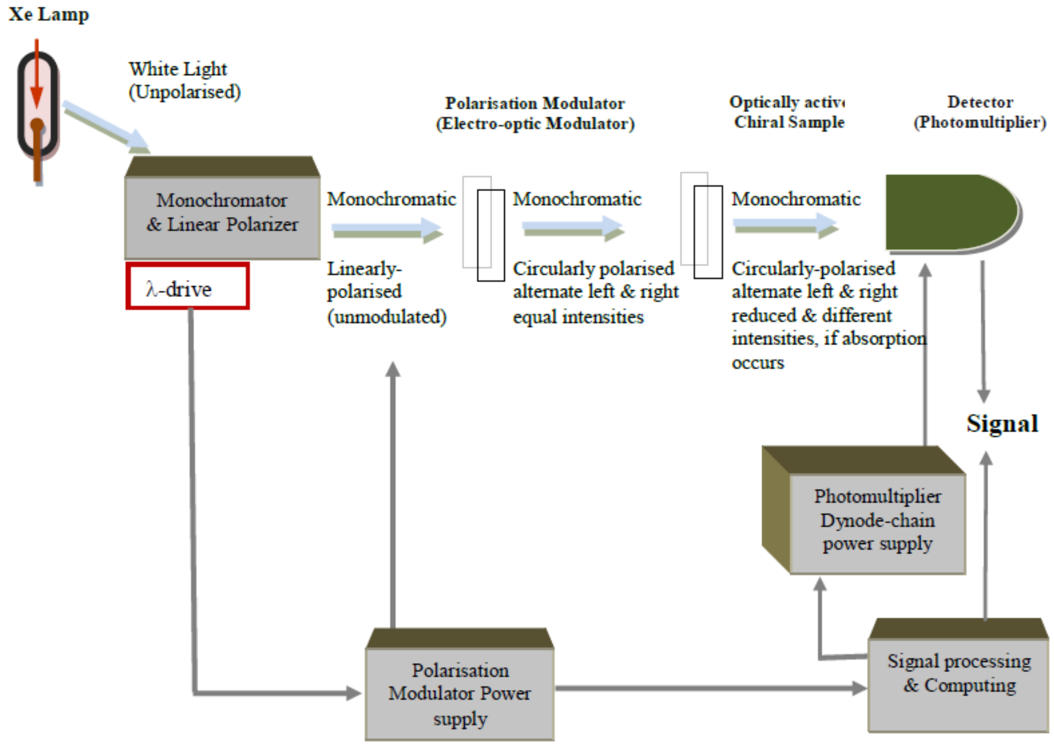


Figure 1.17 Schematic of the principle components of a CD spectrometer.

1.6.2.3 Applications of CD

CD spectra are usually recorded over a range of wavelengths and used to study chiral molecules, especially large molecular weight biological molecules where its most important applications can be exploited. CD spectroscopy has been extensively used in studying the secondary structure/conformation of macromolecules, particularly proteins and DNA. This stems from the fact that secondary structures are sensitive to environment, e.g. temperature or pH, and CD can be used to investigate how different environmental conditions affect the secondary structure of macromolecules or their interactions with other molecules. Structural, kinetic and thermodynamic information about macromolecules can also be derived from CD spectroscopy. In our study of MGBs, CD has been used to determine the binding constants for their association with different sequences of DNA. The presence of a CD signal in the absorption band of a nonchiral ligand demonstrates interaction with a chiral molecule (DNA). The CD signal of ligands bound in the minor groove of B-DNA is strong and always positive and proportional with the affinity of binding.[105] The proportional relationship between the amount of ligand added and the CD response of the resultant complex has been used to determine the association constant of these complexes. Binding constants were calculated by non-linear least squares fitting of Engel′s equation[106] for tight ligand binding to the CD data:

$$\Delta\varepsilon = \Delta\varepsilon_{\max}\left\{\frac{1+k[X]+k[D]}{2k[D]} - \sqrt{\left(\frac{1+k[X]+k[D]}{2k[D]}\right)^2 - \frac{[X]}{[D]}}\right\}$$

Eq. 1.10

where $\Delta\varepsilon$ is the molar ellipticity at each titration point, $\Delta\varepsilon_{max}$ is the molar ellipticity once all the ligand has been titrated into the DNA, *k* is the binding constant , X is the ligand concentration (variable) and D is the DNA concentration (constant).

1.6.3 Nuclear magnetic resonance spectroscopy (NMR Spectroscopy)

Nuclear magnetic resonance (NMR) spectroscopy is a versatile, non-invasive and powerful technique for the elucidation of the three-dimensional structure of biological macromolecules in solution.[107-109] After the observations of nuclear magnetic resonance in 1946,[110-111] enormous progress has been achieved in the applications of NMR spectroscopy. This mainly stems from huge advancement in superconducting high-field NMR magnets combined with state-of-the-art electronics and the computers associated with these instruments. The availability of high resolution NMR spectrometers (i.e. 600 MHz, 800 MHz) coupled with Fourier transform technology make it possible to study biological samples (nucleic acids and proteins) at relatively low concentration (0.1-10 mM) during short analysis times. Computer control of NMR spectrometers has improved both data acquisition and processing and has given rise to advanced multidimensional NMR techniques which allow the determination of the nuclear spin connectivity of nuclei through space (NOESY) and through bonds (COSY). The connectivity through space is a consequence of the nuclear Overhauser effect (NOE), which can be used to establish distances between protons and to deduce a three-dimensional structure of a molecule. The connectivity through bonds is related to the correlations between protons that are coupled to each other. Both techniques have been widely used to study the structures of biological macromolecules and will be discussed later in this chapter.

1.6.3.1 Basics of NMR spectroscopy[112-113]

NMR spectroscopy arises from the fact that certain nuclei of atoms have magnetic properties that can be exploited to generate structural information. Nuclei with an odd mass or atomic number have ″nuclear spin″ similar to the spin property of electrons (e.g. ^{1}H and ^{13}C) . If the number of neutrons and protons are both even (e.g. ^{12}C, ^{16}O, ^{32}S), these spins are paired against each other so that the nucleus of the atom has no overall property of spin. A nucleus with spin can be considered as a charged particle in motion, which therefore generates its own magnetic field. A fundamental parameter of nuclei with ′spin′ is the spin quantum number *I*. For each nucleus, a magnetic moment exists, μ, which is proportional to its spin and can be represented by Eq. 1.11:

$$\mu = \frac{\gamma I h}{2\pi} \qquad \text{Eq.1.11}$$

where γ is the gyromagnetic ratio and *h* is Planck′s constant.

^{1}H and ^{13}C nuclei, for instance, have nuclear spins, $I = ½$. In quantum mechanical terms, a nucleus possessing a spin, *I*, will be able to adopt $2I + 1$ possible orientations. Therefore, ^{1}H and ^{13}C nuclei with spin = ½ will have two possible orientations. In the absence of an external magnetic field, these orientations are of equal energy and randomly orientated. When such nuclei are placed in a strong magnetic field, they line up 'parallel' to the applied field with their spin orientations being aligned with or aligned against the magnetic field. As a consequence, the energy levels split and two spin states (energy levels) are formed: the α (+ ½) and β (- ½) states. The difference in energy between the two spin states is directly proportional to the strength of the external magnetic field. Each level is given a magnetic quantum number, m (Figure 1.18).

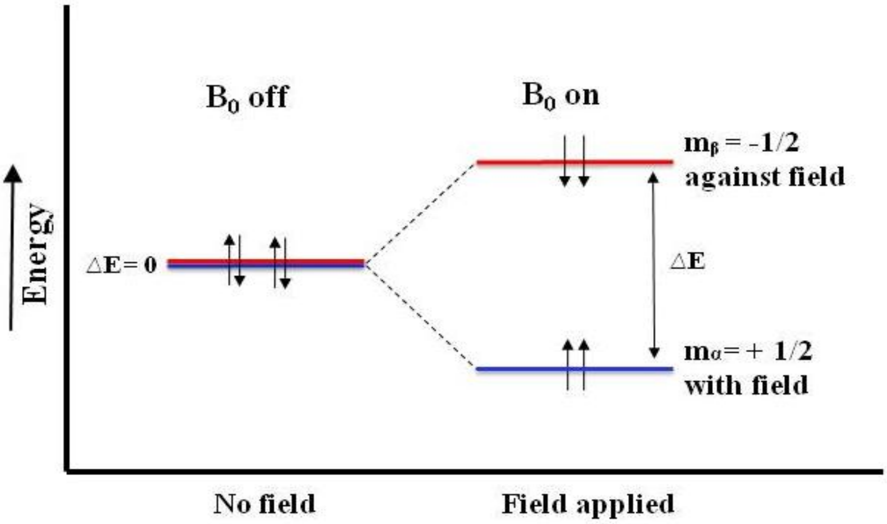


Figure 1.18 Diagram for the ground and excited states of a nucleus with a spin of ½.

Irradiation of atoms, usually protons or carbon-13 atoms, with radio frequency (RF) energy corresponding exactly to the spin state separation of their nuclei will cause excitation of those nuclei in the lower energy +½ state to the higher energy -½ state. The energy value of the spin states can be calculated using the Larmor equation, which describes the relationship between the precessional frequency v_0 (called the Larmor frequency), the strength of the external magnetic field (B_0) and the gyromagnetic ratio (γ) (Eq 1.12):

$$v_0 = \frac{\gamma}{2\pi} B_0$$

Eq.1.12

An external source of electromagnetic radiation (v_1) causes the transition of the nuclear spins between the two energy states. When the v_1 frequency is equal to the exact v_0 value of the nucleus under study, a transition from the more stable α state to the less stable β state takes place.

The range of frequencies used for excitation and the complex patterns of splitting generated are very characteristic of the chemical structure of a molecule. This arises from the fact that nuclei in different parts of the molecule are exposed to different local magnetic fields depending on the molecular structure, and therefore have different frequencies at which they absorb energy. These differences are called chemical shifts. In the proton NMR spectrum, the chemical shifts (or frequency differences) are plotted relative to the signal arising from a standard compound, (e.g. tetramethylsilane -TMS, which is defined to be at 0 ppm (parts per million); for proton magnetic resonance (Eq 1.13):

$$\boldsymbol{Chemical\ shift\ (\delta) = \frac{frequency\ of\ signal - frequency\ of\ reference}{spectrometer\ frequency}\ \mathrm{X}\ 10^6}$$

Eq.1.13

Expressing the spectrum scale as parts per million (ppm) and independent of the spectrometer frequency aims to make the NMR spectrum handy and more manageable.

1.6.3.2 Instrumentation

Figure 1.19 shows the basic layout of an NMR spectrometer. The NMR tube containing the sample is placed in the magnetic field and rotates by means of an air turbine at ca 30 revelations/s to ensure the uniformity of the magnetic field across the sample.

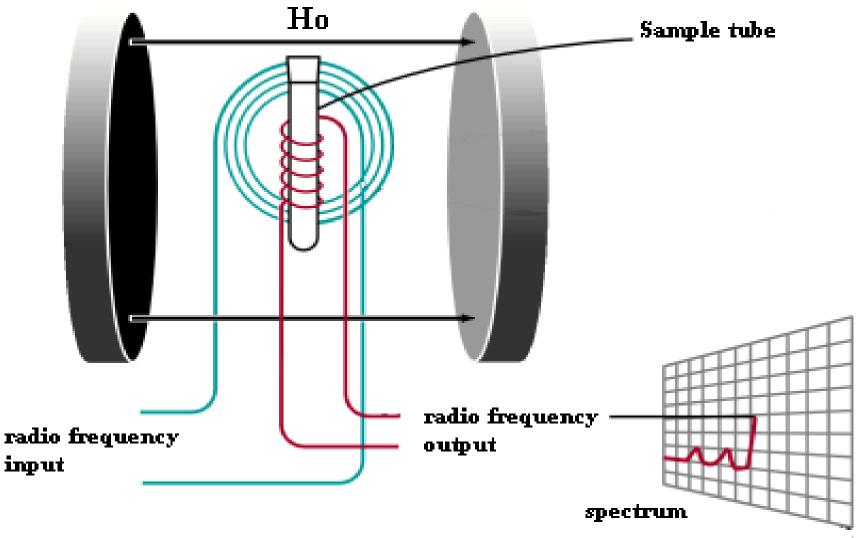


Figure 1.19 Schematic diagram of NMR spectrometer.[114]

The sample is analysed in a deuterated solvent to avoid interference between the sample and solvent protons. Modern instruments use the Pulsed Fourier Transform (PFT) method to excite the sample. In this method, a short RF signal is applied to excite the nuclei of interest and when the value of v_1 produced by electromagnetic radiation matches the Larmor frequency (v_0) of the nucleus of interest, resonance occurs that produces a signal called free induction decay (FID). In the PFT technique (usually called FT-NMR), emission of radiation is exploited to record the output signals as the excited nuclei relax back to the ground state following the short high energy pulse of radiation. A complex FID is usually observed resulting from the interference between v_1 and different v_0 present in a sample with non-equivalent nuclei. A number of scans are usually performed to allow the signals to be identified from the background noise and the collected FIDs are stored in the computer. Fourier transformation analyses (mathematical manipulations) of the complex FID outputs are then performed to produce the NMR spectrum.

1.6.3.3 One dimensional NMR

The conventional 1D NMR spectrum is a plot of intensity (number of atoms per signal) vs. frequency (chemical shift). Figure 1.20 shows the resonance region of various protons of the d(CGATCG) DNA hexamer in the 1D ^{1}H-NMR spectrum acquired in D_2O.[115]

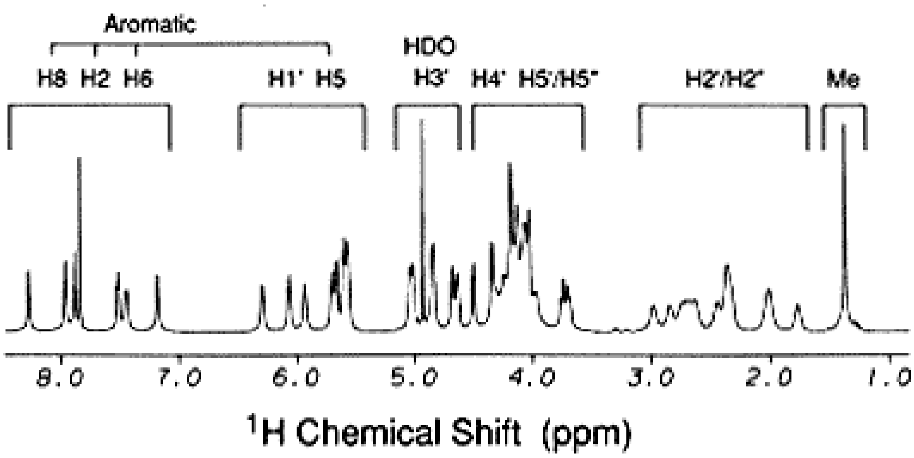


Figure 1.20 1D ^{1}H-NMR spectrum (500 MHz) of the oligonucleotide d(CGATCG) in D_2O at a sample temperature of 298 K and neutral pH.[115]

The non-exchangable protons resonances of nucleic acids usually appear in four main regions of the 1D ^{1}H-NMR spectrum: 7-8.5 ppm for the aromatic protons, H8, H2, and H6; 5.5- 6.5 ppm for the aromatic H5 proton of cytosine and the deoxyribose proton H1′; 3.5-5 ppm for the deoxyribose protons H3′, H4′, and H5′/H5″; 1.2-3.0 for the deoxyribose protons H2′/H2″ and the methyl protons of thymine. The exchangeable protons of the amino and imino protons of guanine and thymine respectively do not appear in the NMR spectrum recorded in D_2O as they exchange with the solvent. However, using water enables the detection of these protons, but the intense proton signal from water at 4.83 ppm needs to be suppressed so that other resonances can be observed.

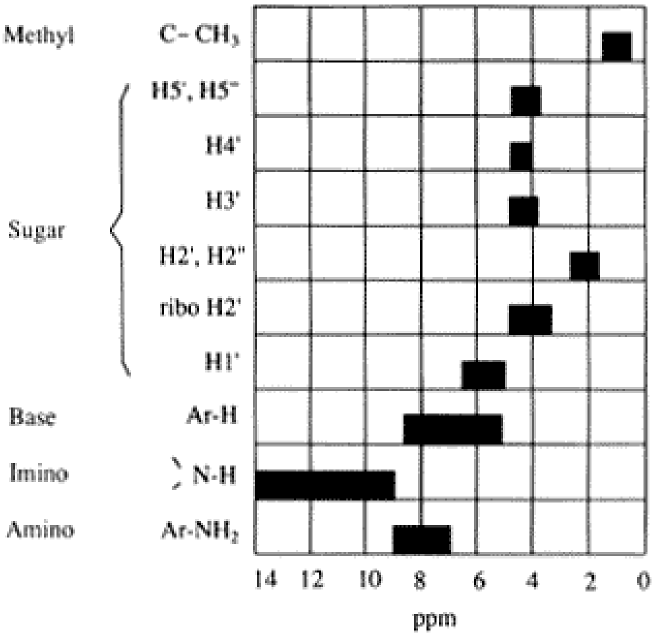


Figure 1.21 A schematic representation showing the characteristic chemical shifts of the various protons in nucleic acids.[115]

1.6.3.4 Two dimensional NMR

The concept of the 2-D NMR experiment was originally introduced by Jeener[116] and extensively expanded by Ernst and others.[117] In 2D NMR spectroscopy, intensity is plotted as a function of two frequencies, F1 and F2 and the FID is obtained as a function of two time variables. This is achieved by introducing an incrementable delay in the pulse sequence which is usually called t_1; the acquisition time during which an FID is acquired is known as t_2. A series of FIDs are recorded as a function of t_2 for each value of $t_{1,}$ and stored separately in the computer. The raw data are then Fourier transformed to generate a 2D NMR spectrum. The spectrum is represented as a contour plot in which each signal has two frequency co-ordinates corresponding to F1 and F2. The 2D experiment is performed in four steps outlined below (Figure 1.22).

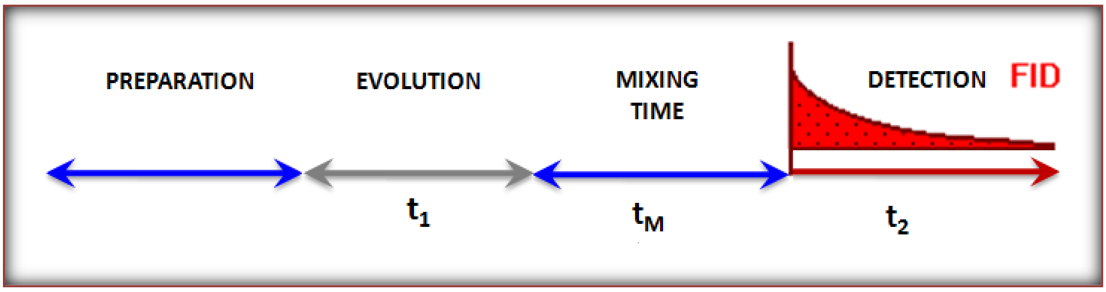


Figure 1.22 A scheme for the time pulse in the 2D NMR experiment.

During the preparation step the spins are allowed to relax to equilibrium. The sample is then excited by a radio frequency pulse. The generated spin magnetization is allowed to evolve into other states during the first time period, t_1. This is followed by another time period called the mixing time during which magnetization is allowed to mix between states without evolution occurring followed by the application of another pulse. After the mixing period, the signal in the form of a FID (Free Induction Decay) is recorded during the acquisition time as a function of the second time variable, t_2. This process is called the pulse sequence and the exact features of the preparation and mixing steps determines the information generated in the 2D NMR spectrum.[112]

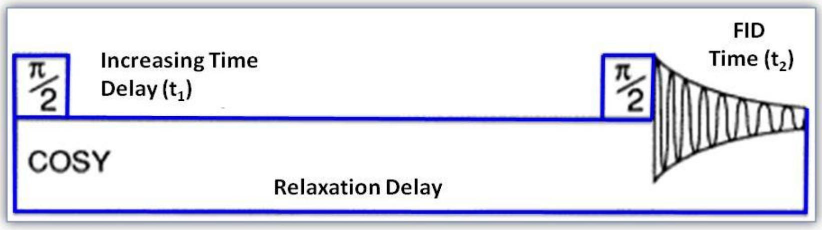
There are different types of 2D NMR experiments that have been developed depending on the magnetization transfer that is measured. The transfer usually takes place through bonds to the same type of nucleus (e.g. COSY, TOCSY) or to a different type of nucleus (e.g. HSQC and HMBC) or through space (e.g. NOESY and ROESY). These experiments have been widely applied to study ligand-DNA complexes and will be described below.

1.6.3.4.1 COSY (COrrelation SpectroscopY) and TOCSY (TOtal Correlation SpectrscopY)[118]

In the 2D [^{1}H, ^{1}H] COSY NMR spectrum, cross-peaks are generated due to the magnetization transfer that occurs through the bonds. This is usually observed between protons belonging to the same spin-system and generally separated by between 2 and 4-chemical bonds. The COSY experiment is used to identify the adjacent aromatic protons of cytosine, H5 and H6 in the NMR assignment of nucleic acids. The 2D TOCSY generates cross peaks between all protons within a spin system even if they are not directly coupled. In COSY-90, which is the most commonly used COSY experiment (Figure 1.23), the sample is excited with an RF pulse, p1, which tilts the nuclear spin by 90° (net magnetic spin polarisation is perpendicular to the applied external field). After applying p1, the sample magnetization is allowed to freely evolve into other states during the evolution period (t_1). Another 90° pulse, p2, is then applied, after which the signal is acquired in the form of a FID (free induction decay). This pulse sequence is repeated using a series of different evolution periods (t_1). All the generated data is then Fourier transformed in each dimension to generate the two dimensional spectrum containing COSY cross peaks which appear as a result of varying evolution periods. The principle of the TOCSY experiment is similar to COSY except that after the evolution period t_1, the magnetization is spin-locked and this makes the spin system to be strongly coupled through scalar coupling. This allows magnetization to be transferred successively over up to 5 or 6 bonds as long as successive protons are coupled. The TOCSY experiment is extremely useful for identifying protons on the sugar rings of the backbone of DNA.

The COSY and TOCSY experiments are not reserved only for homonuclear ^{1}H NMR spectroscopy (^{1}H-^{1}H) but they can also used with ^{31}P and ^{19}F for instance where there is high natural abundance of the spin active nucleus which therefore provides a quick method to establish through bond conectivities .

(A)



(B)

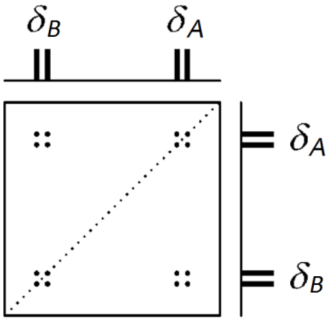


Figure 1.23 (A) simplified schematic representation showing the pulse sequences for 2D COSY experiment **(B)** Basic representation of a 2D COSY NMR spectrum of an HA/HB system.

1.6.3.4.2 NOESY (Nuclear Overhauser Effect Spectroscopy)[118]

NOESY is an indispensable tool for the structure determination of nucleic acid structures and their complexes in solution. In 2D [^{1}H,^{1}H] NOESY NMR spectroscopy, cross-peaks are generated due to the magnetization transfer through space between protons that are spatially close to one another (within a distance of 5 Å), even if they are not connected by chemical bonds. Thus intramolecular and intermolecular interactions are observed and structural information can be obtained for ligand-ligand or ligand-DNA association. The difference between NOESY and COSY experiments is the application of a second 90° RF pulse and delay (t_m), termed the mixing time, in the NOESY experiment (Figure 1.24). The protons in the molecule have an opportunity to cross relax amongst themselves during the mixing period, leading to the exchange of spins. This cross-relaxation is known as the nuclear Overhauser enhancement and is observed as the off-diagonal cross-peaks in the 2D-NOESY spectrum. Figure 1.25 shows the complete 2D-NOESY spectrum of the nonexchangeable protons of d(CGATCG)$_2$.

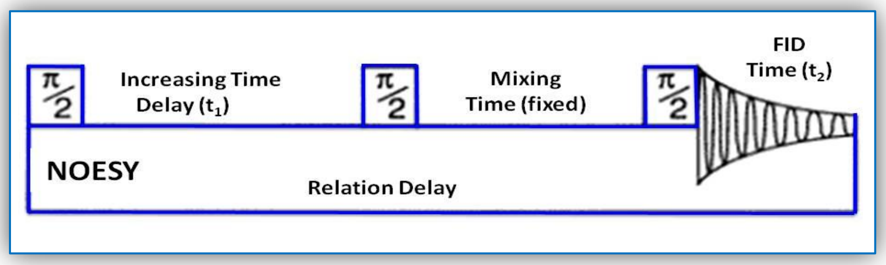


Figure 1.24 A simplified schematic drawing showing the pulse sequences for 2D nuclear Overhauser effect spectroscopy (NOESY).

The rate of cross-relaxation between protons A and B is inversely proportional to the sixth power of the distance between the two protons. This inverse relationship means that the measured NOE becomes stronger as the distance between two protons decreases and *vice versa*. It is worthy to note that the observed cross-peak intensity results from multiple pathways for relaxation within the mixing time period. At very short mixing time, the observed NOE cross-peak volume is mainly generated by the direct relaxation between the two spins, but at longer relaxation times many secondary relaxation pathways become significant, which could affect the interpretation of NOESY cross peak volumes. Therefore, these cross-peaks are interpreted within the context of a relaxation matrix if long mixing times are used to encompass for the rate of relaxation between all nuclei simultaneously.

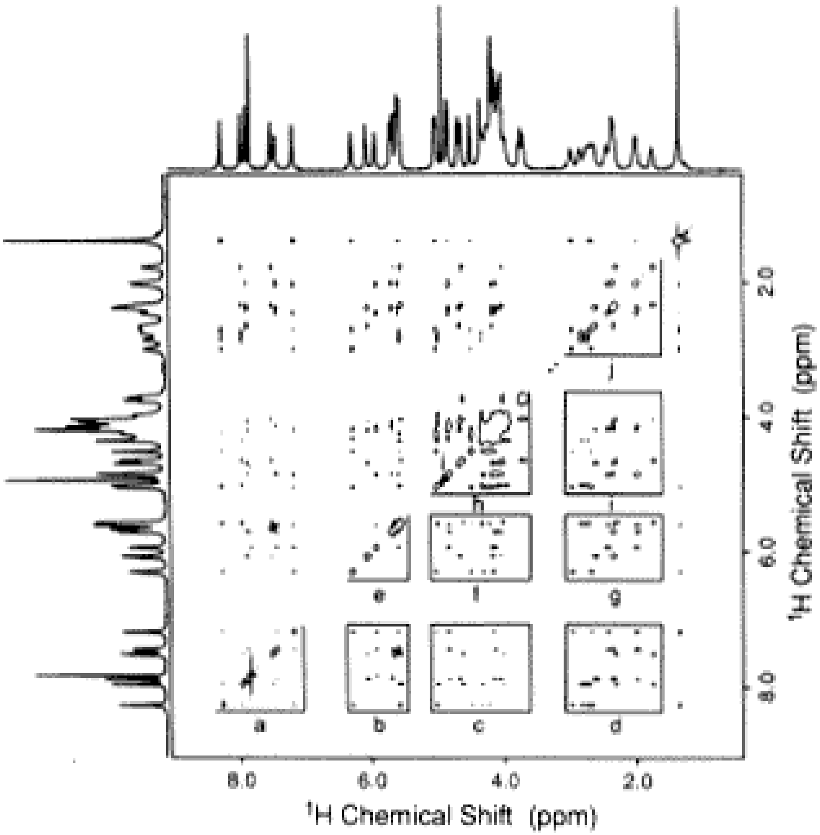


Figure 1.25 2D-NOESY of d(CGATCG)$_2$ in D$_2$O at 10°C. Different regions of the cross-peaks between protons are labelled. (a) aromatic-aromatic. (b). aromatic-H1′/H5. (c) aromatic-H3′, H4′, and H5′/H5″.(d) aromatic-H2′/H2″. (e) H1′/H5-H1′/H5. (f) H1′/H5 -H3′, H4′, and H5′/H5″(g) H1′/H5-H2′/H2″. (h) H3′, H4′, and H5′/H5″-H3′, H4′, and H5′/H5″. (i) H3′, H4′, and H5′/H5″-H3′, H4′, and H5′/H5″. (i) H3′, H4′, and H5′/H5″-H2′/H2″. (j) H2′/H2″-H2′/H2″.[115]

1.6.3.4.2.1 NOESY data assignment

In a double-stranded DNA dodecamer, there are more than 200 hydrogen atoms, which make the NMR spectrum highly overlapped and very complex. Finding certain landmarks in the COSY and NOESY spectra is extremely helpful in tracing the connectivity pathway at the beginning of an NMR data assignment. In the COSY spectrum, the most distinctive cross-peaks are observed between the H5 and H6 proton resonance of the same cytosine base. These protons are easily identified and are used to determine the chemical shifts of cytosine bases and the corresponding NOESY cross peaks in the NOESY spectrum. In 2D NOESY NMR data of DNA, each nucleotide aromatic resonance has a different chemical shift associated with it. In general, A′s are more downfield, T′s are more upfield, and C′s and G′s are intermediate. Before pairwise distances can be derived from 2D NOESY data, cross peaks must be correlated to the correct nuclei, which can be largely achieved by using sequential assignment procedures. This requires knowledge of both the exact sequence of the DNA sample before beginning spectral assignment and the chemical shift of the first nucleotide in the sequence. The studied DNA sequence is usually known and the first nucleotide can be determined using the following rules:

1) The chemical shift of H5′/H5″ protons at the 5′ end is characteristic of the first nucleotide in the sequence . This chemical shift is shielded due to the absence of the phosphate group at the 5′ end and is often observed at 3.7 ppm.

2) Only H2′ and H2″ protons of the first nucleotide will show up its frequency. For example, in the sequence CTAG at the chemical shift for C H6, its own H2′ and H2″ protons will be seen, but none of the other nucleotides. For T, its own 2′ and 2″ protons will be seen, as well as those from C. Once the first nucleotide has been found, it is possible to determine which nucleotide is next in the sequence through the recognisable NOE patterns which are found in B-form DNA.[108] In B-form DNA a deoxy ribos H1′ protons has a weak NOESY cross peak to its pyrimidine H6 or purine H8 base proton , and another to the (n+1) neighbouring base proton, where sequence numbering goes in the 5′ to 3′ direction. The deoxyribose H2′ and H2″ protons display a similar pattern. With this information, protons can be assigned sequentially along each DNA strand until all nucleotides have been assigned. (Figure 1.26)

The advantage of the rapid exchange rate of nitrogen –bound protons in DNA can be exploited to simplify the ^{1}H NMR spectrum for assignment purposes. Dissolving the DNA sample in D_2O rather than H_2O, makes the base amino and imino protons exchange with deuterons thereby effectively disappearing from the spectrum, as deuterons resonate at a very different frequency compared with protons. This technique also significantly reduces the resonance arising from the solvent, simplifying the solvent suppression problem.

The full assignment of the NOESY NMR spectrum provides the entry point for obtaining the entire dataset of the volumes of all NOE crosspeaks. However, a major problem in the measurement of the NOE volume associated with a given proton is that many crosspeaks are severely overlapped.

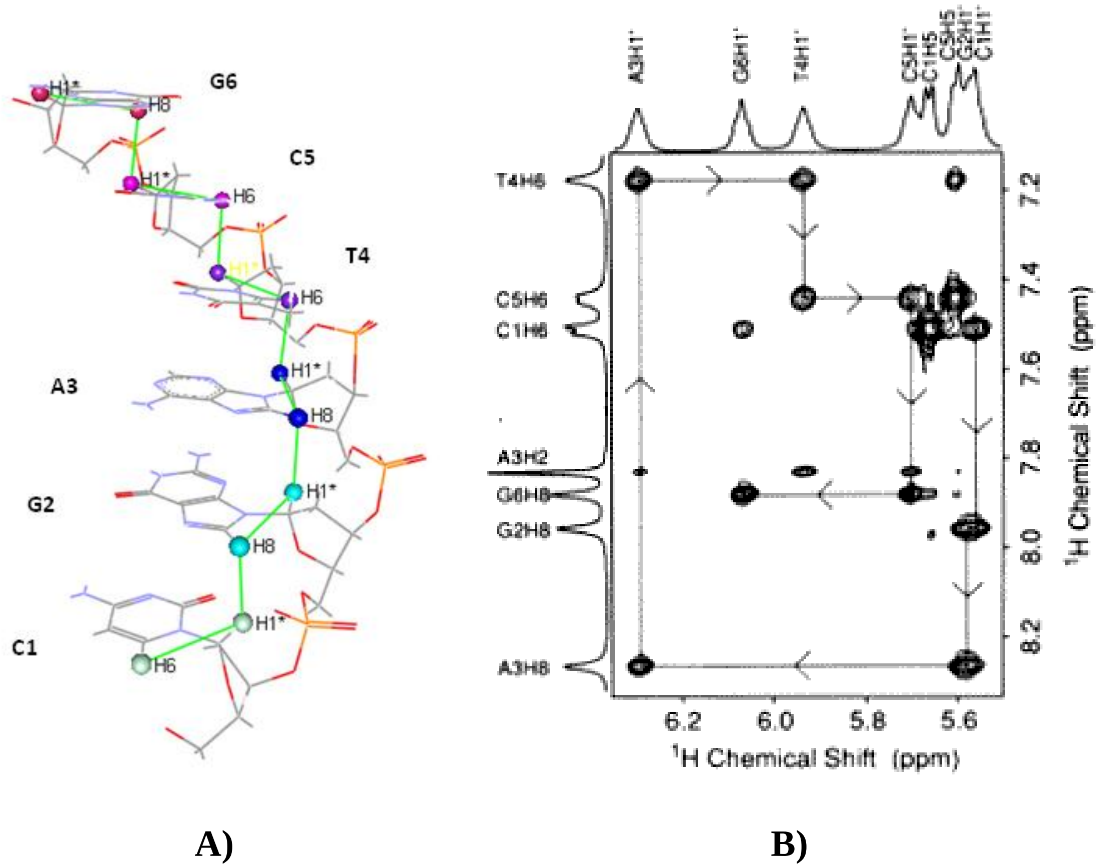


Figure 1.26 A) A representation of a single strand of the d(CGATCG)$_2$ duplex in the B-DNA conformation. The through-space connections between the aromatic H6/H8 and the sugar H1′ protons (both shown as large circles) are represented by a green solid line. B) An expanded 2D [^{1}H,^{1}H] NOESY NMR spectrum of the nucleotides in the d(CGATCG) hexamer , showing the cross-peaks between aromatic and sugar H1′ protons.

1.6.3.4.3 HSQC (Heteronuclear Single Quantum Coherence) and HMBC (Heteronuclear Multible Bond Correlation)

These methods yield correlations between two different nuclides. The generated spectrum is two-dimensional with one axis for ^{1}H and the other for a heteronucleus (an atomic nucleus other than a proton such as ^{13}C, ^{31}P, or ^{15}N). HSQC shows single bond ($^{1}J_{XY}$) connectivities although it can be tuned, for instance for $^{2}J_{XY}$ if the coupling constant is large enough. The 2D [^{1}H, ^{31}P] HSQC and [^{1}H, ^{13}C] HSQC experiments are frequently used in nucleic acid NMR to solve the ambiguity in the assignment of H4′ and H5′/H5″ protons which are overlapped and not well resolved. Moreover, HSQC is also useful for detecting interactions with ligands by comparing the HSQC of the free and bound DNA. If there is an interaction, one should expect changes in the chemical shifts of the peaks, which are most likely to occur in the binding interface. The HMBC experiment differs from the HSQC in that long-range multiple bond correlation (over two or three bonds) can be shown. In HMBC the cross-peaks are between protons and heteroatoms that are two or three bonds away and direct one-bond cross-peaks are suppressed.

1.6.4 Molecular modeling

Molecular modeling is a computational technique that can be used to study the three-dimensional structure, dynamics and properties of molecules. Molecular dynamics is complementary to experimental techniques, such as NMR spectroscopy and X-ray crystallography in the determination of the three-dimensional structure of nucleic acids and their complexes. Molecular dynamic simulations can provide a complete theoretical description of DNA structure and motions which helps in the development of DNA models and the interpretation of the experimental data. Moreover, the force fields of molecular dynamics play a key complementary role in the determination of three-dimensional structures by iterative refinement procedures in crystallography and NMR spectroscopy using the electron density or nuclear Overhouser effect restraints, respectively. In these iterative refinements, the experimental data are fitted to an empirical force field.[119] Explicit solvation models (using periodic boundary conditions and the particle mesh Ewald method for representation of electrostatic interactions) are commonly used in NMR structure refinement. This is a reasonable approach as the NOEs are determined in an aqueous medium and the best representation of the solution phase should be used for accurate structure refinement.

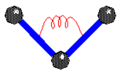
Once the distances between protons have been obtained from NOESY data, they are used as input restraints during the molecular dynamic calculations which generate a three-dimensional structure for the molecule. The most important aspects of these calculations will be summarised here. The primary reference for the following discussions is the user guide for the molecular modeling program Amber (Amber 10 users manual).[120]

1.6.4.1 Force fields

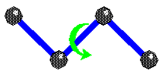
The force fields are analytical expressions of the potential energy surface of a system in terms of nuclear coordinates, and are used to calculate the geometry and the energy of molecules. Force field functions and parameters are empirically derived or generated from quantum mechanics calculations. Amber is one of the most widely used force fields in the simulation of nucleic acids and proteins.[121] The functional form and description of the terms in the Amber energy expression are shown in Figure 1.27.[122] The energy is divided into bonded interactions which occur directly through covalent bonds (terms 1-3) and non-bonded interactions represented by electrostatic (term 4), van der Waals (term 5) and hydrogen bonding (term 6) contributions. All of the atoms in a system which are modelled under the Amber force field are assigned atom types which specify element type, hybridisation, and bonding environment. The force field contains idealised equilibrium values for each individual atom in the system, including bond length, bond angle, and van der Waals radius, in addition to an adequate force constant associated with them. As the simulation changes the global structure of the molecule, an energetic penalty is assigned to structures which deviate from idealised equilibrium geometry. The severity of the energetic penalty is determined by both the amount of deviation from equilibrium, and by the adjustable parameter, the force constant. The quality of the calculated results are determined by the ability of the force field to accurately describe the structure and energetic of the system of the interest.

$$E_{pot} = \sum_{b} K_b(B - B_0)^2 \qquad \textbf{(1)}$$

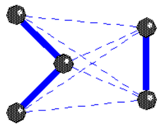
$$+\sum_{\theta} K_\theta \ (\theta - \theta_0)^2 \qquad \textbf{(2)}$$



$$+\sum_{\phi} K_\phi \ [1 + \cos(n\phi \ - \ \phi_0)] \qquad \textbf{(3)}$$



$$+\sum \varepsilon \left[\left(\frac{r^*}{r}\right)^{12} - 2\left(\frac{r^*}{r}\right)^{6} \right] \qquad \textbf{(4)}$$



$$+\sum \frac{q_i q_j}{\mathcal{E}_{ij} r_{ij}} \qquad \textbf{(5)}$$

$$+\sum \ \left[\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right] \qquad \textbf{(6)}$$

Figure 1.27 The functional form of the standard AMBER forcefield.[122] The first three terms of the Amber energy expression characterise the energies needed to stretch bonds, bend bond angles, and twist dihedral angles away from their reference equilibrium values. The fourth and fifth terms describe non-bonded interaction energies, van der Waals and electrostatic interactions, which are a function of internuclear distances. The sixth term is a small hydrogen bond term to account for energies not included in the electrostatic terms. The parameters are B=current bond length, B_0= equilibrium bond length, θ = current bond angle, θ_0 = equilibrium bond angle, ϕ = current dihedral angle, ϕ_0 = equilibrium dihedral angle, ε = depth of the Lennard-Johnes potential well, r*=equilibrium internuclear distance, r = current internuclear distance, q = point charge, ε_0 = permittivity constant.

Because of the large number of atoms in a nucleic acid or a protein (a DNA decamer has approximately 630 atoms), the summation of all interactions in the molecule becomes enormous, and calculations become unmanageable and very computationally inefficient. Many of the pair-wise electrostatic and van der Waals interactions are insignificant because these energies fall off quickly with distance. Therefore, it is common to set a non-bond cut-off distance which excludes non-bonded atom pairs beyond the cut-off distance from the calculation and thus hugely shorten the time of analysis. Two cut-offs are usually used, one for van der Waals interactions (around 9 Å), and the other for electrostatic interactions (longer cut-off of 12-15 Å).

Structural restraints obtained experimentally are included as extra terms in the input files for the molecular dynamic simulations, such as distance restraints obtained from NOESY cross peaks intensities. These distance restraints are input as ranges to allow for the experimental error and uncertainty in the measured distance.

1.6.4.2 Energy minimization

The goal of energy minimisation is to find the coordinates corresponding to the nearest minimum energy of a system.[119] At an energy minimum, if any coordinate is changed, the energy of the system increases, and the force applied according to the force constant drives the conformation back into the local energy well. The energy minimisation is carried out in two main steps. The starting structure is input and its target function is evaluated. The conformation is then altered slightly and the target function is re-evaluated. This process is iterated until the nearest local energy minimum is found.

There are two widely used algorithms that determine how the molecular conformation is adjusted at each iteration. The first is the method of steepest descents (SD) which is the most robust algorithm. This method follows directly the gradient of the potential energy surface. The SD algorithm works best when the conformation is far from the energy minimum (i.e. at the beginning of the calculations) and the gradients are large. As the structure approaches an energy minimum, this method loses efficiency because the gradient approaches zero. At this point, using the second method, the conjugate gradient algorithm (CG), becomes useful. Unlike SD, CG is more computationally intensive as it uses the coordinates from the two previous steps to calculate the following geometry.[123] Therefore, it works better when the system approaches the energy minimum. Generally, energy minimisation involves both algorithms, starting with SD to bring the system roughly near the local minimum, followed by CG to fine-tune the structure.

1.6.4.3 Restrained molecular dynamics

Restrained molecular dynamics is used to find the global energy minimum for the molecule using NOE-derived distance restraints. This approach is used to determine the three dimensional of the molecule in solution. The initial structure may not be close to the final, ″true″ structure, but if the data defines the system well enough, in theory it should be possible to start from any starting structure and arrive at the global energy minimum. Minimisation calculations only find the nearest local minimum in the target potential function. It is necessary to explore the conformational space more effectively by going over energetic barriers on the potential energy surface. Molecular dynamic calculations add kinetic energy to the system so that conformational space can be searched more effectively. There are two phases included in molecular dynamic simulations: the equilibrium and the production phase. In fully solvated systems, the equilibration phase is essential for the relaxation of the solvent and counterions before the production phase to reduce potentially destabilising van der Waals and electrostatic interactions. If the starting model is not optimal, the ligand-DNA structure should be refined in vacuum prior to explicit simulation, to get accurate starting structure and to remove any potentially bad contacts. Equilibrating the solvent and the conterions while the solute (DNA-ligand complex) is restrained reduces the artefacts from the interactions of the non optimised solvent with the complex. The restraints are then removed gradually to slowly allow the system to move freely. In the production phase, the Sander program of Amber is used to run the simulations usually at constant pressure and temperature with the PME summation and a time step of 1 fs. The hydrogen atom bond lengths are usually fixed, using the Shake method.[124] When the energy of the system settles around a value, an average structure can be determined from those structures where the energy has converged. The average structure can then be minimised to give an acceptable conformation of the system.

1.6.4.4 Generation of starting models for the free and bound DNA

The generation of DNA starting model can be done using various software, such as sybyl 6.3, insightII, Discovery Studio, Nucgen or NAB for example. These allow one to generate different forms of DNA (A, B, and Z forms). The starting DNA models may also be obtained from the experimentally determined structures published in the protein database (PDB) or the Nucleic Acid Database (NDB), if they are available. It should be noted that the format of the PDB file and some atom names often need to be changed when generating structures in one program and using another for the simulation because, although there is a standard PDB format, slight variations from software to software are still common. Different programs can also be used to generate the initial three-dimensional model of the ligand. Alternatively, it can be obtained from the Cambridge Structure Databasse or the PDB if a structure of the ligand-DNA complex has already been reported. The selection of the starting position of the ligand in the DNA: ligand complex is usually based on experimental evidence such as NMR NOEs and DNase I footprinting. Moreover, the ligand′s mode of DNA recognition (e.g. an intercalator or groove binder) is often known and this reduces the uncertainty of binding location.

1.6.4.5 Generation of input files for molecular dynamics simulations in Amber

The Amber molecular dynamic package is widely used in the simulation of nucleic acids and their complexes. In addition to the force field parameters of nucleic acids and proteins, it includes a set of generic force field parameters for organic molecules called GAFF[125] (the general Amber force field). The use of this force field is recommended for systems containing small organic molecules such as in ligand-DNA interactions. GAFF has parameters for organic molecules made of C, N, O, H, P, S, F, Cl, Br and I. The compatibility of GAFF with other Amber force fields makes it suitable to study organic molecules that bind to DNA. The antechamber program, which is one of the preparation programs in the Amber suite, uses the three-dimensional structure of the ligand as an input file and automatically assigns charges, atom types, and force field parameters. The program is then used to generate prep (internal coordinate file) and frcmod (force field file) files for the ligand to be read by Leap (the basic preparation program for Amber simulations).

Once an initial three dimensional structure of the complex is generated (as a pdb file), the structure, in addition to the prep and frcmod files of the ligand, are loaded into Leap to solvate the system, add counterions, and then construct the necessary input files (parameter/topology ″prmtop″ and coordinate ″inpcrd″ files) for running Sander, the main molecular dynamics program supplied with Amber.

The NMR distance restraints used during the molecular dynamic simulations are derived from the 2D NOESY data. The Sparky program[126] is usually used to convert the volume of NOESY cross-peaks to an intensity file (INT.1). This file in addition to the PDB file of the starting model are used as an input in a program such as Mardigras to generate the NOE distance restraints file (dist) which is used as input restraints during the molecular dynamic simulations.

1.7 Aims of the project

The aim of this project is to investigate the factors that dictate lexitropsin-DNA associations by studying the thermodynamic binding characteristics and the structural features of these complexes using different biophysical techniques. In this study, isothermal titration calorimetry (ITC) and circular dichroism (CD) were used to obtain a complete thermodynamic profile for lexitropsin interactions with different ODN sequences and that included the determination of the binding affinity (K), stoichiometry (n), enthalpy (ΔH), entropy (ΔS) and free energy of binding (ΔG).The thermodynamic studies aimed to reveal the molecular forces that drive the binding and to establish a link between the structure and the binding affinity by studying thermodynamic binding characteristics of closely related ligand structures to a specific binding site.

NMR spectroscopy and molecular modeling were used to obtain detailed structural information for lexitropsin-DNA complexes. Of particular interest to our study is nuclear Overhauser enhancement spectroscopy (NOESY), a technique that can probe the distances between atoms, and thus be used to determine the three dimensional structure of molecules through the restrained molecular dynamic simulations. When these structural studies are combined with thermodynamic data, a complete picture for the interaction can be revealed that allows potential ligands to be developed based on a rational approach.

2 CHAPTER 2: Experimental

2.1 *Isothermal titration calorimetry*

2.1.1 Chemicals

The self-complementary oligodeoxynucleotides (ODNs) 5´- GCG**CCTAGI**CGC-3´, 5´-GCG**ACTAGT**CGC-3´, 5´-GCG**TCTAGA**CGC-3´, 5´-GCG**CCTAGG**CGC-3´, 5´-GCG**GCTAGC**CGC-3, 5´-GCG**ACATGT**CGC-3´, 5´-GCG**ACGCGT**CGC-3´, and 5´-GCG**TCGCGA**CGC-3´, and the non-complementary oligodeoxynucleotides (ODNs) 5´-GCG**ACAGT**CGC-3´ and 5´-GCG**ACTGT**CGC-3´ were purchased from MWG-BIOTECH AG (Anzinger str. 7a, D-85560 Ebersberg, Germany) as HPLC-purified salt free ODNs, custom synthesized on the 1 μmol scale. The MGBs were prepared as described previously.[127] Millipore-filtered water was used in the preparation of all the solutions. Piperazine-*N*,*N*'-bis(2-ethane sulfonic acid (PIPES), 2-(carbamoylmethylamino) ethansulfonic acid (ACES), ethylenediaminetetraacetic acid (EDTA), and sodium chloride (NaCl) were used to prepare the buffers and purchased from Sigma Aldrich (Poole, Dorset, UK).

2.1.2 Sample preparation

The DNA was prepared by annealing the ODN samples at 90 °C for 12 min, then gradually cooled to room temperature. The ligands and the DNA dodecamers were dissolved in degassed 0.01 M PIPES, 0.02 M NaCl, and 0.001 M EDTA which had been adjusted to pH 6.8. All the solutions were degassed for 20 min before use in a desiccator or in a Microcal Thermovac sample degasser to decrease the noise and obtain a stable baseline. The concentrations of the DNA solutions were determined spectroscopically at λ_{260} using the OD values supplied by the manufacturer. For all DNA sequences, aliquots were taken and diluted to achieve the concentration required for the ITC experiments (15 μM). The ligand solution was prepared as a 0.5 mM solution.

For the ligand self-association study, the MGBs were dissolved in degassed PIPES (0.01 M PIPES , 0.02 M NaCl, 0.001 M EDTA) or ACES (0.01 M ACES , 0.02 M NaCl, 0.001 M EDTA) buffers and the analyses were performed at ligand concentrations of 0.5 mM.

2.1.3 ITC experiments

2.1.3.1 Ligand-DNA titrations

ITC was performed at 25 °C using a Microcal VP-ITC (Microcal Inc., Northampton, USA). The control units were interfaced to PCs equipped with the Origin software package for data manipulation and instrumental control. The DNA concentration in the sample cell was 15 µM and the ligand concentration in the syringe was 0.5 mM. Mixing was carried out by stirring the sample cell at 329 revolutions per minute. A 280 µL rotating syringe with an impeller profiled needle was used to perform 25 repeat 10 µL injections of the ligand with a 300 s delay between the first five injections, a 600 s delay between the subsequent fourteen injections and a 300 s delay between the last six injections. To correct for the heat of dilution of the ligand, control experiments were performed at the same temperature using similar conditions with buffer only. The heats of ligand dilution were subtracted from the subsequent heat obtained for the titration of the DNA dodecamers with the ligand, thereby yielding the heat of binding for the ligand–DNA complexes.

2.1.3.2 Ligand dilution experiments

ITC dilution studies were performed at three different temperatures (i.e. T= 25 °C, 35 °C, 45 °C) using a Microcal VP-ITC (Microcal Inc., Northampton, USA). All the dilution experiments were set up so that 10 µL 0.5 mM ligand solutions were added to buffer in the sample cell every 300 s up to a total of 25 injections. Mixing was carried out by stirring the sample cell at 329 revolutions per minute. The binding constant K and the enthalpy ΔH of the ligand self-association could not be obtained using the Origin package coupled with the ITC instrument because it is not capable of treating data for self –assembly directly. For this reason the software package IC-ITC[128] was used for data analysis.

2.1.4 Data analysis

The heat change upon addition of ligand into the DNA solution is proportional to the power represented by the area under each peak and is given by integration of this peak with respect to time, which gives the enthalpy change as a result of injection. Prior to analysis, data were corrected for the heat of dilution of the ligand into the buffer alone without DNA. The binding isotherms were then fitted to models by nonlinear least squares analysis (Origin7.0, Microcal). The Origin algorithim allows fitting to a one or two independent binding sites model. The model for one independent set of binding sites works for any number of sites, N, if all sites have the same binding constant (K) and enthalpy change (ΔH). If a macromolecule has two separate sites with different K and ΔH values, then the two independent sets of binding sites model must be used. The model for one independent set of sites was applicable for all the data generated from the ITC titrations performed in this study.

In the ligand self-assembly study, the raw data were treated with Origin to generate both integrated heat effects per injection (dh) and molar heat effects per injection (ndh). These heat effects were used to generate a dh file (contains the dh and ndh data in two parallel columns) as a notepad file format in a Microsoft Windows environment. This file, together with the file containing information on the injected volumes used during the dilution experiment (volume file, vol), served as input for Prep4icitc (preparation program for IC ITC) to generate the necessary files for running IC-ITC to analyse the data.

2.2 Circular Dichroism

2.2.1 CD experiments

CD experiments were conducted using an Applied Photophysics Chirascan spectrophotometer at 25 ^{o}C in a 1 cm quartz cuvette. To the ODNs (5 μM, 1.0 mL) in pH 6.8 PIPES buffer (10 mM PIPES, 20 mM NaCl and 1 mM EDTA) ligand solutions were added (0.5 mM) in 1 μL increments to a total of 20 additions. At each titration point the molar ellipticity was measured between 240 to 380 nm using a bandwidth of 1 nm. Binding constants were calculated by non-linearing least squares fitting of Engel′s equation[106] for tight ligand binding to the CD data (Eq.1.10).

2.3 NMR Spectroscopy

2.3.1 Chemicals

The self complementary ODNs d(CG**ACGCGT**CG)$_2$ and d(CG**ACTAGT**CG)$_2$, were supplied by Alpha DNA Ltd. (Montreal, Canada) as desalted, cartridge-purified, ethanol precipitated, lyophilised powders which were used without further purification. *N*-(3-(dimethylamino)propyl)-5-isopropyl-2-(1-methyl-4-(1-methyl-4-(nicotinamido)-1*H*-pyrrole-2-carboxamido)-1*H*-pyrrole-2-carboxamido) thiazole-4-carboxamide (AIK18-51) and 2-(4-(4-acetamido-1-methyl-1*H*-imidazole-2-carboxamido)-1-methyl-1*H*-pyrrole-2-carboxamido)-*N*-(3-(dimethyl amino) propyl)-5-isopropylthiazole-4-carboxamide (thiazotropsin B) were prepared as the TFA salts as described previously.[129]

2.3.2 General NMR sample preparation.

For ′baseline′ data set accumulation, the ODNs were typically dissolved to a concentration of ca. 2 mM in 1100 μL phosphate buffer (pH=7.4) (90% H_2O/ 10% D_2O) and half of this solution was admitted to a 5 mm NMR tube (Wilmad, USA., tube code: 528-PP-7) together with 0.5 μL of stock TSP solution. The remaining solution was freeze dried to remove the solvent and was suspended in 550 μl of D_2O phosphate buffer to examine the free ODN in D_2O solvent.

Following the titration of the ODNs dissolved in 90% H_2O/10% D_2O in phosphate buffer (pH =7.4) by the ligand (thiazotropsin B or AIK18-51), the solution was frozen, freeze-dried and then redissolved in 100 % D_2O to examine the complex in the D_2O solvent.

2.3.3 Complex formation between thiazotropsin B and d(CGACGCGTCG)$_2$

Thiazotropsin B (3.70 mg) was dissolved in 100 μL of 90 % H_2O/10 % D_2O in phosphate buffer to provide a stock solution of ligand at a concentration of 55.0 mM . d(CGACGCGTCG)$_2$ was dissolved in 550 μL of 90% H_2O/10% D_2O in phosphate buffer to give a concentration of 2mM. Thiazotropsin B (40 μL) was added to d(CGACGCGTCG)$_2$ in 10-μL aliquots using a Hamilton syringe. The solution was gently mixed to disperse the immediate precipitate that formed on the first contact between the ligand and DNA solutions. 1D ^{1}H NMR spectroscopy was used to confirm the status of the sample at each stage of ligand addition. 2:1 ligand:DNA complex formation was typically achieved following addition of a total of 40 μL stock ligand solution. A titration end point was apparent when complete disappearance of free DNA imino proton ^{1}H NMR resonances was noted together with simultaneous replacement by a new set of imino proton ^{1}H NMR signals. In this typical state, the sample was used in the accumulation of complete NMR data sets for the purposes of structure determination.

2.3.4 Complex formation between AIK18-51 and d(CGACTAGTCG)$_2$

AIK18-51 (4.04 mg) was dissolved in 100 μL of 90 % H_2O/10 % D_2O in phosphate buffer to provide a stock solution of ligand at a concentration of 55.0 mM . d(CGACTAGTCG)$_2$ was dissolved in 550 μL of 90% H_2O/10% D_2O in phosphate buffer to give a concentration of 2 mM. The procedure described above for thiazotropsin B was then repeated.

2.3.5 NMR spectroscopy experiments

NMR data were acquired on a Bruker AVANCEIII 600 NMR spectrometer operating at 600.13 MHz for H-atom resonance by Dr John Parkinson, Department of pure and applied chemistry, University of Strathclyde. A standard geometry triple-resonance probehead equipped for z-pulsed field gradients was used. Data acquisition was carried out in an identical manner for free ODNs ($d(CGACGCGTCG)_2$ and $d(CGACTAGTCG)_2$) and for the complexes formed between thiazotropsin B and AIK18-51. The probe temperature was 298 K in all instances. One-dimensional ^1H NMR data were acquired using either presaturation (zgesgp pulse program) or a double-pulsed-field-gradient-spin-echo (dpfgse)[130] approach to eliminate the solvent resonance. At 600 MHz, data were typically acquired using digital signal processing with 128 scans over a ^1H frequency width equivalent to 20.0276 ppm centred at 4.692 ppm into 32 K data points (acquisition time: 1.09 s) using a 90° hard pulse (P1= 10.16 μs) and a relaxation delay of 2.0 s. For the dpfgse routine, rectangular soft pulses (bandwidth = 125 Hz) were used for selective inversion at the solvent frequency together with sine-shaped gradient pulses (1 ms duration) in a ratio of 31:11.

Two-dimensional (2D) NMR data sets were acquired as follows: 2D [^1H, ^1H] DQFCOSY NMR data (pulse program: cosydfphpr) were acquired using solvent signal presaturation with 16 transients for each of the 512 States t_1 increments over a ^1H frequency width of 12.0166 ppm in both ω_2 and ω_1 into 4 K complex data points (acquisition time 372 ms) with a recycle time of 2 s for a total data accumulation time of ca. 4.5 h.

2D [^1H, ^1H] TOCSY NMR data were acquired using the dipsi2esgpph pulse program with 16 transients for each of the 512 States-TPPI t_1 increments over an ω_2 frequency width of 20.04 ppm and an ω_1 frequency width of 12 ppm into 8 K complex data points (acquisition time 341 ms) with a recycle time of 2 s (D9= 0.07 s) and a pulsed spin-lock mixing time of 55 ms for a total data accumulation time of 4.5 h.

2D [^1H, ^1H] NOESY NMR data used for assignment purposes were acquired using the noesyesgpph pulse program with 32 transients for each of the 1024 States-TPPI t_1 increments over an ω_2 frequency width of 20.04 ppm in both ω_2 and ω_1 into 8 K complex data points (acquisition time 341 ms) with a recycle time of 2 s (D8 = 0.15 s) and a mixing time of 100 ms for a total accumulation time of ca. 22 h.

One-dimensional ^{31}P-{^{1}H} NMR data were acquired at 161.977 MHz (9.4 T magnetic field) using 64 transients over a frequency width of 810 Hz (5 ppm) into 512 data points (acquisition time 316 ms) with a recycle time of 0.5 s and centered close to the center of the DNA phosphate resonance envelop. GARP composite pulse decoupling was used for ^{1}H decoupling during the acquisition time only.

2D [^{31}P, ^{1}H] correlations were acquired using an inverse INEPT 2D sequence with TPPI (pulse program hxineptp). Transients (80) were acquired for 80 t_1 increments into 128 data points over a ω_2 (^{31}P) frequency width of 486.4 Hz (3 ppm) and an ω_1 (^{1}H) frequency width of 1200 Hz. The evolution period 1/4J was varied for different experiments in order to select for differently sized couplings. J values were chosen in order to compromise between low optimization of couplings and loss of signal during the evolution period through T_2 relaxation. ^{31}P NMR data were referenced indirectly according to Maurer and Kalbitzer.[131]

Natural abundance 2D [^{1}H, ^{13}C] HSQC NMR data were acquired at 14.1 T using gradient coherence selection. Data were acquired with 128 transients over a ^{1}H (ω_2) frequency width of 7 kHz into 2 K complex data points (acquisition time 146 ms) for each of 512 States-TPPI t_1 increments over a ^{13}C (ω_1) frequency width of 16 kHz using a recycle time of 2 s between transients. All NMR data were processed on a Dell Precision 340 workstation running under Microsoft Windows 2000 using Xwin-nmr (version 3.5, Bruker Biospin, Karlsruhe, Germany) with appropriate processing parameters and imported into SPARKY (version3.111)[126] for full data analysis.

2.3.6 NMR data assignment strategy

2D [^{1}H, ^{1}H] NOESY NMR data were assigned for both the free DNA duplexes and the ligand-DNA complexes by using established assignment strategies[108] for right-handed B-form DNA. 2D [^{1}H, ^{1}H] DQFCOSY NMR data were used to establish resonance assignments for specific protons of the ligands in their complexes with the DNA duplexes. ^{31}P resonance assignments were achieved on the basis of observed correlations between ^{31}P and H3′, H4′, and H5′/H5″ resonances. 2D [^{1}H, ^{13}C] HSQC

NMR data assisted with resolving assignment ambiguities and for establishing geminal proton resonance pairings.

2.4 *Molecular modeling*

2.4.1 Generation of starting models

Molecular models for thiazotropsin B and AIK18-51 complexes with d(CGACGC GTCG)$_2$ and d(CGACTAGTGC)$_2$, respectively were generated in Sybyl 6.3 using the Tripos 5.4 force field[132] running on Silicon Graphics work stations. These models were created on the basis of NOE data in which a head-to-tail pair of ligand molecules were bound in the minor groove, in order to construct a crude starting model for both complexes. The free DNA models were constructed using the Biopolymer module of Sybyl 6.3 from standard B-form DNA, which is consistent with the NMR data (as indicated by the presence of imino proton 1H NMR resonances in the NMR data of free and bound DNA).

2.4.2 Generation of input files

The NMR distance restraints were derived from the 2D NOESY data. The Sparky program was used to generate an intensity file (INT.1) from the volume of NOESY cross peaks. This file in addition to the PDB file of the starting model was used as an input in a program called Mardigras to generate the NOE distance restraints file (dist) which was used as an input restraints during the molecular dynamic simulations.

The generated PDB files of the starting models were also used to construct the necessary input files; parameter/topology (prmtop) and coordinate (inpcrd) files to perform minimization and restrained molecular dynamic simulations using Sander the main program supplied with Amber 10. The input files were created using Leap, the basic preparation program for Amber simulations. Antechamber was used to generate prep (internal coordinate file) and frcmod (force field file) files for the ligands (thiazotropsin B and AIK18-51) to be read by Leap using GAFF.

2.4.3 NMR structure refinement

To obtain accurate starting structures for the restrained molecular dynamic calculations in explicit water, the free decamers and complexes were firstly refined in vacuum using Sander. Before the restraint MD production phase in vacuum, minimisation was carried out to relax the systems and relieve any unfavourable clashes between atoms. Here minimization (imin=1) was performed for 500 steps (maxcyc) (250 steps of steepest descent, SD, (ncyc = 250) followed by 250 steps of conjugate gradient, CG), using a nonbonded cutoff of 12 Å (cut), and no periodic boundary (ntb=0) (the sander input files can be found in appendix 3 & 4). During the molecular dynamics production phase, the Langevin dynamics approach with a collision frequency of 1 ps^{-1} was used at a constant temperature (300 K) for 50 ps using a time step of 1 fs. The production phase was allowed to extend for a total time of 400 ps. During molecular dynamic simulations, the distance restraints file derived from the NOE data was used to refine the structure. The final restart file was used to create a pdb file which was used as a starting structure for carrying out the final NMR structure refinement in explicit solvent. The simulated systems were neutralized by the addition of 16 Na+ and 18 Na+ counterions for the complex and the free DNA, respectively. Each system was placed in a periodic octahedral box solvated with TIP3P water with outer edges approximately 10 Å in each direction from the closest solute atom. Periodic boundary conditions with a 12 Å cutoff for nonbonded interactions were applied, with the particle mesh Ewald (PME) method[133] applied to account for the long range electrostatic interactions.

Before the MD production phase, minimization and equilibration were carried out in three stages as follows. (i) The solvent and ions were minimized whilst the DNA and the ligand dimer were restrained by a force constant of 500 kcal mol^{-1} $Å^{2}$ for 1000 steps (500 steps of steepest descent, SD, (ncyc = 500) followed by 500 steps of conjugate gradient, CG), using a nonbonded cut off of 10 Å (cut), and periodic boundary (ntb=1); intermediate results were printed every 50 steps (ntpr). (ii) Next, the restraints on the DNA and the ligand dimer were removed and the whole system was minimised for 2500 steps (1000 steps of steepest descent, SD, (ncyc = 1000) followed by 1500 steps of conjugate gradient, CG), using a nonbonded cutoff of 10 Å (cut), and periodic boundary conditions (ntb=1). (iii) Equilibration was then performed for 20 ps and the NOE distance restraints were applied. The system was heated using the NVT ensemble and Langevin dynamics with a collision frequency of 1 ps^{-1} from 0 to 300 K over 20 ps. The MD production phase involved the NVT ensemble at a constant temperature and pressure (300 K and 1 atm, respectively) for 100 ps using a time step of 2 fs and the SHAKE[124] algorithm to constrain hydrogen to heavy atom bonds. The MD calculations were repeated three times using the restart file from the previous run. The NMR distance restraints were applied during these calculations. The trajectories were analysed using ptraj and an average pdb structure created from the MD production phases where the energy of the system has converged. The MD calculations were performed on the UK National Grid Service computers using 16 CPUs in parallel.

2.4.4 Structure analysis

The program CURVES (version 5.2)[134] was used to analyse the resulting structures by applying the helicoidal parameters used to describe a nucleic acid duplex, as defined according to the EMBO workshop on DNA curvature and bending.

3 CHAPTER 3: Thermodynamics of lexitropsin-DNA interactions

3.1 ITC results for thiazotropsin A

3.1.1 Sequence recognition

The thermodynamics of binding by thiazotropsin A to seven dodecamers (d[5´-GCG**ACTAGT**CGC-3´]$_2$, d[5´-GCG**TCTAGA**CGC-3´]$_2$, d[5´-GCG**GCTAGC** CGC-3´]$_2$, d[5´-GCG**CCTAGG**CGC-3´]$_2$, d[5´- GCG**CCTAGI**CGC-3´]$_2$, d[5´-GCG**ACATGT**CGC-3´]$_2$ and d[5´-GCG**ACAGT**CGC-3´] were examined using ITC to provide insight into the energetic basis for recognition and affinity by the ligand. Using NMR, we had previously found that thiazotropsin A recognises a 5′-ACTAGT-3′ sequence embedded in the decameric duplex 5′-CGACTAGTCG-3′ as a dimer, side by side, in a head to tail fashion (Figure 3.1).[38]

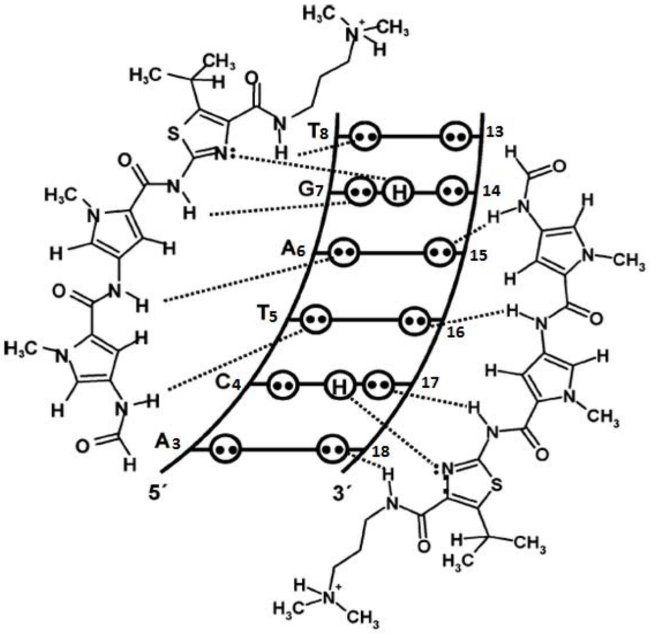
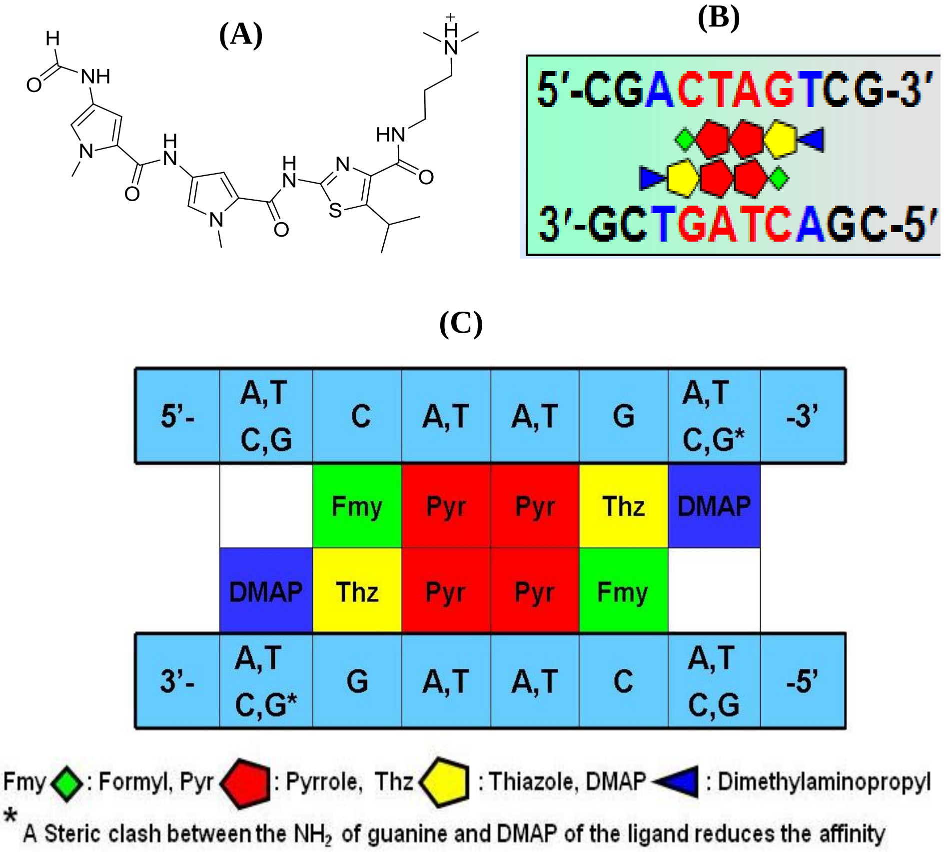


Figure 3.1 Deduced arrangement of hydrogen bonding between thiazotropsin A and d(CG**ACTAGT**CG)$_2$ showing the ligand binds to the minor groove of DNA as an antiparallel dimer.[38]

Our latest NMR and ITC studies reported herein have revealed that the dimeric association of thiazotropsin A exists in two configurations; the slipped dimer and overlapped dimer. The slipped configuration has two of the three aromatic rings from each ligand stack on the top of each other, allowing the ligand to read six base pairs (Figure 3.2.). The ligand binds specifically to the central four base pairs (CTAG) whilst the other two bases at the edge of the binding site are covered by the dimethylaminopropyl (DMAP) tail, which cannot differentiate between the four DNA bases. The pyrrole ring of each ligand recognizes A/T bases and does not distinguish between them. Thus, reversal of the position of the central T/A bases to 5′-AC**AT**GT-3′ means the ligand can still can bind (Figure 3.2.C).



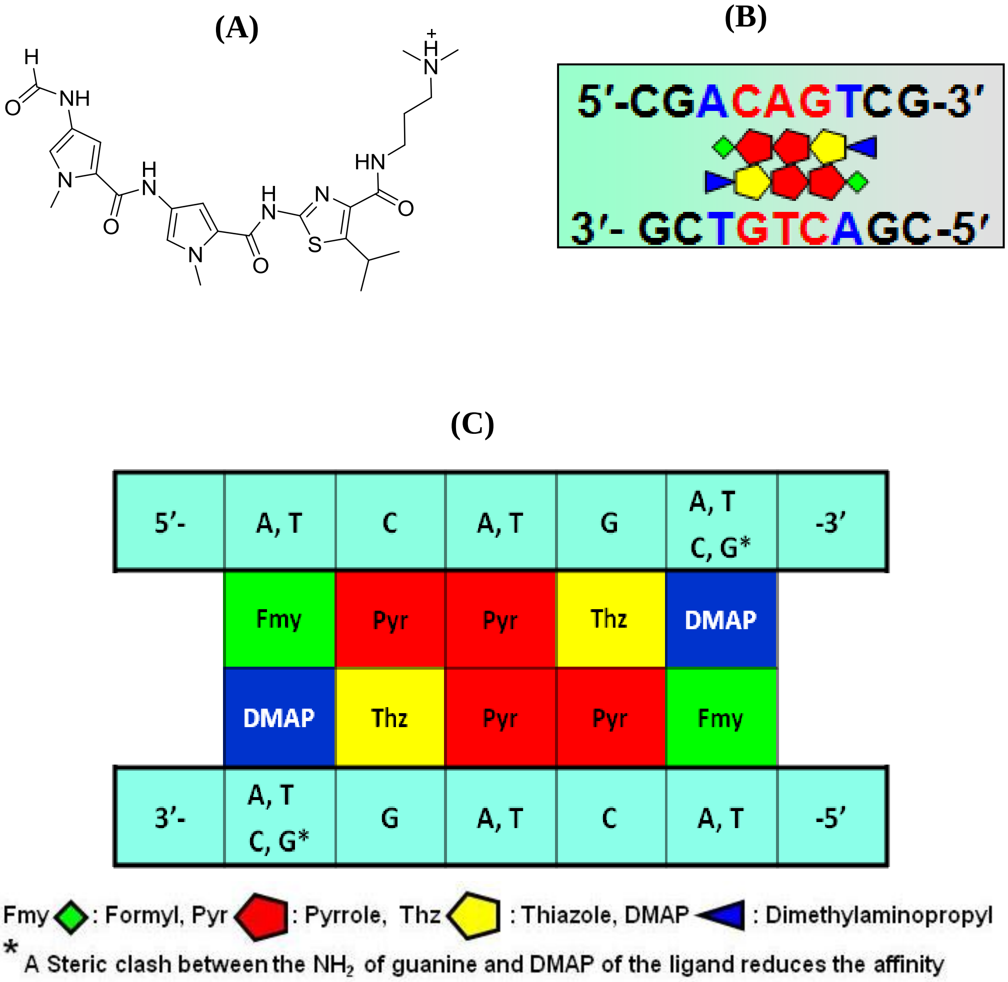
5’-	A,T C,G	C	A,T	A,T	G	A,T C,G*	-3’
		Fmy	Pyr	Pyr	Thz	DMAP	
	DMAP	Thz	Pyr	Pyr	Fmy		
3’-	A,T C,G*	G	A,T	A,T	C	A,T C,G	-5’

Fmy ◆: Formyl, Pyr ⬟: Pyrrole, Thz ⬡: Thiazole, DMAP ◀: Dimethylaminopropyl

* A Steric clash between the NH_2 of guanine and DMAP of the ligand reduces the affinity

Figure 3.2 Schematic representation for thiazotropsin A (slipped dimer) associations with DNA (6 base pairs). **(A)** Structure of thiazotropsin A, **(B)** a schematic shows thiazotropsin A- ACTAGT association, and **(C)** a schematic shows all the possible DNA recognitions by the slipped dimer of thiazotropsin A**.**

In the overlapped configuration, all the aromatic rings of the ligands are stacked on the top of each others, which enables the ligand dimer to recognise a shorter five base pair sequence (5′-ACAGT-3′) (Figure 3.3). The Pyr/Thz pair recognises a C-G base pair, whereas the Thz/Pyr combination targets a G-C base pair. Both the Pyr/Pyr pair and the Fmy/DMAP pair do not distinguish T-A from A-T base pairs. The Fmy/DMAP pair binds C-G base pairs with a reduced affinity due to a steric clash between the NH_2 of guanine and the DMAP of the ligand (Figures 3.2 & 3.3).



5′-	A, T	C	A, T	G	A, T C, G*	-3′
	Fmy	Pyr	Pyr	Thz	DMAP	
	DMAP	Thz	Pyr	Pyr	Fmy	
3′-	A, T C, G*	G	A, T	C	A, T	-5′

Fmy ◆ : Formyl, Pyr ⬟ : Pyrrole, Thz ⬡ : Thiazole, DMAP ◀ : Dimethylaminopropyl

* A Steric clash between the NH_2 of guanine and DMAP of the ligand reduces the affinity

Figure 3.3 Schematic representation for thiazotropsin A overlapped dimer associations with DNA. **(A)** Structure of thiazotropsin A, **(B)** a schematic shows thiazotropsin A-ACAGT association, and **(C)** A schematic shows all the possible DNA recognitions by thiazotropsin A overlapped dimer.

3.1.2 Thermodynamics of thiazotropsin A-DNA associations

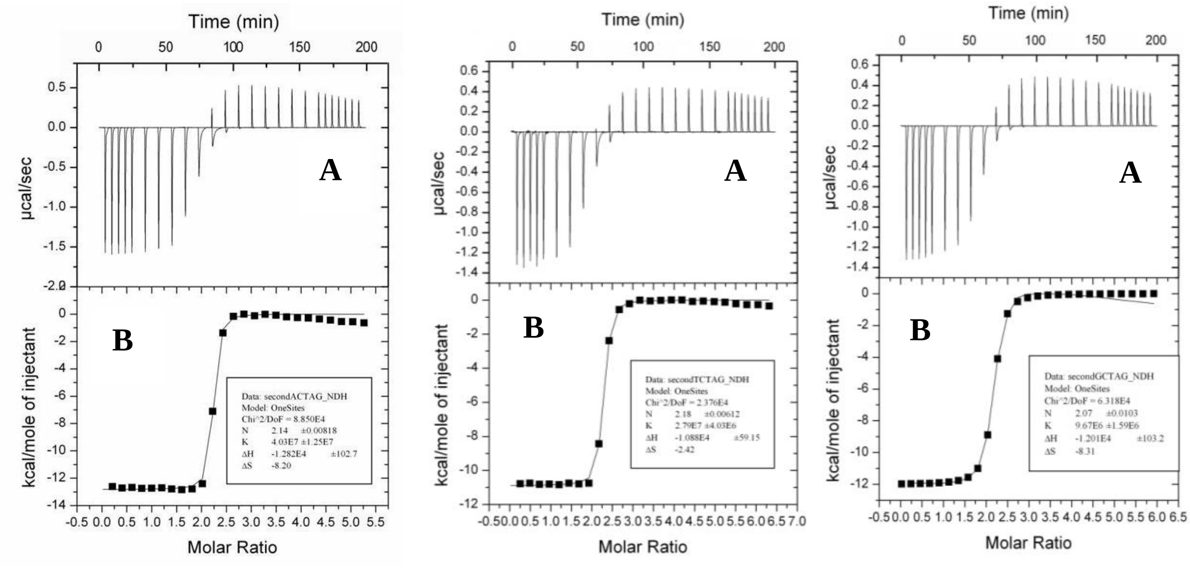
Titration of thiazotropsin A with the seven dodecamers in PIPES buffer at 25 °C was clearly associated with an exothermic process (Figures 3.4 & 3.6). Figure 3.4 (A panels) shows representative ITC profiles resulting from the injection of thiazotropsin A with the DNA dodecamers. Each of the heat burst curves in Figure 3.4 (A panels) corresponds to a single drug injection. The areas under these heat burst curves were determined by integration to yield the associated injection heats. These injection heats were corrected by subtraction of the corresponding dilution heats derived from the injection of identical amounts of drug into buffer alone. Dilution peaks (Figure 3.5) were all endothermic, and their intensity decreased as more ligand was added, indicating that aggregation of the thiazotropsin A occurs in buffered solution (ligand self-aggregation will be discussed in section 3.2.7). Figure 3.4 (B panels) shows the resulting corrected injection heats plotted as a function of the [ligand]/[duplex] ratio. In this figure, the data points reflect the experimental heats while the solid lines reflect the calculated fits of the data. The injection heat data corresponding to the titration of the seven dodecamers with thiazotropsin A were fitted with a model for one binding site to yield a complete thermodynamic profile for the binding interaction, which includes the determination of binding affinity, K, binding enthalpy, ΔH, binding free energy, ΔG, and entropy changes, ΔS. The fits of the ITC data shown in Figure 3.4 (B panels) were derived using ΔH and K as free floating parameters. The obtained value of K was used to calculate the binding free energies (ΔG) using the following standard relationship:

$$\Delta G = -RT\ln(K) \qquad \text{Eq. 1.7}$$

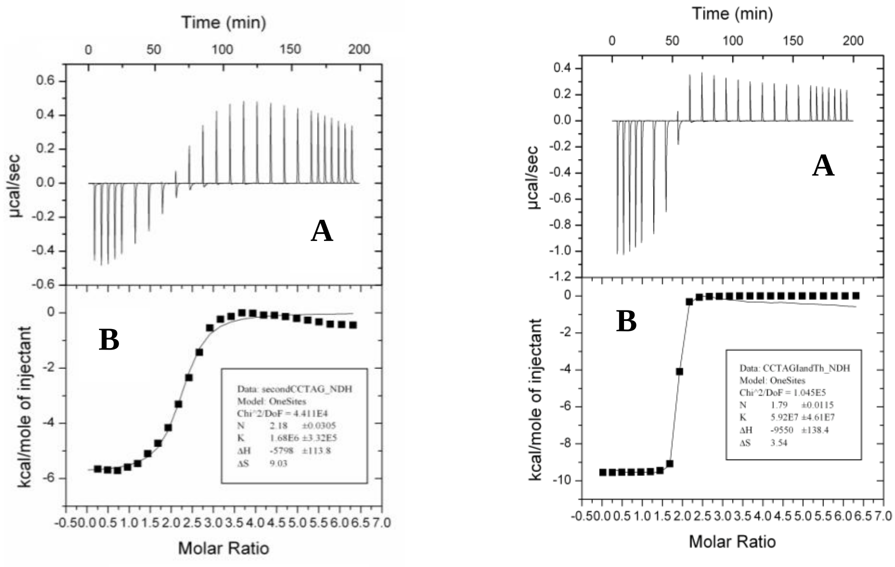
The binding free energies, coupled with the binding enthalpies derived from the fitted ITC data allowed the determination of the corresponding entropic contributions to binding ($T\Delta S$) using the standard relationship:

$$T\Delta S = \Delta H - \Delta G \qquad \text{Eq. 1.8}$$

The thermodynamic binding parameters were derived automatically using the Origin package software.



(1) ACTAGT **(2) TCTAGA** **(3) GCTAGC**



(4) CCTAGG **(5) CCTAGI**

Figure 3.4 The ITC titrations of thiazotropsin A to five DNA sequences in PIPES buffer at 25 °C (pH 6.8). **(A)** Raw data for the titration of thiazotrosin A into: **(1)** GCG**ACTAGT**CGC, **(2)** GCG**TCTAGA**CGC, **(3)** GCG**GCTAGC**CGC, **(4)** GCG**CCTAGG**CGC, and **(5)** GCG **CCTAGI**CGC. **(B)** Enthalpogram retrieved from A, corrected for the heat of dilution; the line represents the least-squares-fit to the single-site binding model.

The experiments were carried out at low salt concentration in order to avoid the competition between the ions of salt and thiazotropsin A molecules, which would be more pronounced at higher salt concentration. The thermodynamic parameters of thiazotropsin A titrations into five dodecamers have been summarised in Table 3.1. These results represent means ± SEM of duplicate experiments. Significant heat changes were observed upon addition of thiazotropsin A to the five ODN duplexes, which is characteristic of tight minor groove binding.

The interaction of thiazotropsin A with the 5′-ACTAGT-3′ containing sequence involved one binding process (Figure 3.4 (**1**)). Our analysis generated values for ΔG of -10.2 kcal mol^{-1}, ΔH of -12.9 kcal mol^{-1}, ΔS of -9.1 cal mol^{-1} K^{-1} ($T\Delta S$ = -2.7 kcal mol^{-1}), K of 3.0 x10^{7} M^{-1} and a binding stoichiometry of 2:1 (ligand to DNA). These results indicated that the binding interaction was enthalpically driven via hydrogen bonding and/or van der Waals interactions, and that there was an entropic penalty associated with the complexation process.

Figure 3.4 (**2**) and Figure 3.4 (**3)** show the enthalpograms for the titration of thiazotropsin A to dodecamers containing the central sequences, 5′-TCTAGA-3′ and 5′-GCTAGC-3′, respectively. The interaction of thiazotropsin A with both duplexes showed remarkably similar binding behavior to that observed with the 5′-ACTAGT-3′ sequence. The interactions were enthalpically driven, implying that the hydrogen bonding and van der Waals forces play an important role in the binding process. Our results revealed that the binding affinities of thiazotropsin A with DNA duplexes were in a decreasing order of ACTAGT> TCTAGA> GCTAGC> CCTAGG. A relatively weak binding was observed with 5′-CCTAGG-3′ (ΔG –8.5 Kcal mol^{-1}; Figure 3.4 (**4**)) compared with other sequences due to the presence of the guanine NH_2, which sterically impedes the ligand from being fully accommodated into the minor groove of DNA. This observation was confirmed by using an equivalent dodecamer containing inosine instead of guanine, which has no exocyclic amino group protruding from the groove floor. This change was accompanied by a significant increase in binding affinity from a ΔG of -8.5 Kcal mol^{-1} to -11.1 Kcal mol^{-1} (Figure 3.4 & Table 3.1).

Table 3.1 Thermodynamic parameters for the interaction of thiazotropsin A with five DNA sequences in 10 mM PIPES buffer at 25 °C (6.8 pH) presented as mean±SEM of duplicate experiments.

Binding site	***ΔH*** ***(Kcal.mol^{-1})***	***TΔS*** ***(Kcal.mol^{-1})***	***K*** ***(M^{-1})***	***ΔG*** ***(Kcal.mol^{-1})***
GCG**ACTAGT**CGC	-12.8 ± 0.03	-2.7 ± 0.3	$3.0x10^{7}$±$1.1x10^{7}$	-10.2 ± 0.2
GCG**TCTAGA**CGC	-10.4 ± 0.5	-0.3 ± 0.2	$2.0x10^{7}$± $0.8x10^{7}$	-9.9± 0.2
GCG**GCTAGC**CGC	-12.0 ± 0.2	-2.5± 0.4	$9.6x10^{6}$± $0.9x10^{6}$	-9.5 ± 0.1
GCG**CCTAGG**CGC	-5.5 ± 0.3	3.03 ± 0.3	$1.7x10^{6}$± $0.03x10^{6}$	-8.5 ± 0.01
GCG**CCTAGI**CGC	-9.6 ± 0.4	1.5 ± 0.4	$1.1x10^{8}$± $0.5x10^{8}$	-11.1 ± 0.1

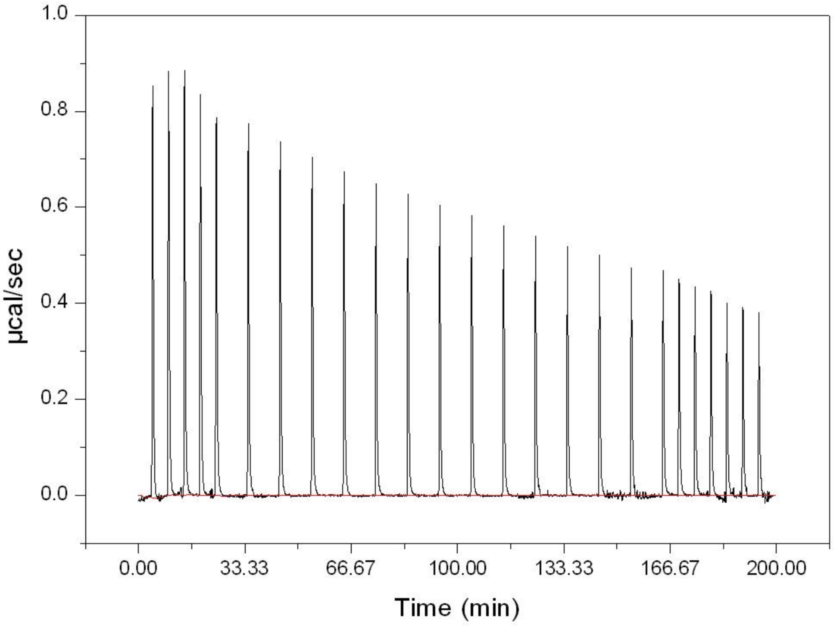


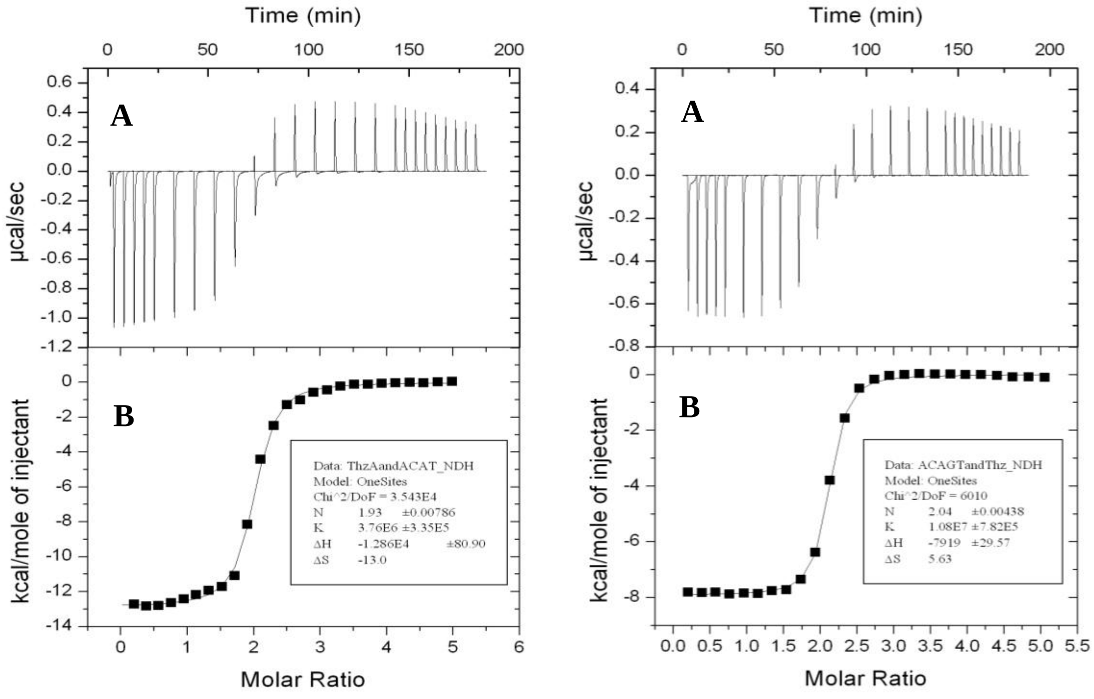
Figure 3.5 ITC heats of dilution of thiazotropsin A in PIPES buffer. Each heat burst curve is the result of a 10 µL injection of 0.5 mM ligand into the PIPES buffer solution. The solution conditions were 10 mM PIPES, 20 mM NaCl, and 1 mM EDTA.

Analysis of the heat effects resulting from the binding of thiazotropsin A to the five dodecamers at ligand:DNA ratios (r) varying between 0 and 7 (Figure 3.4) revealed that when $r \leq 2$, the binding enthalpy remained constant. A dodecamer can potentially provide two separate binding sites for a small molecule with dimensions similar to Hoechst 33258,[135] which raises the question: ′does thiazotropsin A bind to two individual binding sites as a monomer or to one site as a dimer?′. If monomeric binding occurred in relation to different base pair sequences of the two potential binding sites, a noticeable difference in the enthalpy of binding would be observed as the first and subsequently the second site was occupied. Since the measured ΔH values remained constant when $r \leq 2$, we propose that in this range, the ligand binds to the dodecameric DNA sequence exclusively in a non-stepwise dimeric 2: 1 mode, which is consistent with our previous NMR study.[38]

We attempted to assess the generality of the pyrrole blocks in recognising A·T and T·A base pairs by reversing the position of the central TA bases of the 5′-AC**TA**GT-3′ sequence. ITC results showed that the ligand can still bind to 5′-AC**AT**GT-3′ (Figure 3.6 (**1**)), but with slightly lower affinity. The binding free energy for thiazotropsin A-ACATGT complexation is -8.986 Kcal mol^{-1}, which is lower than the corresponding binding energy of the thiazoropsin A-ACTAGT association (-10.15 Kcal mol^{-1}). The initial assumption would be that the change in the position of the central DNA bases (A and T) would change the geometry of the hydrogen bond contacts between the NHs of the ligands and the acceptors at O2 and N3 of thymine and adenine, respectively. The ACTAGT frame would therefore provide better hydrogen bond contacts with the ligand compared with ACATGT and explain the higher affinity. However, the enthalpy changes of the two complexes were within 0.015 Kcal mol^{-1} of each other, suggesting non-bonded interactions of similar magnitude were formed. The difference in the binding free energy arises from the higher entropic penalty of thiazotropsin A-ACATGT association compared with the thiazotropsin A-ACATGT complex (-3.874 Kcal mol^{-1} and -2.697 Kcal mol^{-1} respectively). The higher loss of entropy associated with the former complex reflects more conformational changes in either or both species in order to achieve better hydrogen bond contacts between the ligand and the DNA duplex (induced fit), which are entropically unfavourable.[84,136-137] Large unfavourable entropy is often indicative of an ″induced fit″ during the interaction process.

We also attempted to quantify the ligand′s binding to the DNA duplex as an overlapped dimer (5 base pair recognition sequence) as opposed to the six base pair sequence read by the slipped dimer (Figures 3.2 & 3.3) using ITC (Figure 3.6 **(2)** and Table 3.2). Although thiazotropsin A binds to the 5′-ACAGT-3′ binding

site with approximately 3x lower affinity (K) than the 6 base pair binding site (5′-ACTAGT-3′), the overall binding energy is around 0.5 Kcal mol^{-1} less favourable. However, there are noticeable differences in the enthalpy/entropy balance for the two complexes. For the thiazotropsin A-ACAGT complexation process, there is a significant reduction in ΔH (-7.919 kcal mol^{-1} compared with -12.845 Kcal mol^{-1}) and a favourable associated TΔS contribution (+1.677 Kcal mol^{-1} compared with -2.697 Kcal mol^{-1}). There are two possible explanations for this change in the enthalpy /entropy balance: the 6 base pair binding site may provide more hydrogen bond and van der Waals contacts with the slipped ligand dimer compared with the 5 base pair sequence and the overlapped ligand configuration. The reduction in the number of non-bonded contacts could lead to a decrease in the observed ΔH. Moreover, the bulkiness of the overlapped dimer may have the effect of expelling more bound water and/or counterions from the minor groove upon ligand binding leading to a favourable positive TΔS. This increase in ΔS itself could result in a decrease in the observed ΔH due to the enthalpy/entropy compensation phenomenon which has been discussed earlier in section 1.6.1.2.



(1) ACATGT **(2) ACAGT**

Figure 3.6 (A panels) ITC profiles at 25 °C for the titration of thiazotropsin A to the binding sites; **(1)** GCG**ACATGT**CGC [6 bp] **(2)** GCG**ACAGT**CGC [5 bp], at pH 6.8. Each heat burst curve is the result of a 10 µL injection of 0.5 mM ligand .The DNA concentration was 15µM, and the solution conditions were 10 mM PIPES, 20 mM NaCl, and 1 mM EDTA. (**B panels)** Corrected injection heats plotted as a function of the [ligand]/[DNA] ratio. The corrected injection heats were derived by integration of the ITC profiles shown in Panels A, followed by subtraction of the corresponding dilution heats derived from control titrations of drug into buffer alone. The data points reflect the experimental injection heats, while the solid reflect calculated fits of the data. The data were fit with a model for one binding site.

Table 3.2 Thermodynamic parameters for the interaction of thiazotropsin A with ACATGT and ACAGT in 10 mM PIPES buffer at 25 °C (6.8 pH).

Binding site	ΔH ($Kcal.mol^{-1}$)	$T\Delta S$ ($Kcal.mol^{-1}$)	K (M^{-1})	ΔG ($Kcal.mol^{-1}$)	N (L/DNA)
GCG**ACATGT**CGC	-12.9	-3.9	$3.8x10^{6}$	-9.0	2.0
GCG**ACAGT**CGC	-7.9	1.7	$1.1x10^{7}$	-9.6	2.0

The negative unfavourable entropy changes observed for the binding of thiazotropsin A with ACTAGT and TCTAGA, GCTAGC, and ACATGT are indicative of an ″induced fit″ interaction. The association itself is also expected to constrain the complex structure and to incur an entropic penalty due to the losses in the rotational and translational degrees of freedom when two molecules are bound together to form a complex.[138] Such induced fit and allosteric regulation are common phenemona in regulating biological processes.[139-140]

In ligand-DNA interactions, favourable (positive) entropy is mainly derived from the desolvation of the binding interface, and it appears that this was not a significant event when thiazotropsin A bound to the minor groove. The process was driven primarily by enthalpy, indicating that hydrogen bonding and van der Waals forces dominate this interaction. The formation of hydrogen bonds between the ligand and the DNA duplex counteracts the entropic contributions to the binding in two ways: firstly, restricting the degrees of freedom of both the ligand and the DNA; and secondly, trapping some water molecules at the binding interface, which reduces the desolvation of the binding site.

The NH_2 groups of guanine which protrude into the floor of the minor grooves play an important role in assisting the binding when they form hydrogen bonds with the thiazole nitrogen, or opposing it when a steric clash between NH_2 and ligand occurs. The *N*-methyl of the pyrrole and the isopropyl of the thiazole moieties do not impede the binding as they protrude outside the minor grooves, whereas, the DMAP tail of the ligand could interfere with binding through its bulkiness. The reduced affinity of thiazotropsin A with the $5'$-$C_3C_4T_5A_6G_7G_8$-$3'$duplex occurs due to a steric clash between the NH_2 of G8 with the DMAP tail, which prevents the ligand from being fully accommodated into the minor groove. As a consequence, weak hydrogen bonds and van der Waals contacts are formed in this complex, which lead to a significant reduction in the binding enthalpy. This substantial loss in the binding enthalpy decreased the binding affinity despite the favourable entropy contribution to ΔG, which is likely to be caused by a gain in the conformational freedom of either the ligand and/or the DNA as a result of the weak non-bonded interactions formed in this complex, which are not enough to constrain or change the complex structure. A significant increase in the free binding energy was observed when the guanine responsible for the steric clash with DMAP tail of the ligand was replaced with inosine, which has no exocyclic amine. This increase in the free energy of bining was mainly enthalpic in origin (ΔH -9.555 $kcal.mol^{-1}$) with a small contribution from entropy (-TΔS -1.535 $kcal.mol^{-1}$). The small favourable entropy observed with the inosine containing sequence is likely to be caused by the release of more water and /or counter ions from the minor grooves because the ligand fits more snugly into the minor groove. Furthermore, binding to the minor groove of GC rich sequences is expected to release more water molecules and induce less conformational changes in the DNA structure because such sequences have a wider minor groove that can accommodate the ligand without inducing large perturbations in the DNA structure. This may explain why sequences such as CCTAGI and A/TCGCGT/A (see Section 3.2.3) have favourable entropy while others do not.

Overall, the interaction of thiazotropsin A with the binding sites, 5′-ACTAGT-3′, 5′-TCTAGA3′, 5′-GCTAGC3′, and 5′-ACATGT3′ was driven by large favourable enthalpy contributions and opposed by entropy. The unfavourable entropy suggests that the crescent-shaped ligand induces conformational changes in the DNA structure. The favourable enthalpy indicates that the hydrogen bonding and van der Waals forces are the main forces that drive thiazotropsin A-DNA association. There are two types of ligand-DNA base hydrogen bonds that may be critical to the sequence affinity and specificity exhibited by the dimeric ligand complex: hydrogen bonding between the thiazole nitrogen and the 2-amino hydrogen of guanine; and hydrogen bonding between the ligand amide hydrogens and either the O2 oxygen of thymine, the O2 oxygen of cytosine, or the N3 nitrogen of adenine. Furthermore, the presence of water molecules trapped between the ligand and the binding site could facilitate hydrogen bond formation by enabling the ligand to interact indirectly with DNA base pairs, sugars or phosphates through a connecting water molecule. Trapping of these water molecules could also explain the unfavourable entropies.

The differences observed in the thermodynamics of thiazotropsin A binding to the seven duplexes may be due to the ability of the different ligand dimer configurations to form these types of hydrogen bonds with their corresponding binding sites. The differences in van der Waals contacts and solvation could also be important contributors to the binding process.

3.2 ITC of thiazotropsin A analogues

We have used ITC to study the binding of the thiazotropsin A analogues P36, P22, thiazotropsin B, AIK18-51 and HA10 with DNA (Figure 3.7) to establish a link between the structure of the ligand and the thermodynamics of binding. These analogues vary in their structures in one or two positions with respect to thiazotropsin A. Comparing the thermodynamic binding characteristics of closely related ligand structures to a specific binding site can help establish how modifications in the structure influence binding affinity.

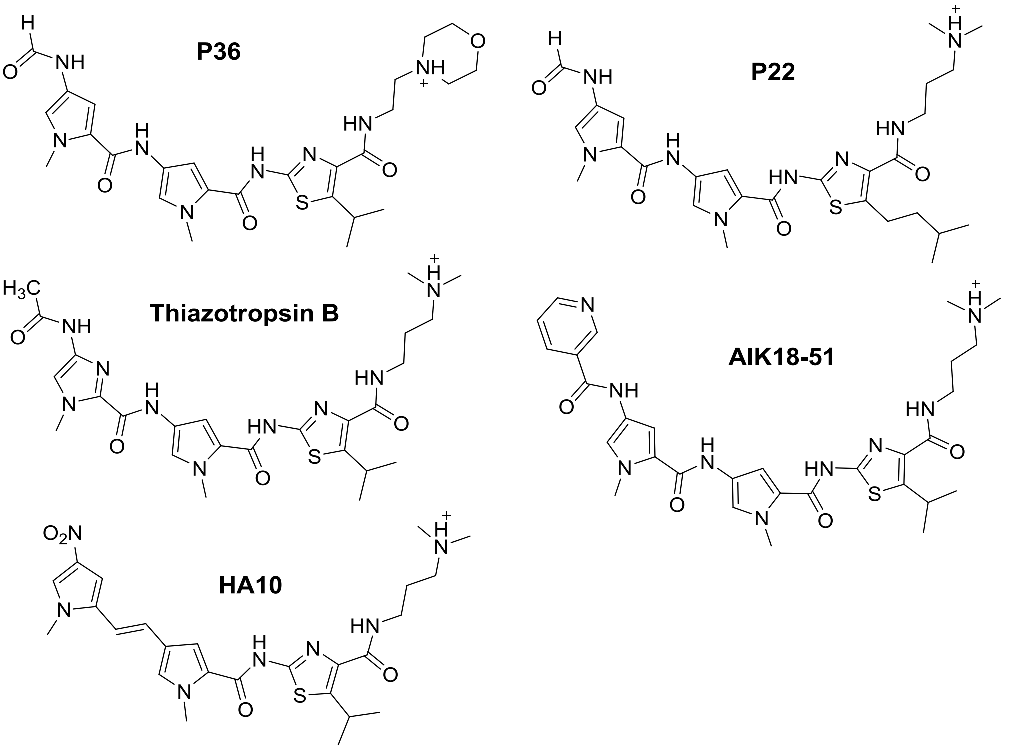


Figure 3.7 Structures of the thiazotropsin A analogues used in the study.

3.2.1 Thermodynamics of P36-DNA interactions

ITC was used to characterise the binding of P36 to four oligodeoxynucleotides (ODNs) containing the central sequences: 5′-ACTAGT-3′, 5′-TCTAGA-3′, 5′-GCTAGC-3′, and 5′-CCTAGG-3′, in PIPES buffer at a constant Na^+ concentration of 20 mM and a pH of 6.8. Figure 3.8 (A panels) shows the raw data of the titration of P36 to the four ODNs sequences at constant temperature 25 °C. Binding of P36 with the four ODN sequences is clearly associated with an exothermic process. The first 10 injections of the ligand (Figure 3.8 **(1)**) generated exothermic signals, followed by endothermic signals as result of the heats of dilution of the ligand in the buffer. Dilution peaks (data not shown) were all endothermic and their intensity decreased as more ligand was added indicating that at this concentration, aggregation of the P36 occurs in the buffered solution. Fitting the ITC data to a single site binding model allowed us to obtain the thermodynamic binding profiles summarised in Table 3.3.

The thermodynamic binding characteristics of P36 are similar to those observed with thiazotropsin A. The ligand formed complexes with the four dodecamers in a stoichiometry of 2:1, which were driven by large enthalpy changes and opposed by entropy. Again, this suggests that hydrogen bonding and /or van der Waals forces are the main molecular forces that drive the binding process. The apparent binding affinities of P36 follows the hierarchy: 5′-ACTAGT-3′ > 5′-TCTAGA-3′ > 5′-GCTAGC-3′ >5′-CCTAGG-3′ which is in agreement with the binding affinities of thiazotropsin A, although with slightly lower affinity.

A remarkable improvement in the binding enthalpy of P36 was observed with the four dodecamers. Figure 3.9 shows a comparison between the binding enthalpy of P36 and thiazotropsin A to their DNA targets. The greater exothermic binding enthalpies of P36 compared with those of thiazotropsin A was of the magnitude *-2.8, -5.3, -1.4* and *-3.7* $kcal.mol^{-1}$ for the binding sites 5′- ACTAGT-3′, 5′-TCTAGA-3′, 5′-GCTAGC-3′, and 5′-CCTAGG-3′, respectively.

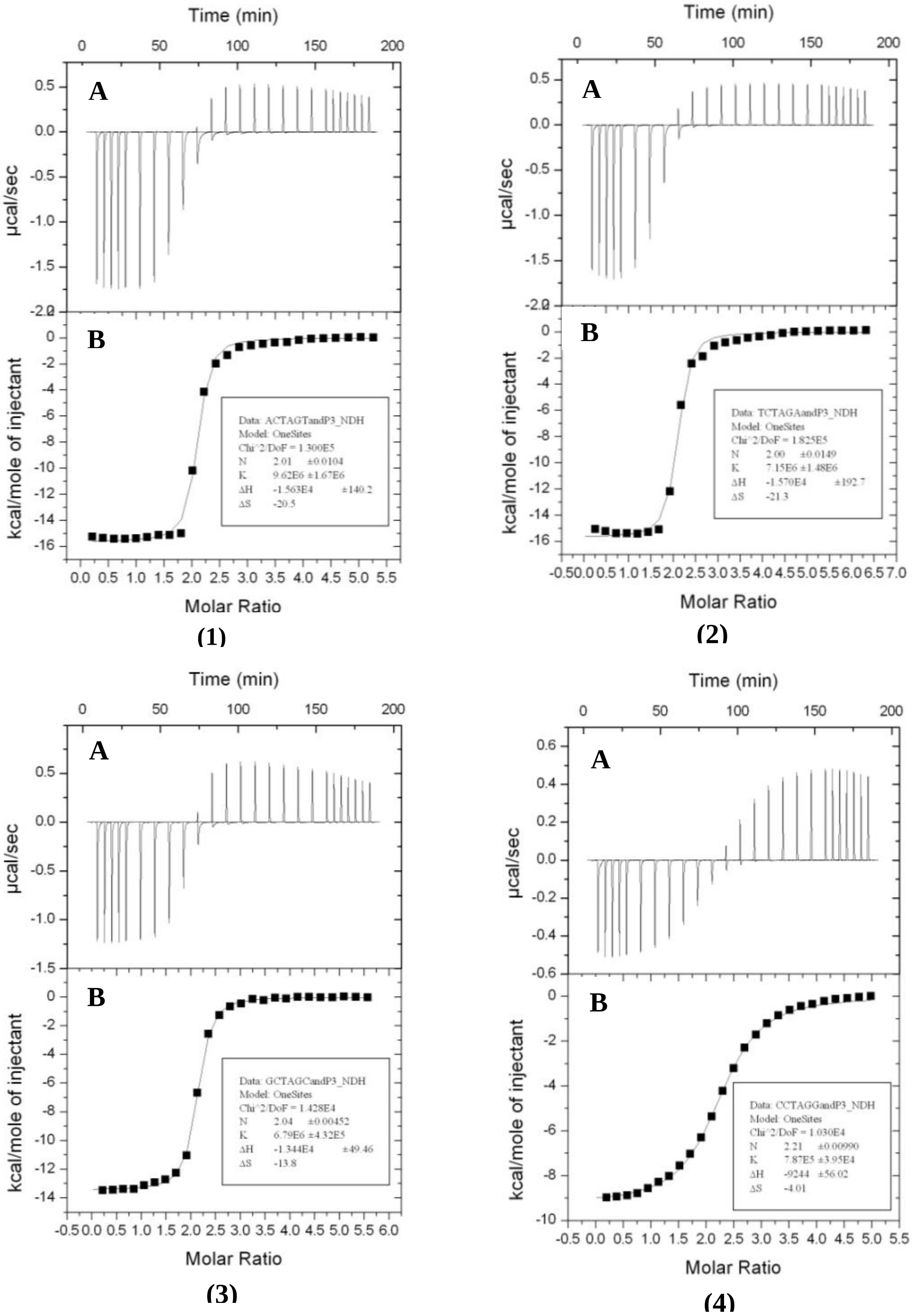

(1)
(2)
(3)
(4)

Figure 3.8 (A panels): ITC profiles at 25 °C for the titration of P36 into a solution of (1) GCG**ACTAGT**CGC (2) GCG**TCTAGA**CGC (3) GCG**GCTAGC**CGC and (4) GCG**CCTAGG**CGC at pH 6.8. Each heat burst curve is the result of a 10 µL injection of 0.5 mM ligand. The DNA concentration was 15µM, and the solution conditions were 10 mM PIPES, 20 mM NaCl, and 1 mM EDTA. **(B panels)** Corrected injection heats plotted as a function of the [ligand]/[DNA] ratio. The corrected injection heats were derived by integration of the ITC profiles shown in Panels A, followed by subtraction of the corresponding dilution heats derived from control titrations of drug into buffer alone. The data points reflect the experimental injection heats, while the solid reflect calculated fits of the data. The data were fit with a model for one binding site.

Table 3.3 Thermodynamic parameters for the interaction of P36 with four dodecamers in 10 mM PIPES buffer at 25 °C and pH of 6.8.

Sequence	**ΔH (Kcal.mol$^{-1)}$**	**TΔS (Kcal.mol^{-1})**	**K (M^{-1})**	**ΔG (Kcal.mol^{-1})**	**N (L/DNA)**
GCG**ACTAGT**CGC	-15.6	-6.1	9.6x10^{6}	-9.5	2.0
GCG**TCTAGA**CGC	-15.7	-6.3	7.5x10^{6}	-9.4	2.0
GCG**GCTAGC**CGC	-13.4	-4.1	6.8 x10^{6}	-9.3	2.0
GCG**CCTAGG**CGC	-9.2	-1.2	7.9x10^{5}	-8.0	2.0

P36 and thiazotropsin A differ only with respect to the tail substituent; P36 has an ethylmorpholine group in place of the DMAP group (see Figure 3.7). Thus, the presence of an oxygen atom in the morpholine ring, the ring itself, and a shorter chain appears to suggest an enthalpically driven enhancement in binding free energy. The molecular basis for this may reflect contributions from enhanced hydrogen bonding and van der Waals interactions. For example, the oxygen in the morpholine ring is capable of hydrogen bonding with the bases, the sugars or water in the complex. The reduced enthalpies with 5′-GCTAGC-3′ and 5′CCTAGG-3′ may be due to the more bulky morpholine ring (compared with DMAP) in the shorter tail clashing with the exocylic guanine-NH_2 on the groove floor, which is more exacerbated with the contiguous GG sequence, as seen with thiazotropsin A. Overall, despite the large favourable enthapic contribution, significant entropic penalties tended to oppose these gains, perhaps via water structuring via the morpholine oyygen (Figure 3.10). The lowest entropic penalty was observed with CCTAGG; if the exocylic guanine-NH_2 on the groove floor forces the tail out into solvent, there will be more conformational degrees of freedom associated with this binding mode, and therefore reduced entropic loss on complexation. These results illustrate the hurdle which enthalpy/entropy compensation poses to the optimisation of binding affinity. Mathematically, this is due to the Gibbs free energy equation (ΔG = ΔH –

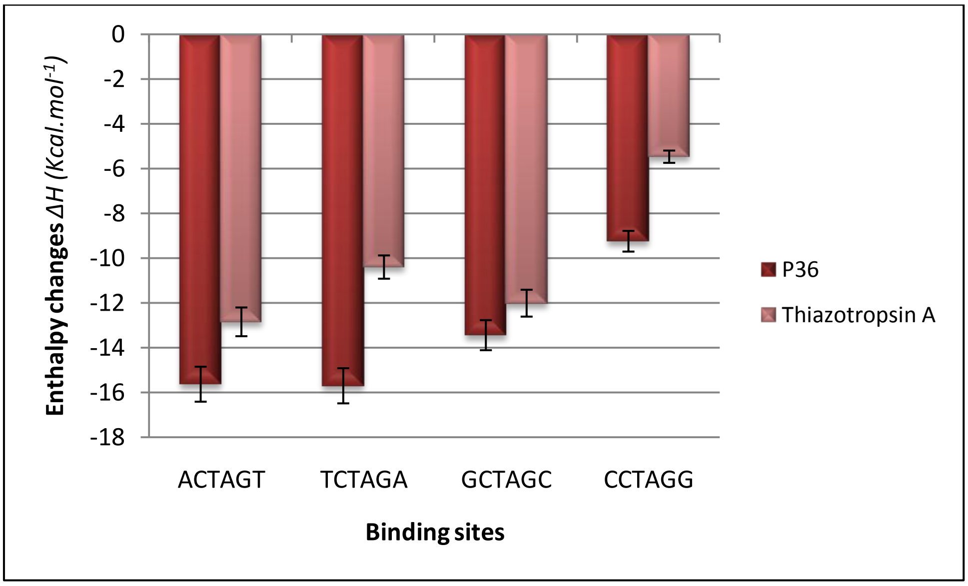


Figure 3.9 Comparison of the binding enthalpy of P36 with thiazotropsin A.

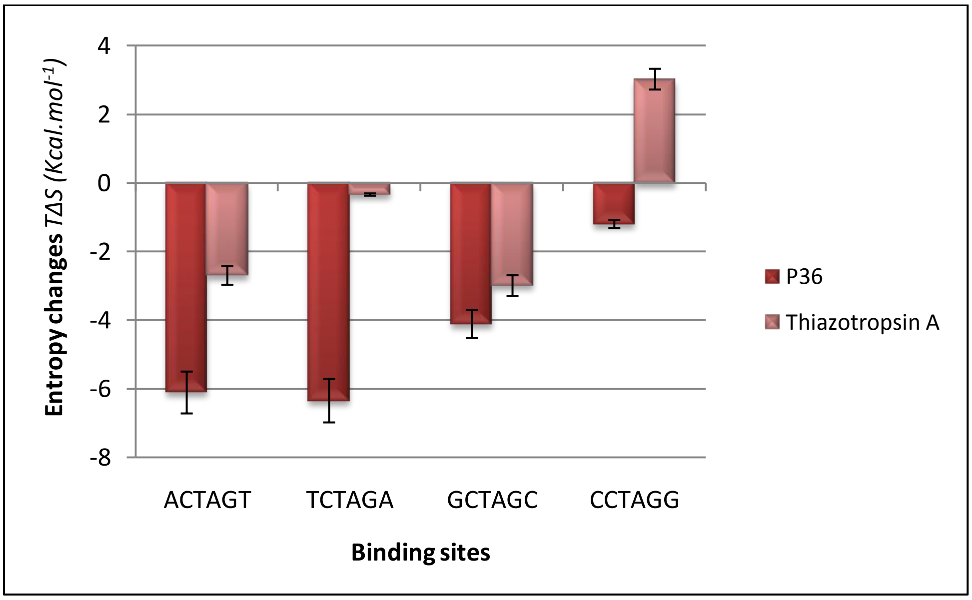


Figure 3.10 Comparison of the binding entropy of P36 with thiazotropsin A.

TΔS): as the enthalpy changes (ΔH) and the entropy changes (ΔS) have opposite signs, when both terms are increased, there will be very little change in the free energy of binding (ΔG). For example, if the binding enthalpy of the P36-DNA complexe is enhanced through the formation of an extra hydrogen bond, this act of forming or even strengthening bonds will limit movement (e.g. rotational and vibrational) and incur an entropic penalty. Moreover, some water molecules may be trapped at the binding interface to facilitate hydrogen bond formation which could contribute to unfavourable entropy via reduced desolvation.

3.2.2 Thermodynamics of P22-DNA interactions

ITC was used to characterize the binding of P22 to four dodecamers containing the central sequences, 5′-ACTAGT-3′, 5′-TCTAGA-3′, 5′-GCTAG-3′, and 5′-CCTAGG-3′. Figure 3.11 shows the ITC profile for the titration of P22 into a solution of 5′-ACTAGT-3′. Panel B shows the resulting corrected injection heats plotted as a function of the [ligand]/ [duplex] ratio. The ITC profile has two apparent phases, indicative of two distinct binding events. An initial exothermic tight 2:1 specific minor groove binding event, which once these sites are saturated, is followed by a second endothermic binding event at a stoichiometry of 5-6:1 ligand to DNA).

There are two possible explanations to account for these results:

1) There is only one DNA binding event and the sigmoid curve we see upon saturation of DNA binding sites is in fact due to ligand self-association in the ITC cell as more and more free ligand accumulates. If the ligand self-associates in the syringe, upon injection into the cell there is dilution of the ligand and hence dissociation and then subsequently self-association again. These all have heats and equilibrium constants that need to be accounted for. If it is ligand self-association, then the heats of dilution (ligand into buffer alone) should be the same shape as binding curve from about the sixth injection onwards, i.e. small positive heats saturating to small negative heats (Figure 3.11 A). As the heat signals at the end of titration are not the same sign and magnitude as the ligand heats of dilution into the same buffer (Figure 3.12), we can conclude that the second binding event was not due to ligand self-association.

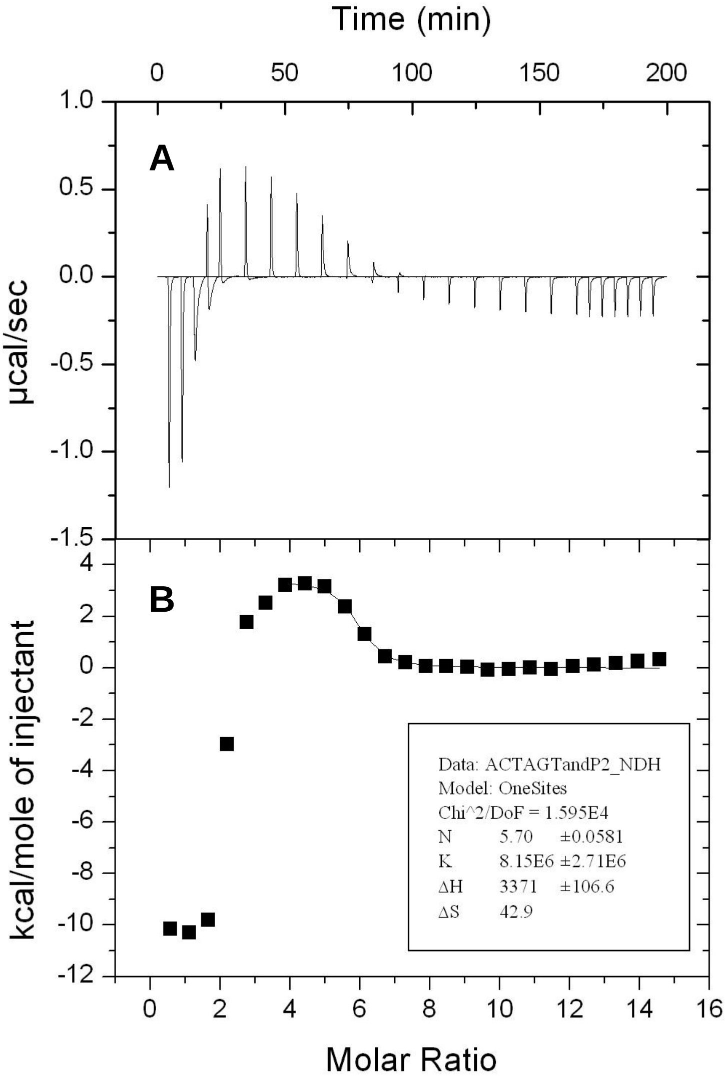


Figure 3.11 (A panels): ITC profiles at 25 °C for the titration of P22 into a solution of GCG**ACTAGT**CGC at pH 6.8. Each heat burst curve is the result of a 10 µL injection of 0.5 mM ligand .The DNA concentration was 7µM, and the solution conditions were 10 mM PIPES, 20 mM NaCl, and 1 mM EDTA. **(B panels)** Corrected injection heats plotted as a function of the [ligand]/[DNA] ratio. The corrected injection heats were derived by integration of the ITC profiles shown in Panels A, followed by subtraction of the corresponding dilution heats derived from control titrations of drug into buffer alone. The data points reflect the experimental injection heats, while the solid line reflects calculated fits of the data. The first sex data points were ignored, and the remaining data points were fit to a model for single binding sites to obtain the thermodynamic profile for the second binding event.

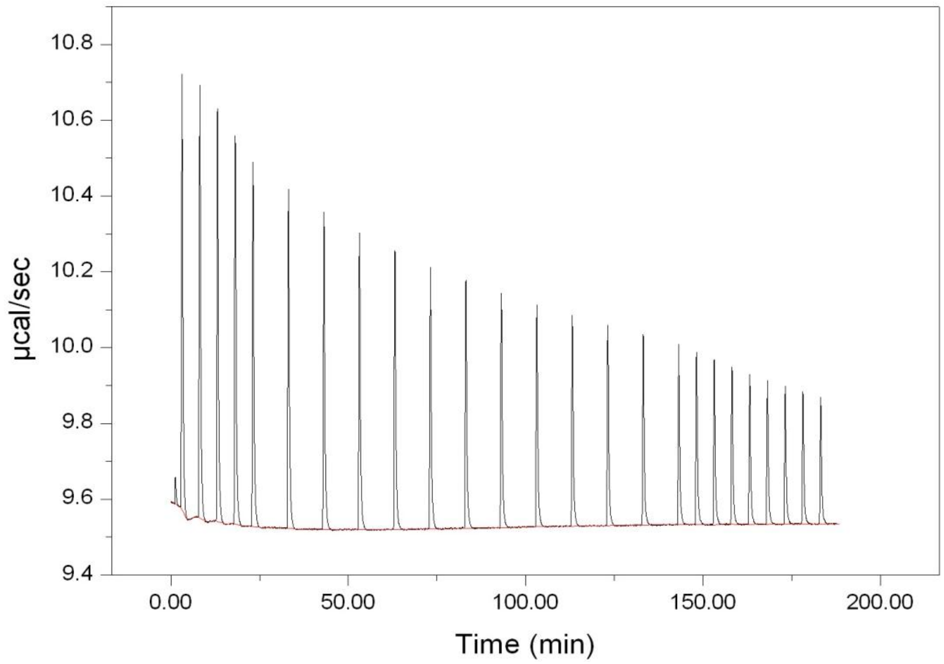
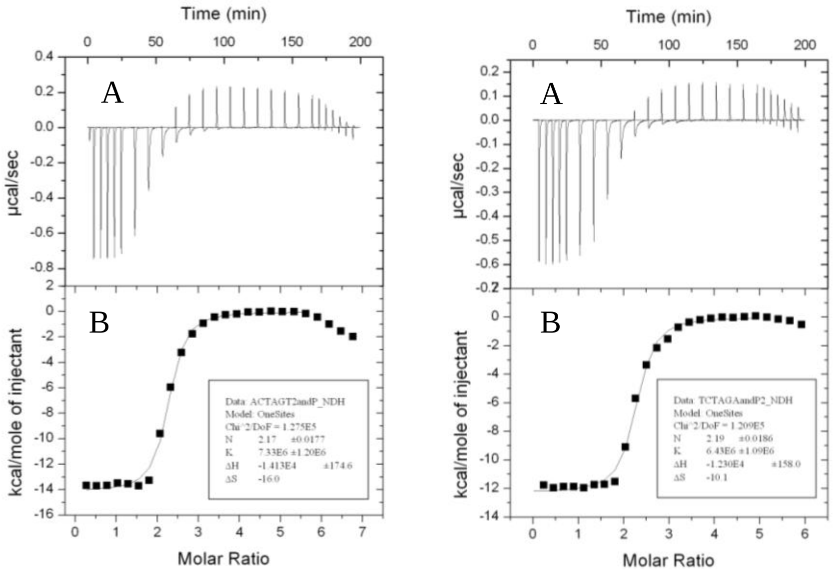


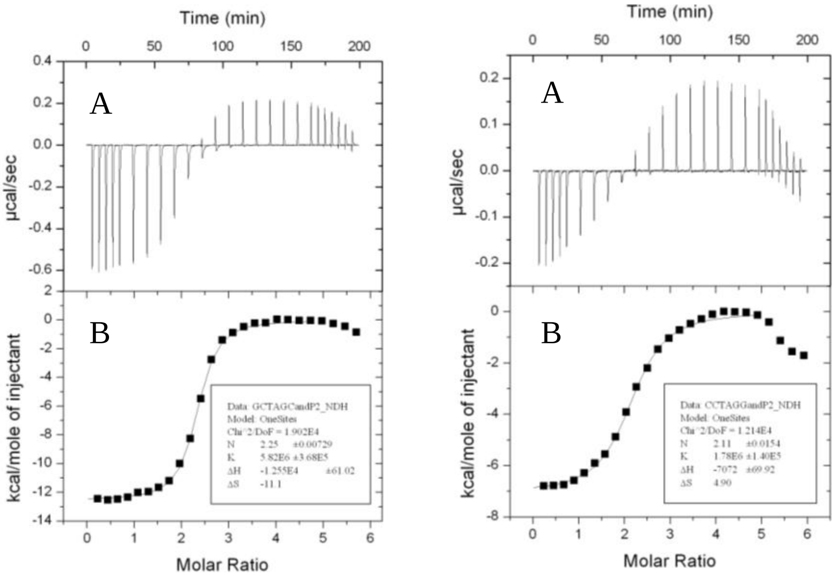
Figure 3.12 ITC dilution heats of P22 in PIPES buffer. Each heat burst curve is the result of a 10 µL injection of 0.5 mM ligand into the PIPES buffer solution. The solution conditions were 20 mM PIPES, 10 mM NaCl, and 0.1 mM EDTA.

2) There are two sets of binding sites on the DNA duplex for the ligand. After the initial exothermic tight 2:1 specific binding event, these sites become saturated and a second, lower affinity, endothermic binding event occurs at a different binding site. This type of second binding event has been observed with other MGBs interacting with short oligonucleotides at high stiochiometry (ligand to DNA) and has been attributed to non-specific interactions,[128] possibly with the DNA backbone. To investigate this possibility, the experiment was repeated at a different salt concentration (see section 3.2.2), since the second binding event might be more electrostatic in nature and as such more sensitive to salt.

The thermodynamic profiles were obtained by fitting experimental data points of the two binding event separately to a single site binding model. For the second binding event, the first six data points were ignored, and the remaining data points were fit to a model for single binding sites to obtain the thermodynamic profile for the second event (see Section 3.2.2.1). More data points were required for fitting the first binding event accurately which was achieved by decreasing the concentration of the injected ligand (Figure 3.13). Each heat burst curve is the result of a 10 µL injection of 0.25 mM ligand. The DNA concentration was 7µM, and the solution conditions were 10 mM PIPES, 20 mM NaCl, and 1 mM EDTA. In these ITC profiles, the last four data points (belong to the second binding event) were ignored and the remaining data points were fit to a model for single binding sites to obtain the thermodynamic profiles for the first binding event summarised in Table 3.4 . Except for the second binding event, the thermodynamic binding characteristics of P22 in the first binding mode were similar to those observed with thiazotropsin A and P36. The binding interaction of P22 with the four dodecamers were driven by large enthalpy changes and opposed by entropy. These results suggest that the hydrogen bonding and /or van der Waals forces are the main molecular forces that drive the binding process and that the apparent binding affinities follows the hierarchy: 5′-ACTAGT-3′ > 5′-TCTAGA-3′ > 5′-GCTAGC-3′ >5′-CCTAGG-3′, which is in agreement with the hierarchy noted for the binding affinities of thiazotropsin A and P36. However, the binding affinity of P22 to the binding sites is lower than the binding affinity of thiazotropsin A for the corresponding binding sites. Both P22 and thiazotropsin A recognise the 5′-CCTAGG3′ sequence with a comparable binding affinity of $1.7\text{x}10^{6}$ M^{-1}. Like P36, an increase in enthalpy was observed in the binding of P22 with the four dodecamers which could be related to the enhancement of van der Waals contacts between the ligand and the minor grooves of DNA. The structure of P22 and thiazotropsin A differ at the thiazole substituent, with this functionality containing an isopentyl group in P22 and an isopropyl group in thiazotropsin A (see Figure 3.7). The presence of a longer alkyl group in P22 seems to improve van der Waals interactions with the sugar moieties lining the groove walls, which leads to an enhancement in the binding enthalpy. Figure 3.14 shows a comparison between the binding enthalpy of P22 and thiazotropsin A with the four dodecamers. The increase in the binding enthalpies of P22 compared with those of thiazotropsin A was in a magnitude of -1.3, -1.9, -0.5, and -1.6 $kcal.mol^{-1}$ for the binding sites; 5′-ACTAGT-3′, 5′-TCTAGA-3′, 5′-GCTAGC-3′, and 5′-CCTAGG-3′, respectively.



(1) ACTAGT **(2) TCTAGA**



(3) GCTAGC **(4) CCTAGG**

Figure 3.13 (A panels): ITC profiles at 25 °C for the titration of P22 into a solution of (1) GCGACTAGTCGC (2)GCGTCTAGACGC (3) GCGGCTAGCCGC and (4) GCGCCTAGGCGC at pH 6.8. Each heat burst curve is the result of a 10 µL injection of 0.25 mM ligand. The DNA concentration was 7µM, and the solution conditions were 10 mM PIPES, 20 mM NaCl, and 1 mM EDTA. **(B panels)** Corrected injection heats plotted as a function of the [ligand]/[DNA] ratio. The corrected injection heats were derived by integration of the ITC profiles shown in Panels A, followed by subtraction of the corresponding dilution heats derived from control titrations of drug into buffer alone. The data points reflect the experimental injection heats, while the solid line reflects calculated fits of the data. The last four data points were ignored , and the remaining data points were fit to a model for single binding sites to obtain the thermodynamic profile for the first binding event.

Table 3.4 Thermodynamic parameters for the first binding event of P22 with four dodecamers in 10 mM PIPES buffer at 25 °C and pH of 6.8.

Sequence	**ΔH (Kcal.mol^{-1})**	**TΔS (Kcal.mol^{-1})**	**K (M^{-1})**	**ΔG (Kcal.mol^{-1})**	**N (L/DNA)**
GCG**ACTAGT**CGC	-14.1	-4.7	7.33x10^{6}	-9.4	2.0
GCG**TCTAGA**CGC	-12.3	-3.0	6.43x10^{6}	-9.3	2.0
GCG**GCTAGCC**GC	-12.5	-3.3	5.82x10^{6}	-9.2	2.0
GCG**CCTAGG**CGC	-7.1	1.4	1.78x10^{6}	-8.5	2.0

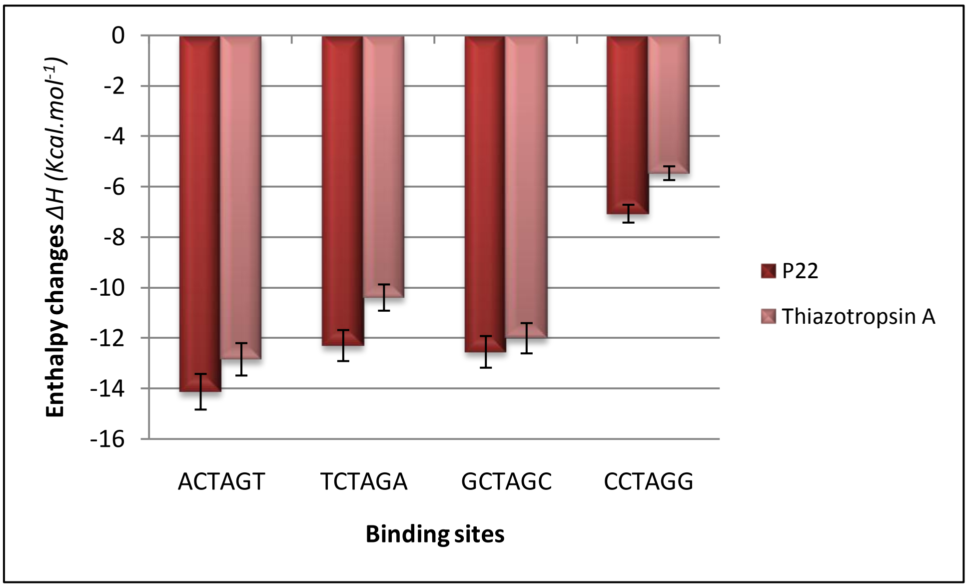


Figure 3.14 A comparison of the binding enthalpy of P22 with thiazotropsin A.

These differences in enthalpies, particularly between ligand binding with 5′-GCTAGC-3′, and 5′-CCTAGG-3′, where there is a more marked difference between P36 and thiazotropsin A, than between P22 and thiazotropsin A, confirms the steric role that the tail unit plays when binding to regions containing G/C base pairs. However, the overall binding affinity was not improved despite the increase in the enthalpic contribution to binding, which was accompanied by a significant penalty in entropy (Figure 3.15). This again illustrates the obstacle which enthalpy/entropy compensation poses to the optimisation of binding affinity. The gain in enthalpy through van der Waals forces was compensated by unfavorable entropy changes, probably through the conformational restriction of the more flexible isopentyl group upon complexation.

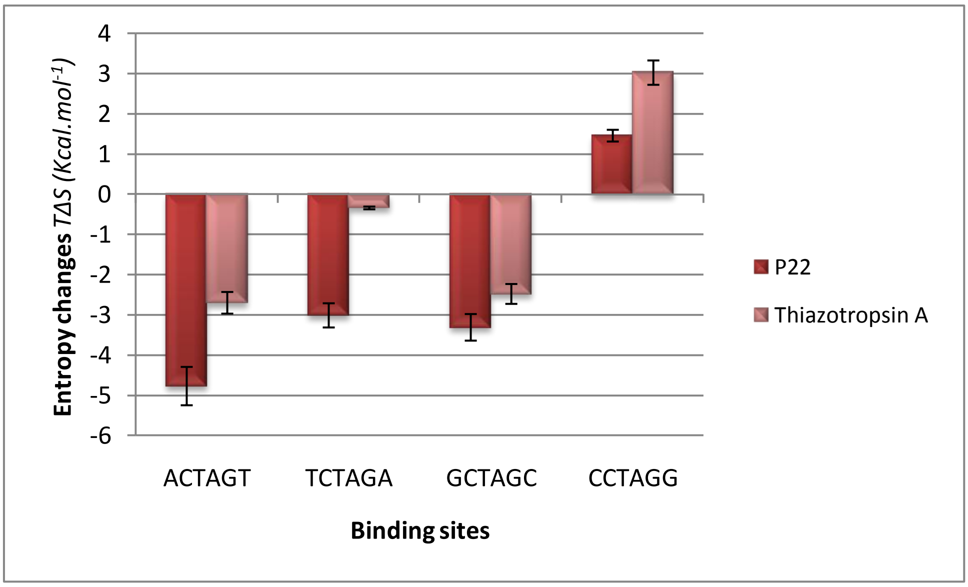
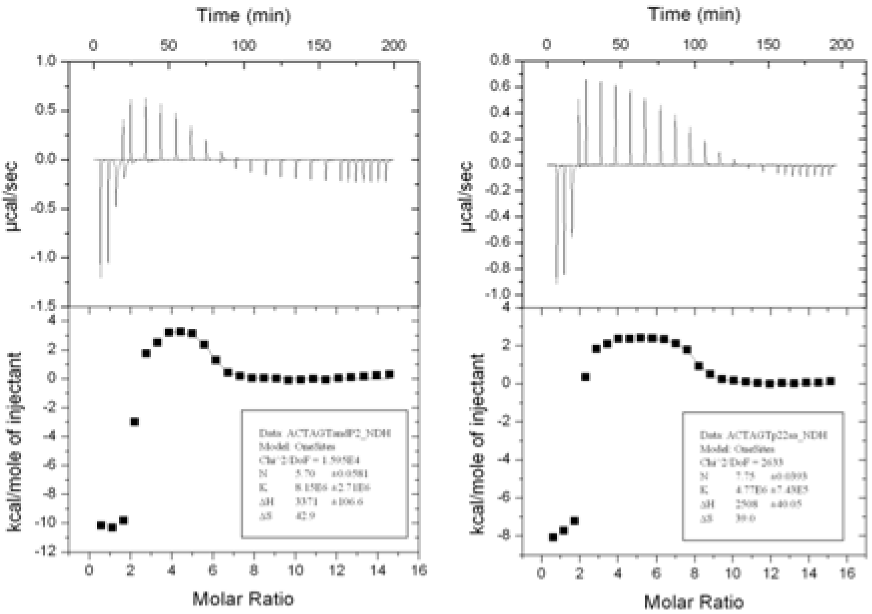


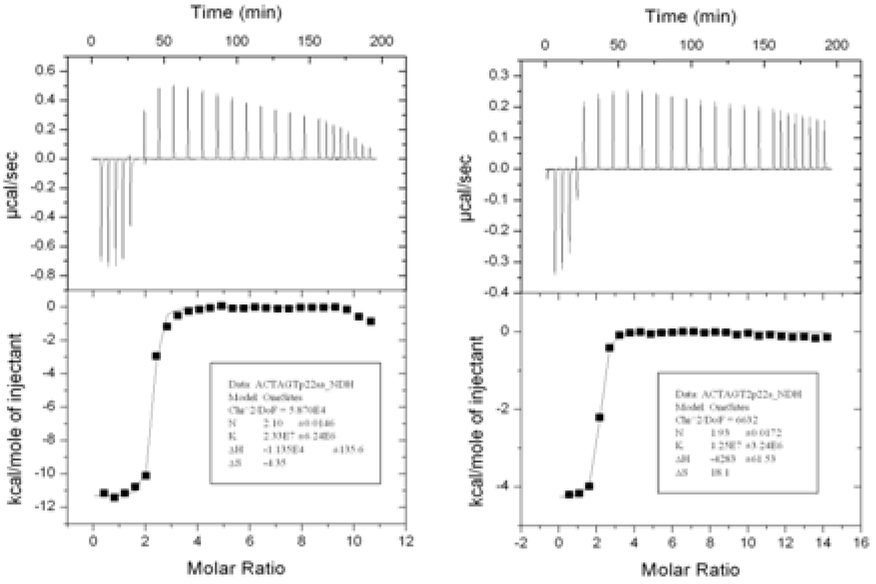
Figure 3.15 A comparison of the binding entropy of P22 with thiazotropsin A.

3.2.2.1 Effect of salt concentration on P22-DNA interaction

ITC studies of P22-DNA interactions revealed the presence of two distinct binding events; an initial exothermic tight 2:1 specific minor groove binding event followed by a second non-specific endothermic binding event at high ligand concentration. Non-specific binding modes, which include electrostatic backbone binding, are sensitive to salt concentration, and so to establish whether the second binding event was electrostatic in nature and more sensitive to salt, an ITC study as a function of salt concentration was therefore carried out to characterise the binding of P22 to the dodecamer containing the 5′-ACTAGT-3′ sequence. Figure 3.16 shows the enthalpograms for the titration of P22 to the ACTAGT oligonucleotide in PIPES buffer at different salt concentrations and at a constant pH of 6.8. ITC revealed that the second binding event was acutely affected by salt concentration. Increasing Na^+ concentration shifted the second binding event to the right. At 20 mM [NaCl], the



(1) 20 mM NaCl **(2) 30 mM NaCl**



(3) 50 mM NaCl **(4) 100 mM NaCl**

Figure 3.16 (A panels): ITC profiles at 25 °C for the titration of P22 into a solution of 5′-GCGACTAGTCGC-3′- sequence at different salt concentrations **(1)** 20 mM NaCl **(2)** 30 mM NaCl **(3)** 50 NaCl **(4)** 100 mM NaCl. Each heat burst curve is the result of a 10 µL injection of 0.5 mM ligand. The DNA concentration was 7µM, and the solution conditions were 10 mM PIPES and 1 mM EDTA at pH 6.8. **(B panels)** Corrected injection heats plotted as a function of the [ligand]/[DNA] ratio. The corrected injection heats were derived by integration of the ITC profiles shown in Panels A, followed by subtraction of the corresponding dilution heats derived from control titrations of drug into buffer alone. The data points reflect the experimental injection heats, while the solid reflect calculated fits of the data. The data were fit with a model for one binding mode.

second

second endothermic binding event started to appear at stoichiometry of 5:1 (ligand to DNA), but when Na^+ concentration was increased to 30 mM, it appeared at stoichiometry of 7:1. A further increase in Na^+ concentration to 50 mM and 100 mM shifted the second binding event beyond the titration curve. These results clearly indicate that the positively charged ligand competes with the Na^+ counter ions in neutralizing the negatively charged backbone of DNA. In solution, positive ions are condensed around the polyanionic DNA helix to reduce the net charge and help stabilise the duplex. Interaction between a positively charged molecule and DNA has the effect of dismissing one or more of the cations around the DNA, since the positively charged ligand provides competing backbone neutralisation. Our results show that as the concentration of counterions increases, a higher concentration of ligand is required in order for the second binding event to take place. Therefore, we attribute the second binding event to non-specific electrostatic backbone binding.

The thermodynamic profiles for the first and second binding process at different salt concentrations are listed in Table 3.5. These profiles were obtained by fitting experimental data points of the two binding event separately to a single site binding model. ITC data show that the first binding event is enthalpy driven and entropy opposed at Na^+ concentration range of 20-50 mM and driven by both enthalpy and entropy at a high concentration of 100 mM. The large negative enthalpy suggests that there are a large number of favourable hydrogen bond contacts and/or van der Waals interactions between the ligand and DNA. The unfavourable entropy suggests a conformational change in either or both of the reactants which includes loss of rotational, vibrational and translational degrees of freedom during complex formation. Although a dramatic decrease in the enthalpy of the first binding event was observed when the salt concentration was increased, the ligand binding affinity did not decrease due to either a decrease in the entropic penalty (at 30 and 50 mM) or an increase in the entropy contribution to the free energy of binding (ΔG) (at 100 mM). This contribution from entropy mainly stems from expelling counterions and probably water from the minor grooves upon ligand binding. These results also show that the tight 2:1 specific minor groove binding was not affected by the salt concentration at which the ITC experiments were conducted, contrary to the second binding event which was significantly affected by salt concentration. The ITC profile of the second binding event not only showed that the Na^+ salts were competing with the positively charged ligand to neutralise the negatively charged backbone of the DNA helix, but also increasing the salt concentration reduced the binding affinity of the second binding mode (Table 3.5). This salt dependency of K_b for the second binding event indicates that electrostatic forces play an important role in this interaction. However, the appearance of a second binding event only with the isopentyl substituted ligand (P22) and its absence with thiazotropsin A suggests that the hydrophobic effect could play a key role in this interaction. Burying the solvent-exposed surface area of the large lipophilic isopentyl units by complexation with DNA, coupled with their removal from bulk solvent could also account for the observed favourable entropy, with the release of water as well as counterions upon binding of the cationic ligand with DNA.

Table 3.5 Thermodynamic parameters for the interaction of P22 with a dodecamer containing the central ACTAGT binding site in 10 mM PIPES buffer at different Na^+ concentrations and a constant temperature of 25 °C and pH of 6.8.

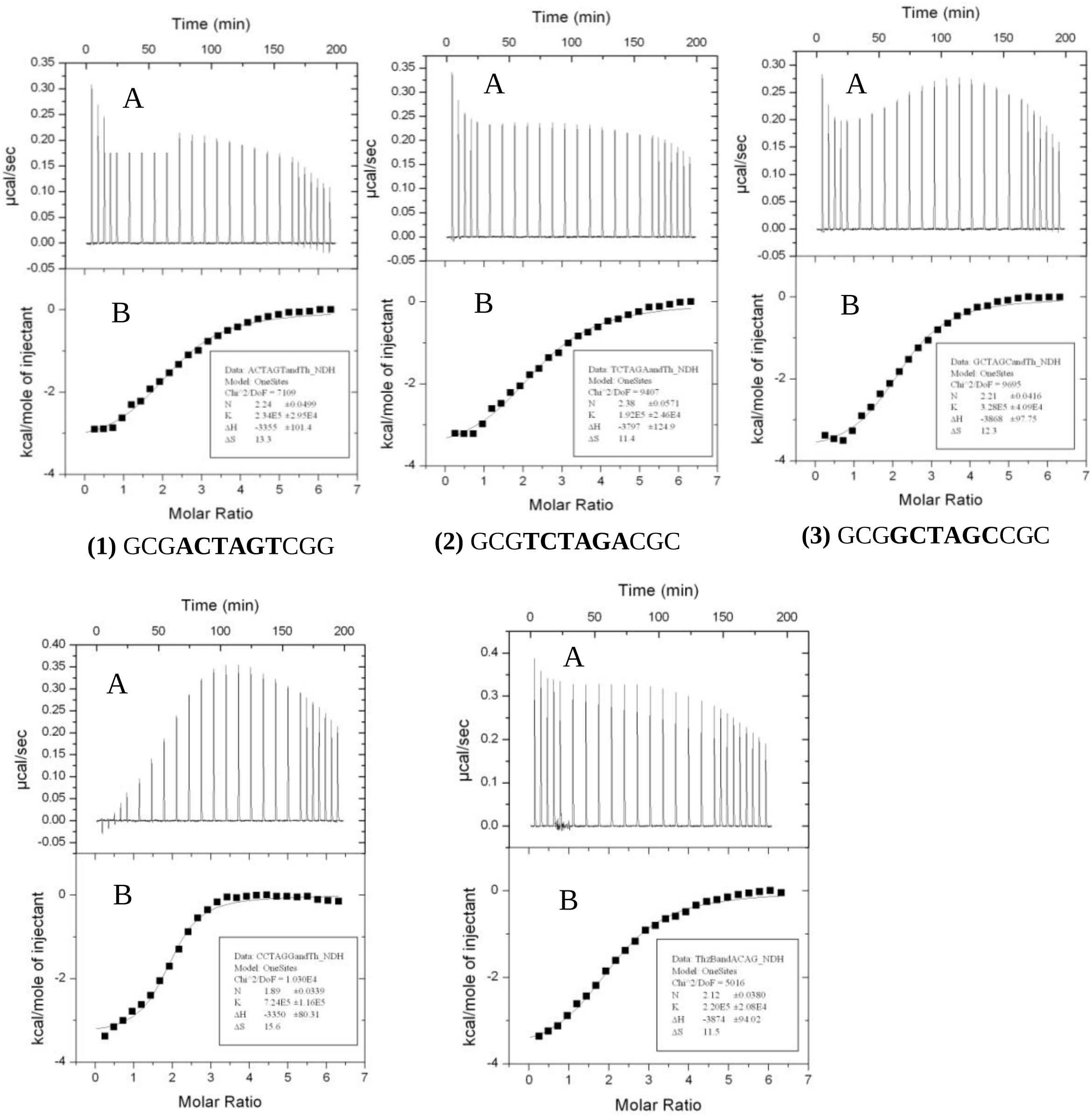
[NaCl] *(mM)*		**ΔH** *($Kcal.mol^{-1}$)*	**$T\Delta S$** *($Kcal.mol^{-1}$)*	**K** *(M^{-1})*	**ΔG** *($Kcal.mol^{-1}$)*	**N** *(L/DNA)*
1^{st} binding site	**20**	-14.1	-4.8	$7.3x10^{6}$	-9.3	2.0
	30	-12.6	-2.9	$1.4x10^{7}$	-9.7	2.0
	50	-11.4	-1.3	$2.3\ x10^{7}$	-10.1	2.0
	100	-4.3	5.4	$1.3x10^{7}$	-9.7	2.0
2^{nd} binding site	**20**	3.4	12.8	$8.2x10^{6}$	-9.4	5
	30	2.5	11.6	$4.8x10^{6}$	-9.1	7
	50	n/a	n/a	n/a	n/a	10
	100	n/a	n/a	n/a	n/a	n/a

n/a: not applicable because the 2nd binding event located beyond the titration curve

3.2.3 Thermodynamics of thiazotropsin B-DNA interactions

ITC has been used to characterise the binding of thiazotropsin B with seven DNA duplexes containing the central sequences 5′-ACTAGT-3′, 5′-TCTAGA-3′, 5′-GCTAGC-3′, 5′-CCTAGG-3′, 5′-ACGCGT-3′, 5′- TCGCGT-3′, and 5′-ACAGT-3′. (Figure 3.17 & 3.20). The profiles are summarised in Table 3.6. Inspection of the ITC data reveals enormous reductions in the binding affinities of thiazotropsin B to the binding sites; 5′-ACTAGT-3′, 5′-TCTAGA-3′, 5′-GCTAGC-3′, 5′-CCTAGG-3′ and 5′-ACAGT-3′ compared with those observed for thiazotropsin A. These low affinities stem from a huge reduction in the binding enthalpy (Figure 3.18), which results undoubtedly from the changes introduced in the structure of thiazotropsin B; namely, replacing one of the pyrrole rings and the formyl head of thiazotropsin A, with an imidazole ring and acetyl head group, respectively. The imidazole nitrogen lies parallel to the O2 oxygen of T^6 and the O2 oxygen of T^{18} (Figure 3.19), placing two H-bond acceptors in proximity and creating repulsive effects which reduce affinity.

Replacing thymines at position 6 and 18 with guanines leads to a remarkable increase in the binding affinity (Table 3.6 & Figure 3.20). The 2-amino hydrogen bond donor of guanine is capable of forming specific hydrogen bonds with both the imidazole nitrogen and the thiazole nitrogen, which markedly enhances binding affinity. The preferential binding of thiazotropsin B to the two duplexes (5′-ACGCGT-3′ and 5′-TCGCGA-3′) is primarily enthalpically driven (although also entropically favourable), while the reduced binding to the other four duplexes is due to less favourable binding enthalpies. The data listed in Table 3.6 allow us to evaluate the thermodynamic consequences of binding by changing the central two base pairs in the low-affinity, 5′-AC**TA**GT-3′ and 5′-TC**TA**GA-3′ binding sites. Inspection of these data reveals that changing the central two base pairs TA to GC in 5′-AC**TA**GT-3′ (to produce 5′-AC**GC**GT-3′) and 5′-TC**TA**GA-3′ (to produce 5′-TC**GC**GA-3′) (Figure 3.21) result in a gain of 4.06 and 2.89 Kcal mol^{-1} of binding enthalpy, and a



(1) GCG**ACTAGT**CGG

(2) GCG**TCTAGA**CGC

(3) GCG**GCTAGC**CGC

(4) GCG**CCTAGG**CGC

(5) GCG**ACAGT**CGC

Figure 3.17 (A panels) ITC profiles at 25 °C for the titration of thiazotropsin B into a solution of **(1)** GCG**ACTAGT**CGC; **(2)** GCG**TCTAGA**CGC; **(3)** GCG**GCTAGC**CGC; **(4)** GCG**CCTAGG**CGC; and **(5)** GCG**ACAGT**CGC at pH 6.8. Each heat burst curve is the result of a 10 µL injection of 0.5 mM ligand into 15µM of DNA. **(B panels)** Corrected injection heats plotted as a function of the [ligand]/[DNA] ratio. The corrected injection heats were derived by integration of the ITC profiles shown in Panels A, followed by subtraction of the corresponding dilution heats derived from control titrations of drug into buffer alone. The data points reflect the experimental injection heats, while the solid line reflects calculated fits of the data which were fit to a model for single binding sites to obtain the thermodynamic profiles.

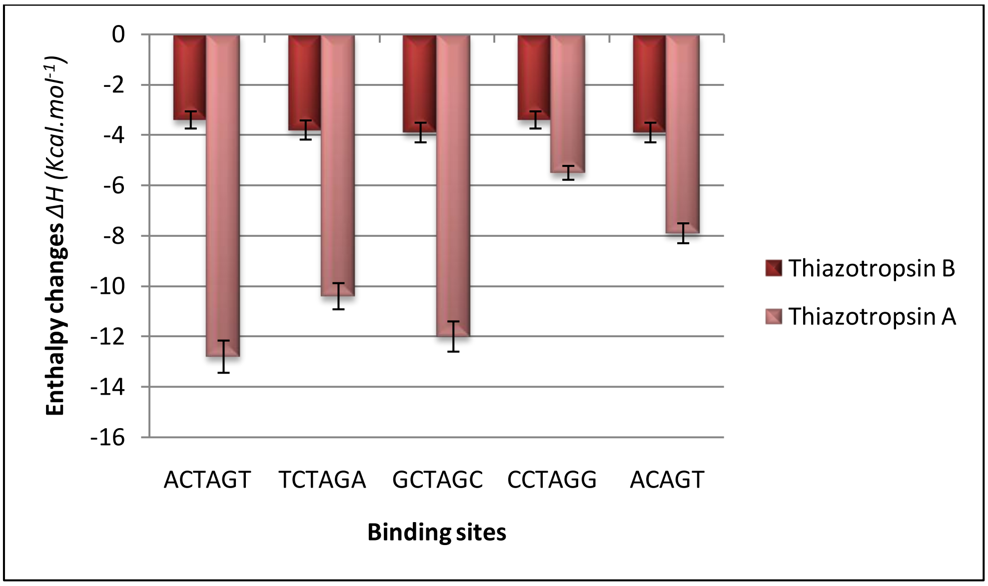


Figure 3.18 A comparison of the binding enthalpy of thiazotropsin B with thiazotropsin A.

Table 3.6 Thermodynamic parameters for the interaction of P22 with four dodecamers in 10 mM PIPES buffer at constant temperature of 25 °C and constant pH of 6.8.

Sequence	*ΔH* *(Kcal.mol^{-1})*	*TΔS* *(Kcal.mol^{-1})*	*K* *(M^{-1})*	*ΔG* *(Kcal.mol^{-1})*	*N* *(L/DNA)*
GCG**ACTAGT**CGC	-3.4	4.0	2.3x10^5	-7.4	2.0
GCG**TCTAGA**CGC	-3.8	3.4	1.9x10^5	-7.2	2.0
GCG**GCTAGC**CGC	-3.9	3.6	3.3 x10^5	-7.5	2.0
GCG**CCTAGG**CGC	-3.4	4.6	7.2x10^5	-8.0	2.0
GCG**ACAGT**CGC	-3.9	2.4	2.2x10^5	-6.3	2.0
GCG**ACGCGT**CGC	-7.4	1.8	5.3x10^6	-9.2	2.0
GCG**TCGCGA**CGC	-6.7	2.1	2.6x10^6	-8.8	2.0

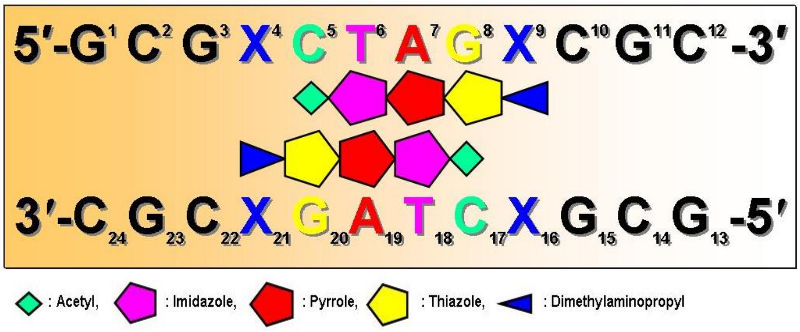
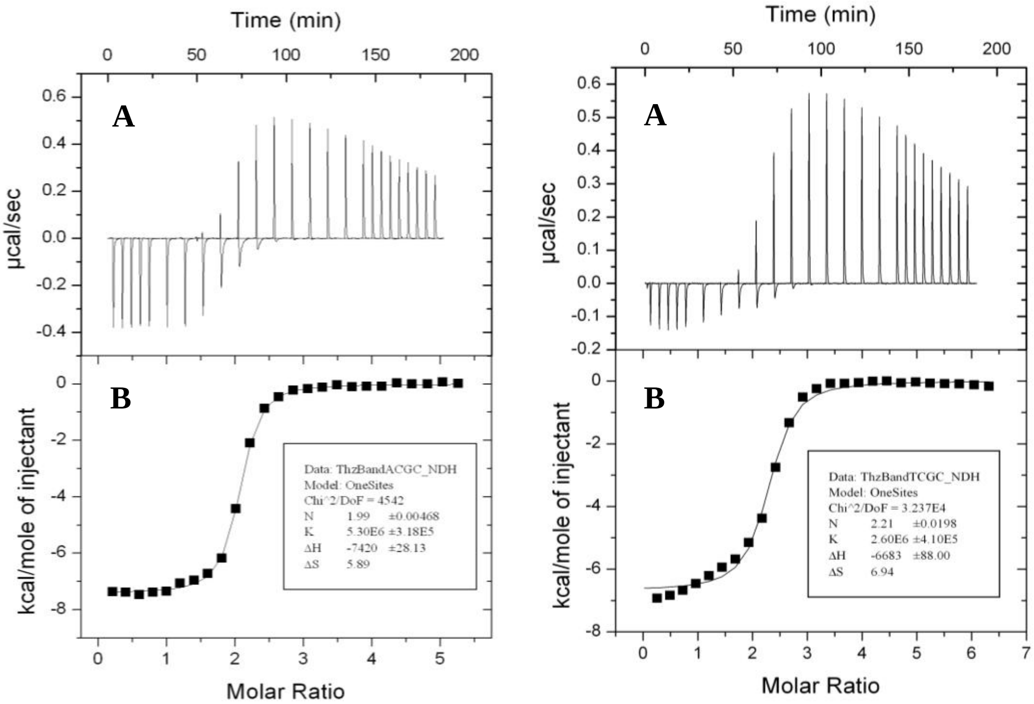


Figure 3.19 A Schematic representation for the binding of thiazotropsin B with dodecamers containing the central sequence XCTAGX, where X is T, A, C, or G.



(1) GCG**ACGCGT**CGC

(2) GCG**TCGCGA**CGC

Figure 3.20 (A panels): ITC profiles for the titration of thiazotropsin B into a solution of **(1)** GCG**ACGCGT**CGC **(2)** GCG**TCGCGA**CGC at 25 °C and at pH 6.8. Each heat burst curve is the result of a 10 μL injection of 0.5 mM ligand into 15μM of DNA. **(B panels)** Corrected injection heats plotted as a function of the [ligand]/[DNA] ratio. The corrected injection heats were derived by integration of the ITC profiles shown in Panels A, followed by subtraction of the corresponding dilution heats derived from control titrations of ligand into buffer alone. The data points reflect the experimental injection heats, while the solid line reflects calculated fits of the data which were fit to a model for single binding sites to obtain the thermodynamic profiles.

loss of 2.208 and 1.329 Kcal mol^{-1} of binding entropy, respectively. Despite these entropic penalties compared with the other sequences, entropy still favours complex formation, and the gains in binding enthalpy translate into increments in binding free energy of 1.857 and 1.557 Kcal mol^{-1} for the duplexes 5′-AC**CG**GT-3′ and 5′-TC**CG**GA-3′ relative to the duplexes 5′-AC**TA**GT-3′ and 5′-TC**TA**GA-3′ respectively. Thus, the enhanced affinity exhibited by thiazotropsin B for the 5′-AC**CG**GT-3′ and 5′-TC**CG**GA-3′ binding sites is enthalpic in origin. These results show clearly the importance of hydrogen bonding for sequence affinity and specificity in minor groove recognitions by small polyamide molecules. Examination of the complex between thiazotropsin B and d(CGACGCGTCG)$_2$ (see Section 4.2) solved by NMR spectroscopy revealed that there is little disruption to the DNA backbone compared with thiazotropsin A binding, and that the DNA helical parameters are similar to the free ODN, including the width of the monor groove. This suggests that the entropic penalties that arise through conformational distortion and induced fit seen with thiazotropsin A binding are not evident with this complex.

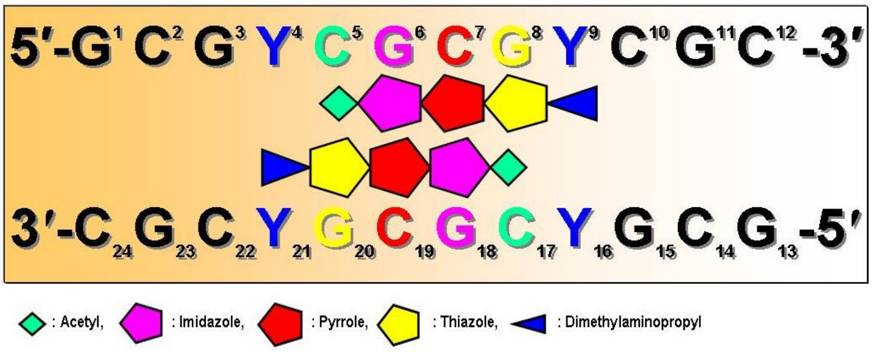


Figure 3.21 A schematic representation for the binding of thiazotropsin B with dodecamers containing the central sequence YCGCGY, where Y is T or A.

Further observations worthy of note are firstly that the binding affinity of thiazotropsin B to 5′-ACGCGT-3′ is higher than its affinity to 5′-TCGCGA-3′. This may be related either to optimised van der Waals interactions between the methyl group at the acetyl head of thiazotropsin B and the lipophilic H2 hydrogen of adenine at positions A^{4} and A^{16}, and/or the ACGCGT frame may provide better hydrogen bond contacts with the ligand compared with TCGCGA (Figure 3.21). Secondly, in contrast to thiazotropsin A, thiazotropsin B cannot recognise the five base pair binding site, 5′-ACAGT-3′, with high affinity. This may be due to either ligand-DNA repulsion forces between the unshared pair of electrons of the imidazole nitrogen and the O2 oxygen of cytosine at positions C^{5} and C^{16} (Figure 3.22), and/or inter-ligand-ligand repulsion forces between the unshared pair of electrons of the imidazole nitrogen and thiazole nitrogen which overlay on top of each other in case of a thiazotropsin B overlapped dimer configuration. This makes the slipped dimer of thiazotropsin B the preferred configuration for the dimeric recognition of DNA minor groove sequences (Figure 3.21).

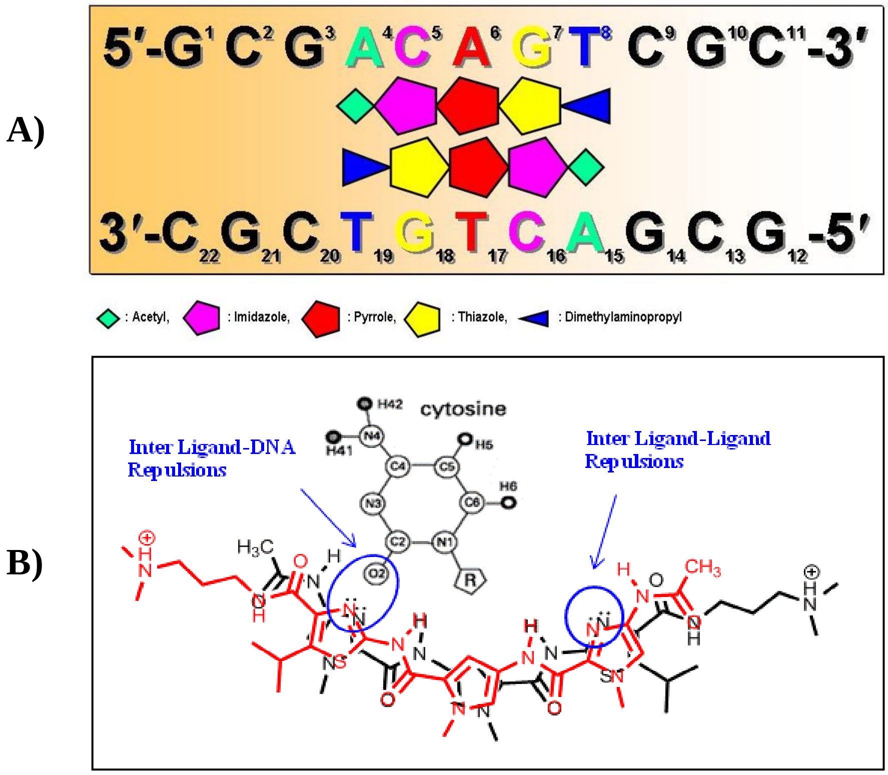


Figure 3.22 A) Schematic representation for the binding of thiazotropsin B with ACAGT site. **B)** A schematic shows the possible inter ligand-DNA and ligand-ligand repulsion forces caused by the unshared pairs of electron of ligand (*N*-imidazole) and DNA (O2 oxygen of cytosine) in thiazotropsin B- ACAGT association.

3.2.4 ITC of AIK18-51-DNA interaction

ITC has been used to characterise the binding of AIK18-51 with a dodecamer containing the central sequence 5′-ACTAGT-3′ (Figure 3.23). Dilution peaks were all endothermic, and their intensity decreased as more ligand was added to the buffer. Like P22, the ITC profile of AIK18-51 has two apparent phases indicative of two distinct binding events; an initial exothermic specific minor groove binding event, and a second endothermic non-specific backbone binding event before saturation of the first site. We were unable to obtain the thermodynamic profile of AIK18-51-ACTAGT association because the two binding events overlapped and fitting the ITC data was not possible. The ITC profiles of AIK18-51 and P22 are similar in that they both have two distinct binding events, however, they are well separated in the P22-ACTAGT association. The appearance of the second binding event of AIK18-51 at lower stoichiometry (2:1) compared with P22 (5:1) suggests that the driving forces for the second binding event are stronger in the case of AIK18-51. The molecular forces that drive the second endothermic binding event are the hydrophobic and the electrostatic interactions The higher lipophilicity of the AIK18-51 structure, which has an isopropyl substituted thiazole and an extra pyridine ring, compared with the structure of P22, which has an isopentyl substituted thiazole, seems to enhance the hydrophobic interactions between AIK18-51 and the backbone of the DNA helix. This may explain the higher efficiency of AIK18-51 in competing with counterions to neutralise the negatively charged backbone of DNA. These results emphasise that hydrophobic interactions play a key role in the occurrence of the second binding event as well as electrostatic interactions. The lack of samples did not allow us to repeat the ITC experiment as a function of salt concentrations in order to separate the two binding events and to obtain the thermodynamic parameters for AIK18-51-DNA interaction.

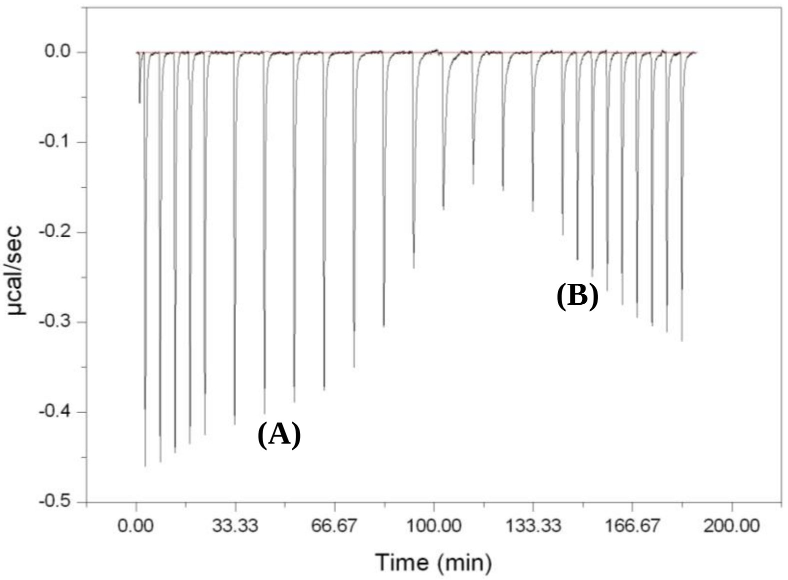


Figure 3.23 ITC profile for the titration of thiazotropsin B into a solution of dodecamers containing the central sequence GCG**ACTAGT**CGC at 25 °C and pH of 6.8. Each heat burst curve is the result of a 10 µL injection of 0.5 mM ligand into 15µM of DNA. The enthalpogram shows two distinct binding modes; exothermic binding process **(A)** followed by endothermic one **(B)**.

3.2.5 ITC of HA10-DNA interaction

We attempted to synthesise an alkene linked analogue of thiazotropsin A by replacing one of the amide links with an alkene group to investigate the effect of removing a hydrogen-bonding functionality on the binding affinity. However, because we were unable to introduce the formyl head onto the ligand (see Section 5.3.4), we decided to test the binding characteristics of the synthetic precursor HA10. ITC has been used to characterise the binding of HA10 to ODN duplexes containing the central sequences 5′-ACTAGT-3′, 5′-TCTAGA-3′, 5′-GCTAGC-3′, and 5′-CCTAGG-3′, in PIPES buffer at a constant Na^+ concentration of 20 mM and a constant pH of 6.5 (Figure3.24). In comparison with the other ligands, the dilution peaks of HA10 were all exothermic, and their intensity decreased as more ligand was added to the buffer, indicating that no aggregation of the HA10 occurs in buffered solution. Furthermore, the ITC results showed that there was no binding between HA10 and the four ODNs. Given that the heterocyclic units are compatible with the

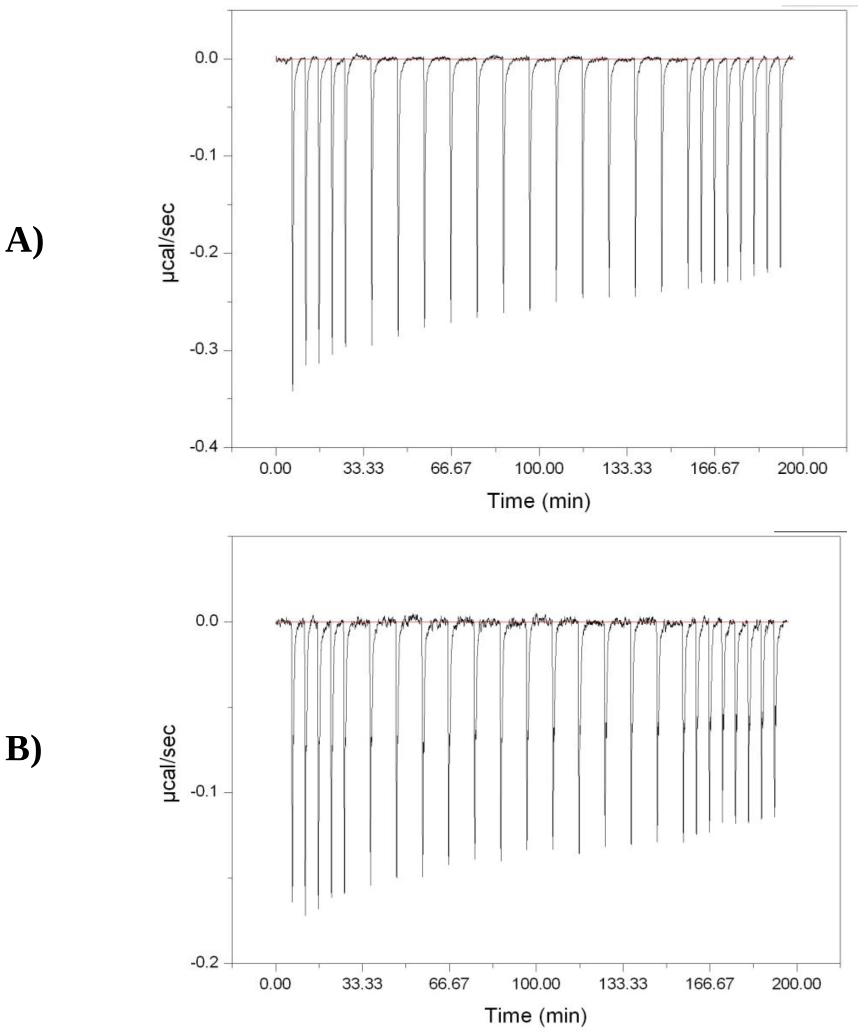


Figure 3.24 A) ITC dilution heats of HA10 in PIPES buffer. Each heat burst curve is the result of a 10 µL injection of 0.5 mM ligand into the PIPES buffer solution. The solution conditions were 10 mM PIPES, 20 mM NaCl, and 1 mM EDTA. **B)** ITC titration of HA10 to a solution of GCGACTAGTCTC sequence in PIPES buffer at 25 °C and a pH of 6.5. Each heat burst curve is the result of a 10 µL injection of 0.5 mM ligand into 15µM of DNA.

central CTAG sequence in a slipped or overlapped binding motif, then there are two possible reasons why this ligand does not bind to the sequences it was assessed against: firstly, the amide link towards the *N*-terminus of the ligand has been replaced by an alkene moiety, and key hydrogen bonds with the N3 atoms of the central A^6 and A^{16} bases on the groove floor seen with thiazotropsin A (see Figure 3.1) are lost. Moreover, introducing an alkene group in the structure of the ligand might change its curvature and flexibility, which are important for fitting the ligand in the minor groove. The amide NH′s of the formyl head groups of thiazotropsin A form hydrogen bond with O2 atoms of the central T^5 and T^{15} bases of the decamer used in the NMR study[38] (Figure 3.1). HA10 has a nitro group in the head group position and is therefore unable to form an equivalent interaction. In fact, the nitro group, being a hydrogen bond acceptor, is more likely to be repelled by the O2 atoms of these bases, which could also hinder binding to the groove. Finally, the absence of aggregation could imply that the ligands are unable to associate in a side-by-side fashion, which also has implications for binding.

The introduction of lipophilic moieties into polyamide MGBs is important to improve their cell permeability; however, such modifications should not alter the binding affinity. We have demonstrated that the more lipophilic thiazole heterocycle and the introduction of lipophilic substituents such as isopropyl and isopentlyl groups can address these issues without being detrimental to binding, primarily because these groups point out the minor and play no hydrogen bonding role. However, our preliminary studies appear to suggest that the amide links play a fundamental role in maintaining binding affinity through hydrogen bonding, and are therefore not amenable to replacement. However, this can only be truly established with an appropriate amide as a head group so that is the effect of only one variable in the ligand structure being assessed.

3.2.6 Heat capacity changes associated with thiazotropsin A /GCGACTAGTCGC and thiazotropsin B/GCGACGCGTCGC complex formation

The heat capacity change (ΔC_p) associated with ligand-DNA interaction can be determined from the temperature dependence of the observed binding enthalpy using the standard relationship:

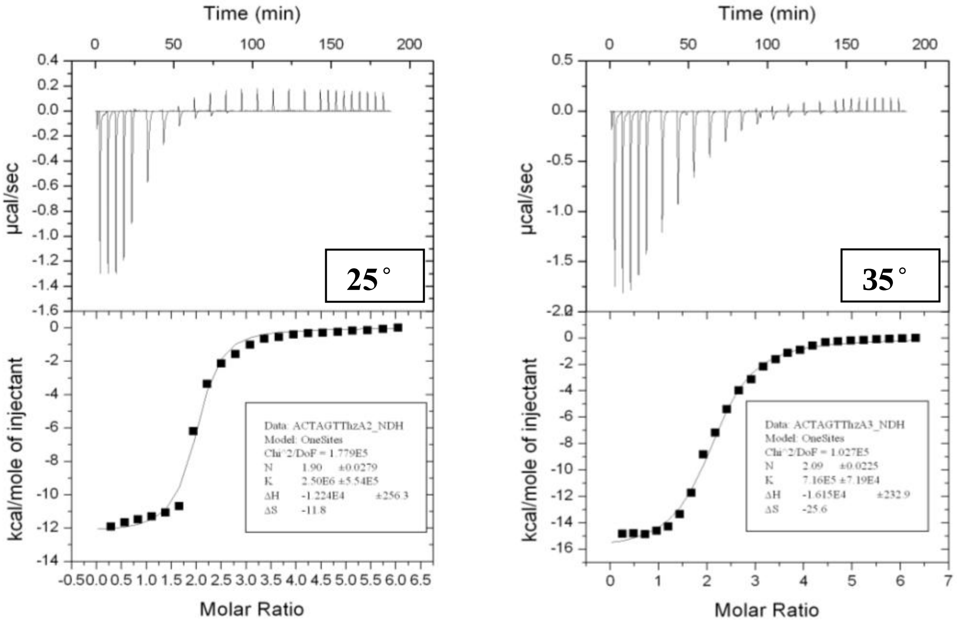
$$\Delta C_p = \frac{\partial \Delta H_{obs}}{\partial T}$$

Eq. 3.1

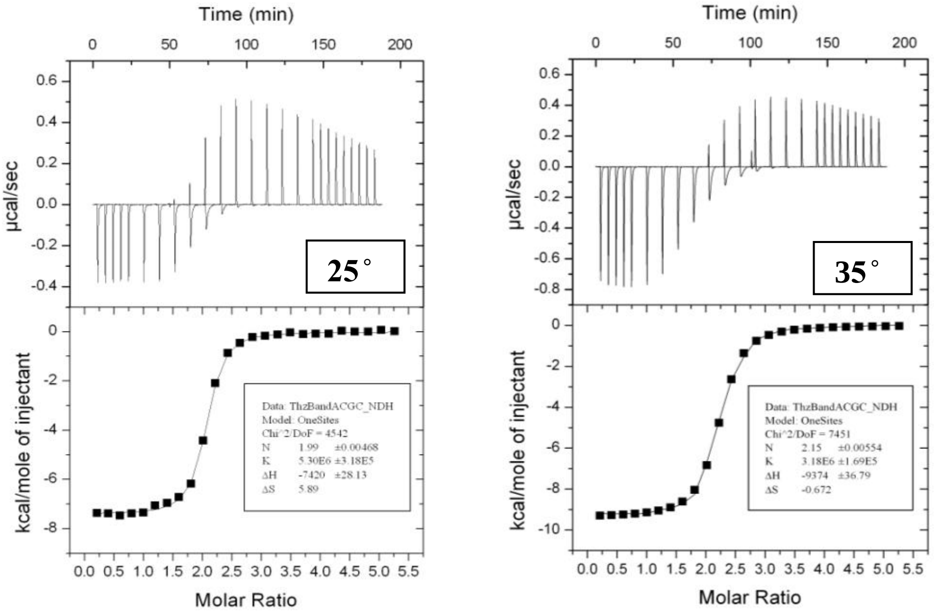
where $\partial \Delta H_{obs}$ is the difference between the enthalpy changes at two different temperatures and ∂T is the difference between the two temperatures.

We carried out ITC experiments for thiazotropsin A and B interactions with dodecamers containing the central sequences 5′-ACTAGT-3′ and 5′-ACGCGT-3′ respectively, both at 25°C and 35°C (Figure 3.25). Table 3.7 summarises the ΔH_{obs} values derived from the ITC profiles shown in Figure 3.25. Figure 3.26 graphically portrays the data in Table 3.7 in the form of ΔH_{obs} vs. temperature. The data points in these plots were fit by linear regression, with the slopes of the resulting lines yielding estimates of ΔCp for DNA binding of the two ligands. This analysis yielded ΔCp values of -391 and -195 cal $mol^{-1}K^{-1}$ for the binding of thiazotropsin A and B respectively (Table 3.7). The magnitude and sign of these ΔCp values fall within the range of -100 to -550 $calmol^{-1}K^{-1}$, which is typically observed for both ligand–nucleic acid as well as ligand–protein interactions.[63,72,86]

Reductions in solvent accessible surface are believed to have an impact on the value of ΔC_p. The burial of non-polar surfaces makes ΔC_p values more negative while the burial of polar surfaces makes them more positive.[63,72,86,138] The binding interactions accompanied with changes in solvent-accessible surface are usually associated with large positive changes in the binding entropy and slight positive changes in enthalpy as a result of releasing water from the minor grooves upon binding. It is tempting to attribute the negative ΔCp values we observe for thiazotropsin A–ACTAGT and thiazotropsin B-ACGCGT complexations to binding-induced reductions in nonpolar solvent accessible surface area. However, the ITC studies described earlier in this chapter showed little or opposing contributions from the binding entropy and huge negative enthalpy changes upon binding, an observation consistent with complex formation being accompanied by little or no changes in solvent- accessible surface. Hence, it is unlikely that our observed negative ΔC_p values reflect binding-induced alterations in solvent accessible surface.



(1) Thiazotropsin A-ACTAGT **(2) Thiazotropsin A-ACTAGT**



(3) Thiazotropsin B-ACGCGT **(4) Thiazotropsin B-ACGCGT**

Figure 3.25 ITC profiles for the binding of thiazotropsin A and thiazotropsin B to dodecamers containing the central sequences 5′-ACTAGT-3′ and 5′-ACGCGT-3′, respectively at 25°C (1 & 3), and 35°C (2 & 4). Each heat burst curve is the result of a 10 µL injection of ligand into a solution of 15 µM of DNA. The experimental solution conditions were 10 m*M* PIPES (pH 7), 1 m*M* EDTA, and 20 m*M* NaCl. The ITC experiments were conducted as described previously.

Table 3.7 Temperature dependence of the enthalpies and corresponding heat capacity changes for the binding of thiazotropsin A and thiazotropsin B to the binding sites; 5′-GCGACTAGTCGC-3′ and 5′-GCGACGCGTCGC-3′, respectively at pH7.

Ligand	***Temperature (^oC)***	***ΔH_{obs} ($Kcal.mol^{-1}$)***	***ΔCp ($cal.mol^{-1}.K^{-1}$)***
Thiazotropsin A	25	-12.24	-391
Thiazotropsin A	35	-16.15	
Thiazotropsin B	25	-7.42	-195
Thiazotropsin B	35	-9.37	

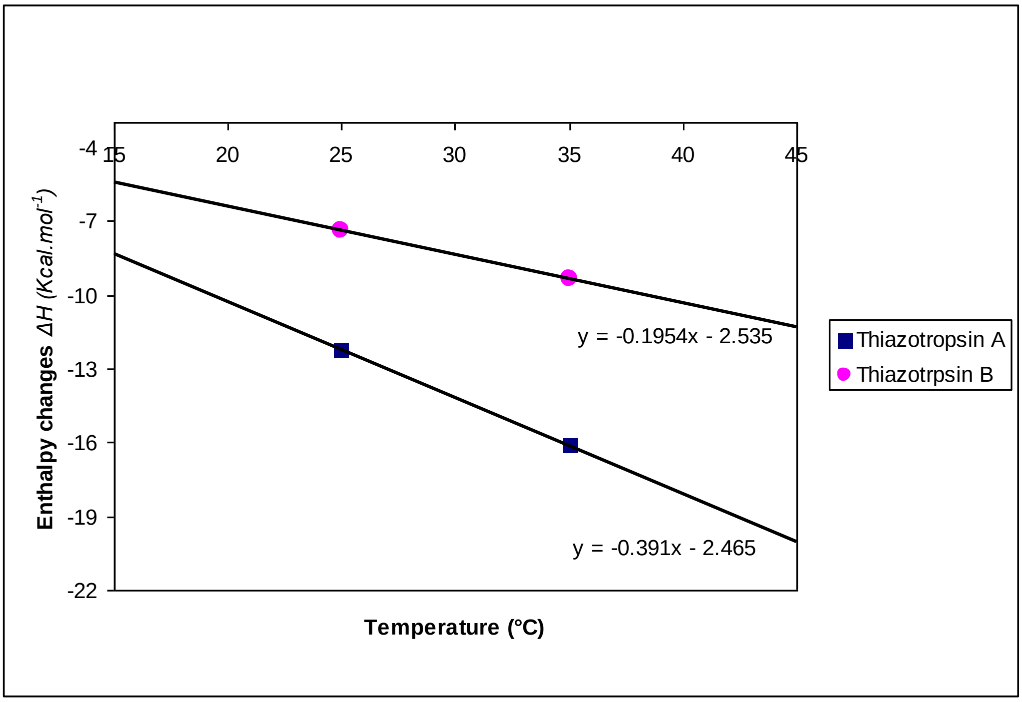


Figure 3.26 Temperature dependence of the observed enthalpies (ΔH_{obs}) for the binding of thiazotropsin A (filled squares) and thiazotropsin B (filled circles) to the binding sites 5′-GCG**ACTAGT**CGC-3′ and 5′-GCG**ACGCGT**CGC-3′, respectively, at pH 7. The experimental data points (which were derived from ITC experiments conducted in PIPES buffer at pH 7 and an NaCl concentration of 20 m*M*) were fit by linear regression (solid lines) and the values of *ΔCp* were obtained from the slope of the regression lines.

As detailed in section 1.6.1.4, there are many factors that have an impact on the observed ΔC_p. Electrostatic interactions can also have an impact on observed ΔCp, although Sharp *et al* [96] have shown that this impact is positive in sign and small in magnitude (15–90 cal $\text{mol}^{-1}\text{K}^{-1}$), which suggests our net negative ΔCp values are not the result of electrostatic interactions, although this does not discount the possibility that such interactions occur. Another potential contributor to ΔCp is the restriction of ″soft″ internal vibrational modes of polar groups and bound water molecules mediating the binding reaction. On being trapped in a ligand–DNA interface as part of a defined network of hydrogen bonds, water molecules and other polar groups can potentially have their ″soft″ vibrational modes dampened. This reduces their heat capacity, and ultimately increases the magnitude of the negative ΔC_p term for complex formation.[93,95] Binding-induced conformational changes can also contribute to ΔCp. In this regard, our structural studies[38] have shown that thiazotropsin A and B binding to DNA binding sites ACTAGT and ACGCGT induces a conformational changes in the backbone of the host DNA and the widening of the DNA minor groove. Eftink *et al.*[141] have shown that binding-linked conformational changes can result in negative heat capacity changes. We therefore suggest that our observed negative heat capacity changes reflect the restriction of ″soft″ internal vibrational modes of polar groups and bound water molecules mediating the binding interaction in addition to ligand-induced changes in the structure of target DNA. The inability of ΔC_p to provide information about solvation decoupled from other effects, and more importantly, the lack of samples did not allow us to carry out comprehensive heat capacity study with the other DNA sequences.

3.2.7 ITC study of ligand self association

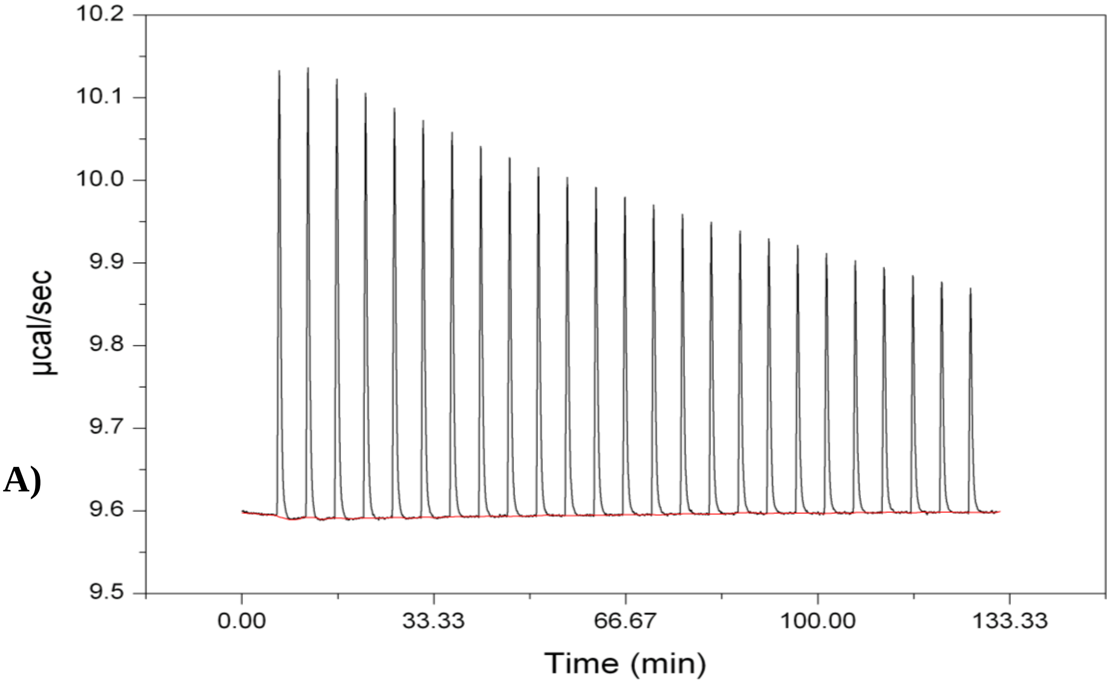
Dilution heats derived from control titrations of thiazotropsin A and its analogues; thiazotropsin B, and AIK18-51 into buffer alone were all endothermic and decrease consistently as more ligand is injected into the sample cell. Figure 3.27 shows representative ITC data for the dilution of 0.5 mM thiazotropsin A in PIPES buffer at pH 6.8. The variation of the dilution heat effects for thiazotropsin A and its analogues is consistent with dimerisation or step-wise aggregation as the mode of self-aggregation of the ligands studied here. Single dilution experiments cannot distinguish between dimerization and step-wise aggregation as the mode of lexitropsin aggregation. The temperature dependence of self-aggregation however, provides a solution. A Clarke-Glew plot for lexitropsin self-aggregation can be constructed using the equilibrium constants for both, assuming stepwise aggregation for one model, or assuming dimerization for a second model (Figure 3.28) by applying the Clarke-Glew equation:[142-143]

$$\mathrm{Ln}(K_T) = \mathrm{Ln}(K_{ref}) + (\Delta H_{ref}/R) * ((1/T_{ref}) - (1/T)) + ((\Delta Cp)/R) * ((T_{ref}/T) - 1 + \mathrm{Ln}(T/T_{ref})) \qquad \text{Eq. 3.2}$$

where K_T is the binding constant at time T, K_{ref} is the binding constant at reference temperature, ΔH_{ref} is the enthalpy change at the reference temperature, R is the gas constant, T_{ref} is the reference temperature, and ΔCp is the heat capacity change.

The heats of dilution of ligands studied were determined in PIPES buffer at a pH of 6.8 and at 25 °C, 35 °C and 45 °C. For all ligands analysed, ITC dilution experiments show a non-constant heat of dilution, indicating the occurrence of ligand self-aggregation in the buffered solution. The effect of ligand protonation or deprotenation on the enthalpies of dilution was excluded by conducting the dilution experiments under conditions where the ligands were monoprotonated (At pH 6.8) in two different buffers (PIPES and ACES). Different buffer types have different ionization enthalpies, and by measuring the enthalpies of dilution in different buffers, we can determine whether the self-association process is affected by ligand protonation or deprotonation. ACES buffer was used because it has a pK_a near to that of the PIPES buffer and a different enthalpy of (de)protonation. The enthalpies of lexitropsin self-interaction in PIPES and ACES buffers (Table 3.9) were essentially equal, indicating that aggregation was not accompanied by (de)protonation.

A)



B)

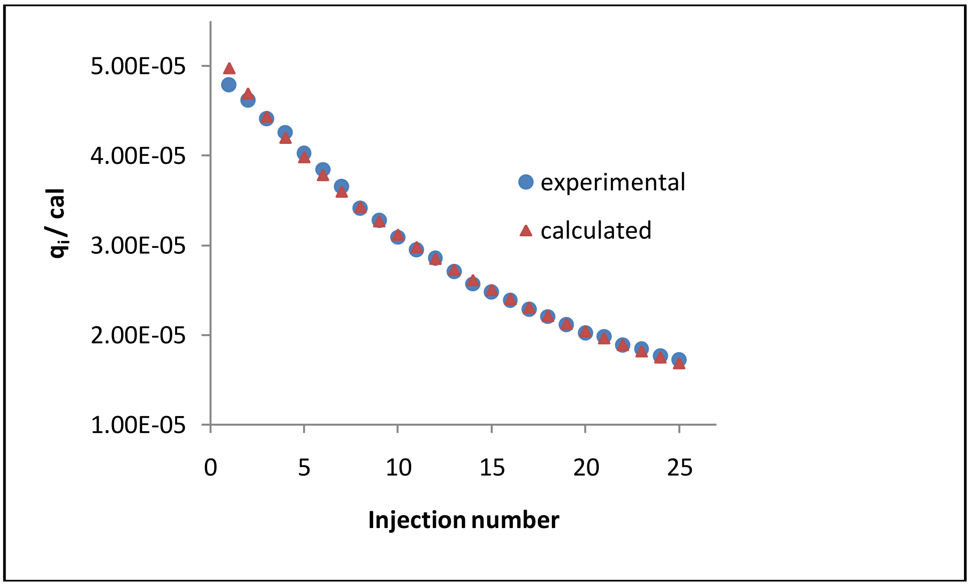


Figure 3.27 A) Representative example of the heats of dilution of 0.5 mM thiazotopsin A in 10 mM PIPES, 20 mM NaCl, 1 mM EDTA, pH 6.8 at 25 °C **B)** A comparison between the experimental and calculated heats of dilution using the IC ITC program to fit the data.

The experimental heats of dilution of lexitropsin ligands were processed using the IC ITC software package[128,144] for analyzing ITC data. This program generates values for the binding constants and enthalpy changes for both the dimerisation and step-wise aggregation models. The heat capacity changes were determined from the temperature dependence of the observed binding enthalpy at 25, 35, and 45 °C. All the thermodynamic parameters for the step-wise aggregation and dimerisation models of thiazotropsin A, thiazotropsin B and AIK18-51 are listed in Table 3.8. Using the equilibrium constants at 35 °C as a reference, together with the enthalpies and heat capacities of self-aggregation, we calculated predicted values for $\ln(K)$ between 20 °C and 50 °C using the Clarke-Glew equation.[142-143] The results revealed that the step-wise aggregation model reproduces the observed data for thiazotropsin A and thiazotropsin B better than the dimerisation model, while the dimerisation model fits the experimental data for AIK18-51 better than the step-wise aggregation model (Figure 3.28). These results show the ability of AIK18-51 to form dimers in the buffered solution, which could be driven by the enhanced stacking of aromatic rings through the additional pyridine ring. The higher association constant (K) of AIK18-51 compared with thiazotropsin A and B (Table 3.8) also supports this conclusion. The negative enthalpy values observed for both the stepwise self aggregation and the dimerisation models indicates that ligand self association is driven by either van der Waals forces between the aromatic rings of the ligand and/or the hydrogen bonding between the amide links.

Table 3.8 ITC-derived thermodynamic data for the step-wise aggregation and dimerisation models of thiazotropsin A, thiazotropsin B and AIK18-51.

Ligand	Tem °C	Step-wise aggregation model				Dimerisation model			
		ΔH (*Kcal.mol^{-1}*)	*TΔS* (*Kcal.mol^{-1}*)	K_{agg} (10^3 M^{-1})	*ΔCp* (*cal.mol^{-1}.K^{-1}*)	*ΔH* (*Kcal.mol^{-1}*)	*TΔS* (*Kcal.mol^{-1}*)	K_{dim} (10^3 M^{-1})	*ΔCp* (*cal.mol^{-1}.K^{-1}*)
Thiazotropsin A	25	-13.4	-8.7	2.6	-637.2	-26.4	-22.1	1.3	-1076
	35	-20.0	-15.9	0.8		-38.7	-35.1	0.4	
	45	-26.5	-22.9	0.3		-47.9	-44.7	0.2	
Thiazitropsin B	25	-17.0	-12.3	3.0	-721	-35.3	-30.9	1.5	-1219
	35	-23.8	-19.5	1.1		-50.2	-46.4	0.5	
	45	-31.4	-27.6	0.4		-59.9	-56.5	0.2	
AIK18-51	25	-9.6	-3.2	49.9	-328.1	-19.6	-13.6	24.9	-733.1
	35	-10.8	-4.9	16.2		-21.9	-16.4	8.1	
	45	-16.2	-11.2	2.5		-33.9	-29.4	1.2	

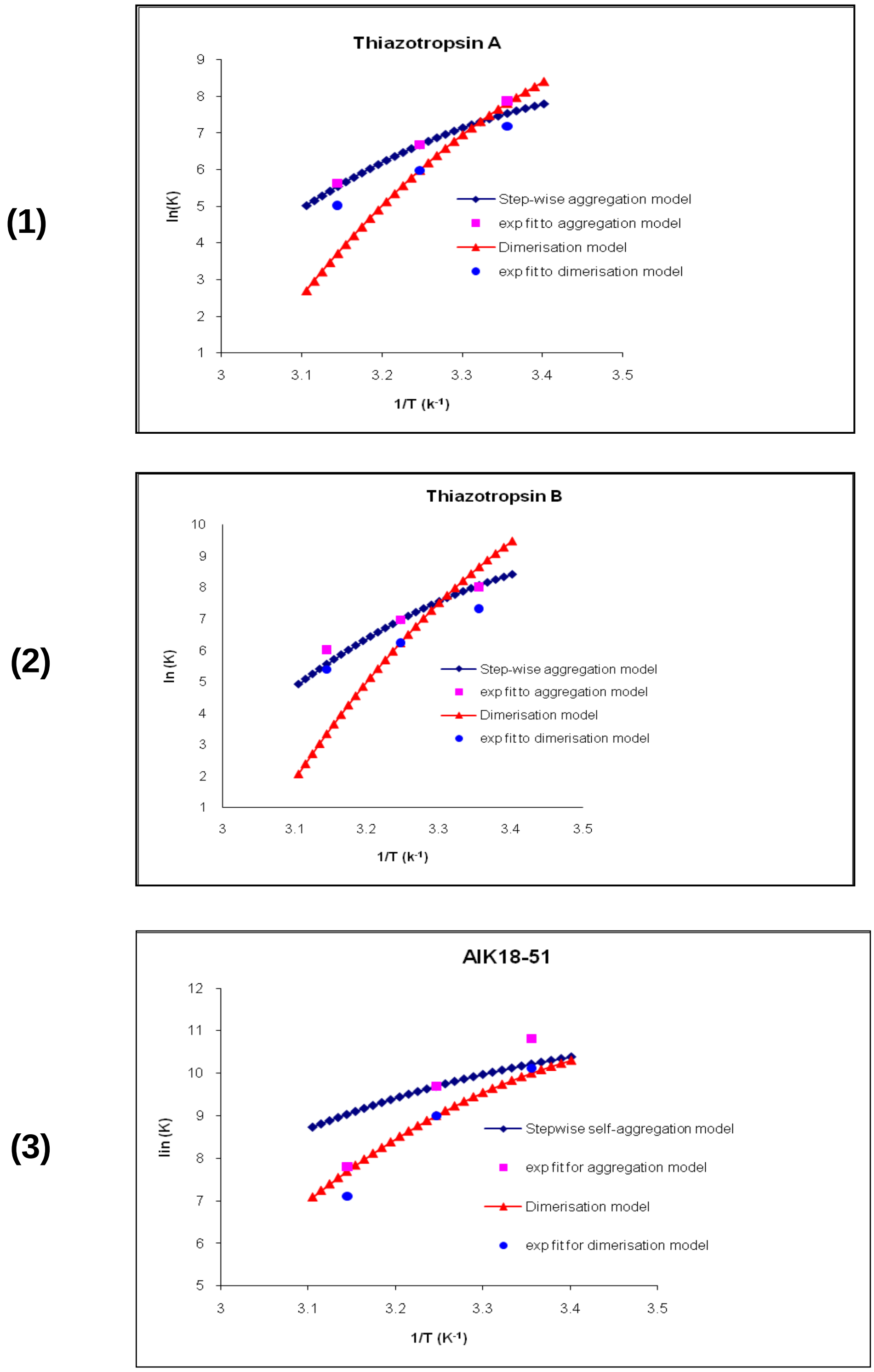


Figure 3.28 The Clarke-Glew plots for **(1)** thiazotropsin A, **(2)** thiazotropsin B, and **(3)** AIK18-51 self-aggregation assuming step-wise aggregation (squares) or dimerization (circles).

Table 3.9 Enthalpies for thiazotropsin A, thiazotropsin B and AIK18-51 in PIPES and ACES buffer. The differences in the enthalpy of lexitropsin self-interaction in PIPES and ACES buffers were less than the difference between enthalpy of (de)protonation of PIPES and ACES buffers indicating that aggregation was not accompanied by (de)protonation.

Ligand	**ΔH (PIPES)** **(*Kcal.mol^{-1}*)**	**ΔH (ACES)** (*Kcal.mol^{-1}*)	**(ΔH_{PIPES} - ΔH_{ACES})** **(*Kcal.mol^{-1}*)**
Thiazotropsin A	-13.4	-15.2	1.8
Thiazotropsin B	-17.0	-15.5	-1.5
AIK18-51	-19.6	-21.7	2.1
Buffer*	-2.7	-7.5	4.8

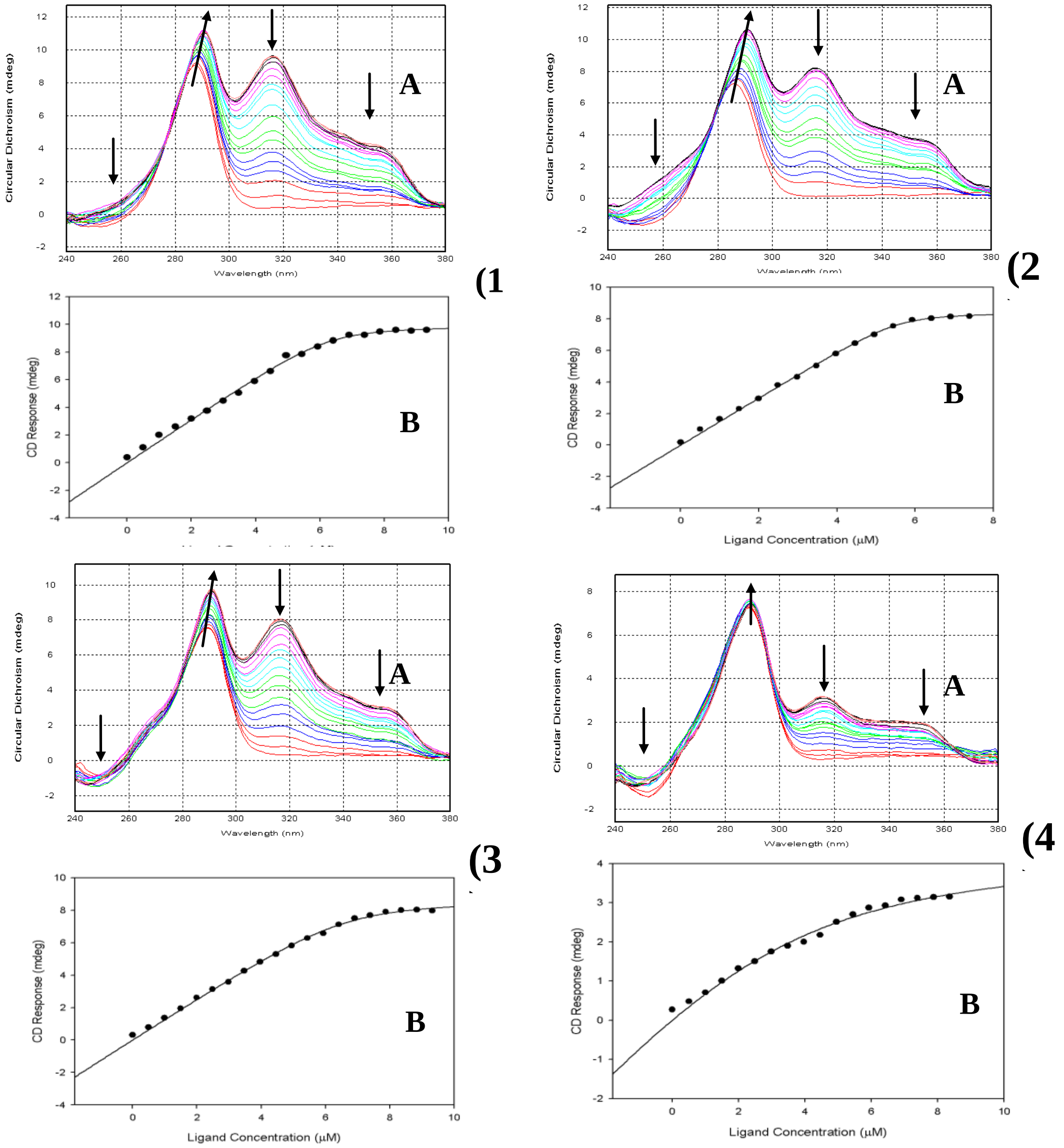
* The enthalpy of (de)protonation of PIPES and ACES buffers.[145]

3.3 Circular dichroism (CD)

3.3.1 Results of CD studies

CD was used to test the reproducibility of the ITC data obtained for the interaction of thiazotropsin A with five dodecamers containing the central sequences 5′-ACTAGT-3′, 5′-TCTAGA-3′, 5′-GCTAGC-3′, 5′-CCTAGG-3′, 5′-CCTAGI-3′.

Under CD analysis, each ODN in PIPES buffer demonstrates typical B-type DNA tertiary structure, namely a negative peak at 255 nm and a large positive peak at 285 nm (Figure 3.29 A). Upon addition of thiazotropsin A, the CD spectrum of each ODN exhibits changes consistent with a widening of the DNA minor groove. For all five ODNs, the negative peak at 255 nm becomes positive with increasing ligand concentration, and the new positive peaks at 316 and 355 nm that increase incrementally with ligand concentration represent minor groove binding. There are, however, some differences between the five ODNs (Figure 3.29 A). Binding of thiazotropsin A to the TCTAGA and ACTAGT ODNs induces a bathochromic shift of the positive peak at 285 nm to 290 nm. For GCTAGC the shift is less pronounced and for CCTAGG there is neither a change in molar ellipticity nor the wavelength of



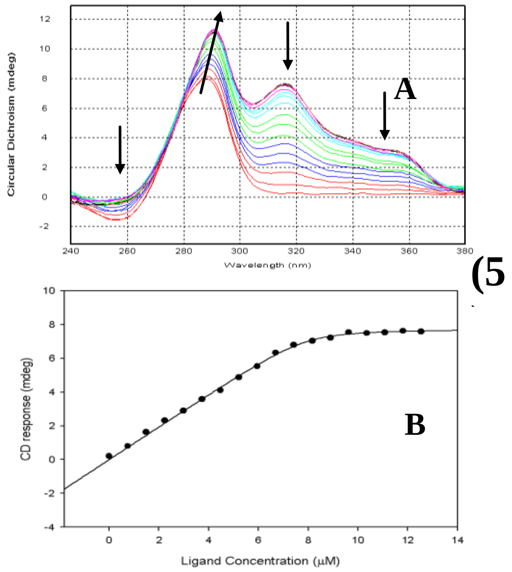


Figure 3.29 A) The CD spectra of thiazotropsin A titrated with five DNA sequences containing the central sequences: **(1)** ACTAGT, **(2)** TCTAGA, **(3)** GCTAGC, **(4)** CCTAGG, and **(5)** CCTAGI. **(B)** The increase in ellipticity at 316 nm as a function of added thiazotropsin A.

the peak at 285 nm. These results are consistent with the NMR and molecular modelling results, indicating that thiazotropsin A binds each sequence in slightly, but sometimes significantly, different ways.

A plot of the molar ellipticity at 316 nm against ligand concentration yields a typical binding curve for all five ODNs (Figure 3.29 B), from which binding constants and binding free energies could be calculated (Table 3.10), and the order of binding was reproduced by both CD and ITC.

Table 3.10 A comparison of the experimental binding constants and binding free energies obtained from ITC and CD for thiazotropsin A complexed with five different DNA sequences containing the central CTAG motif.

Sequence	**Data generated by ITC**		**Data generated byCD**	
	K **(*M^{-1}*)**	**ΔG** **(*Kcal.mol^{-1}*)**	**K** **(*Kcal.mol^{-1}*)**	**ΔG** **(*Kcal.mol^{-1}*)**
GCG**ACTAGT**CGC	$3.0\text{x}10^{7}$	-10.2	$3..0\text{x}10^{7}$	-10.2
GCG**TCTAGA**CGC	$2.0\text{x}10^{7}$	-9.9	$2.6\text{x}10^{7}$	-10.1
GCG**GCTAGC**CGC	$9.3\text{x}10^{6}$	-9.5	$7.9\text{x}10^{6}$	-9.4
GCG**CCTAGG**CGC	$1.7\text{x}10^{6}$	-8.5	$1.1\text{x}10^{6}$	-8.2
GCG**CCTAGI**CGC	$1.1\text{x}10^{8}$	-11.1	$4.6\text{x}10^{7}$	-10.5

3.4 Discussion of the thermodynamic evaluation of ligand-DNA binding

The titration of thiazotropsin A and its analogues with the DNA dodecamers were clearly associated with an exothermic process together with an endothermic dilution process indicative of aggregation of these ligands prior to DNA binding. A favourable enthalpy of interaction (e.g. ΔH of -12.9 kcal mol^{-1} for the thiazotropsin A-ACTAGT complex) is in agreement with the observation that exothermic interactions occur for the vast majority of ligands binding to DNA at room temperature.[91] This favourable ΔH could be explained by the ability of the amide links and the heterocyclic nitrogen of the thiazole/imidazole rings to form hydrogen bonds with the DNA bases on the groove floor, in addition to van der Waals contact between the aromatic rings and their substituents and the minor groove walls. The unfavourable entropies (e.g. $T\Delta S$ of -2.7 kcal mol^{-1} for the thiazotropsin A-ACTAGT complex) means that the association induces conformational changes in the ligand or DNA; restricts the conformational freedom of the complex; and/or does not significantly desolvate the binding site. If this association is not accompanied with the release of water from the minor groove that normally characterizes such binding, then the water molecules could become trapped in the complex and facilitate hydrogen bond formation. Water can contribute to the formation of ligand-DNA complexes through two different ways; firstly, the disruption of the solvent cage around a nucleic acid is entropically favourable and promotes binding affinity;[73,91] and secondly, bridging between the nucleic acid and the ligand through hydrogen bonding is enthalpically favourable and assists complex formation.[74-77] The large favourable ΔH and the unfavourable or slightly favourable ΔS suggests that in the systems we have studied, water assisted the complex formation by bridging through hydrogen bonding and van der Waals interactions. Besides trapping some water molecules at the binding interface, this mode of binding is also expected to release water from the minor groove to assist binding. However, this entropic contribution to binding may not be evident if it is offset by greater opposing factors such as the loss of conformational freedom. This observation does not agree with Chaires,[91] who reviewed a large amount of calorimetric data on groove-binding and intercalating ligands and concluded that minor groove binding was entropically driven, and was unrelated to structure - all minor groove binding compounds exhibited the same binding behaviour despite their large structural variations. Recent studies of HIV-1 protease inhibitors have found that a slight modification to the structure of a ligand can lead to a completely different thermodynamic profile.[78] The suggestion that all minor groove-binding is entropically driven without linking this behavior to specific structural ligand features raises doubts about the validity of this general assumption. Furthermore, some of the data presented in Chaires′ review suggests the opposite is the case: for example, the enthalpic contribution (ΔH -5.8 kcal /mol) of netropsin binding is more than the entropic contribution ($-T\Delta S$ -2.8 kcal/mol). Similarly, the monomeric association of distamycin with DNA has a greater enthalpic contribution than entropy. In the original reference for distamycin binding, the 2:1 binding mode to the minor groove was driven by large favourable enthalpy contributions (ΔH -15.7 kcal /mol) which was significantly opposed by entropy (TΔS -7.8 kcal/mol)[73] and was not commented upon. Therefore, the claim that ″all groove binders have positive (favourable) binding entropies and in no case is the -TΔS term positive.″[91] is not justified. The published literature to date has shown that the minor groove recognition by small molecules can be enthalpically or entropically driven, or both and that the thermodynamic signature of MGBs is highly dependent on ligand structure and the sequence of the binding site, which all of our studies with different sequences and related ligands confirm.

Analysis of the binding isotherms resulting from the titration of thiazotrosin A and its analogues revealed that when $r \leq 2$, the ΔH values remained constant for the binding of these ligands with their target sequence. The ligands bind exclusively in a dimeric 2 : 1 mode, which is consistent with the previous NMR study,[38] and agrees with the detailed comparative study of the thermodynamics of distamycin and netropsin binding performed by Lah and Vesnaver.[87] Moreover, Lah and Vesnaver's study also revealed that distamycin binding in a 2:1 fashion is characterized by a strong enthalpy of similar magnitude to thiazotropsin A (*e.g.* $\Delta H = -12.5$ kcal mol^{-1}) and is also accompanied by a substantial unfavorable entropy contribution ($T\Delta S = -2.0$ kcal mol^{-1}). They also showed that binding free energy was dominated by a combination of non-covalent interactions such as hydrogen bond formation and van der Waals interactions, and through the hydrophobic transfer of the ligand from the surrounding solution to its binding site within the DNA minor groove.

Our ITC studies have revealed that thiazotropsin A could recognise five or six base pairs in an overlapped or slipped binding mode. The slippage of the ligand aromatic rings expands the binding site size for these pyrrole-thiazole polyamides to six base pairs. In this binding mode, the Fmyl-Pyr-Pyr-Thz moieties of the two polyamides bind the central 5′-CTAG-3′ sequence in a 2:1 manner and the DMAP tails bind to the A,T,C, or G flanking bases in a non-adjacent binding mode . A similar observation was reported by Dervan[146] who found that the polyamide Im-Pyr-Pyr-B-Pyr-Pyr-Pyr-DMAP bound to 9 base pairs in the overlapped binding mode or to 13 base pairs in the slipped binding mode. In the ″overlapped″ (5 base pair) binding mode of thiazotropsin A, two Fmyl-pyr-pyr-thz-DMAP polyamides bind directly opposite one another. The fact that a single compound can bind in multiple binding modes is problematic when trying to acheive binding specificity, and it is ncessary to incorporate constraints in the structure of these compounds to specify a single binding mode. Dervan′s group achieved this by coupling the amino- and carboxyl termini of the antiparallel dimers with an aliphatic amino acid (γ) to create a U-shaped motif. The hairpin structure avoids ring slippage and keeps the rings unambiguously paired.[7,146] The downside to this approach is that these hairpin ligands have large molecular weights and high hydrophilicity, which makes them less dug-like for therapeutic applications.

Our synthetic polyamides bind to DNA as an antiparallel dimers and the pairs of Thz/Pyr or Thz/Fmyl recognize G•C, Pyr/Thz or Fmyl/Thz recognize C•G but not A•T or T•A; the Pyr/Pyr pair does not distinguish T•A and A•T but will not bind C•G or G•C; and the DMAP tail covers the flanking bases, A, T, C, or G. A reduced affinity was observed when the DMAP tail lies adjacent to G′s because of a steric clash with exocyclic NH_2 group. These findings are consistent with Dervan′s pairing rules for minor groove recognition.[7]

The isopropyl and isopentyl substituted thiazole motif of our synthetic polyimides exhibited similar binding behaviour to that of imidazole in targeting G′s specifically, and the enhanced lipophilicity of the substituted thiazole rings was intended to offer additional beneficial physical properties by enhancing their ability to cross biological membranes, improve bioavailability and hopefully lead to enhanced efficacy. From the standpoint of the rational design of minor groove binders, these moieties can be introduced into the polyamide structure without sacrificing hydrogen bonding as these substituents interact with the minor groove walls by van der Waals forces, while the NHs of the carboxamides and the thiazole nitrogen points toward the minor groove floor. Replacing the pyrrole ring with an imidazole ring in the structure of thiazotropsin B changed the sequence specificity of this ligand to ACGCGT. The results presented here reveal that the sequence specificity of these ligands is dictated by the NHs of carboxamide pyrrole or formyl which target the A•T/ T•A base pairs and the nitrogen of the thiazole/ imidazole rings which target G•C base pairs when adjacent to a pyrrole ring. There are slight variations in the binding affinity of the ligands studied, which is mainly dictated by the number of hydrogen bonds that can be formed between the ligand and the DNA binding site. As there are only small variations in the structure of our ligands, and all of them are nearly the same length that recognise binding sites of similar size, these ligands are expected to form equivalent numbers of hydrogen bonds, which explains the small differences in their binding affinity.

Enthalpy changes reflect the strength of the noncovalent interactions between molecules relative to those existing with the solvent.[71] Different molecular forces may contribute to the observed enthalpy, such as hydrogen bonding, electrostatic, and van der Waals interactions between the ligand and DNA, or ligand and solvent. Ligand–DNA interactions which have large contributions from the hydrophobic and electrostatic forces are mainly driven by entropy due to the release of water and counter ions from the poly anion DNA duplex upon ligand binding.[72-73] Those interactions in ligand-DNA complexes driven by entropy are characterised by a small positive or negative enthalpy and a large positive entropy. If electrostatic interactions play a role in binding, their contribution is expected to counteract the observed enthalpy by increasing the entropy and reducing the enthalpy. In general, our systems are characterized by large negative enthalpy changes with unfavourable entropies, which suggests that electrostatic effects are not a driving force in the interaction. Electrostatic and van der Waals interactions with the solvent are inevitable, however, since we have subtracted the heat of dilution of the ligand into the buffer alone from the heat of complex formation, we can be sure that the measured ΔH is due to lexitropsins-DNA interactions only.

Overall, the distinct thermodynamic signature of lexitropsin-DNA interactions allowed us to differentiate between the molecular forces responsible for the binding. The large favourable enthalpy (negative) indicates that there are a large number of favourable hydrogen bond contacts or van der Waals interactions between the DNA and ligand. The unfavourable entropy (negative) suggests a conformational change in both or either of the molecules that produce a more restrained complex through an ′induced fit′ process. This would appear to outweigh any solvent rearrangement, desolvation, release of counterions or hydrophobic drive that is characterized by favourable entropy measurements observed with some MGBs.[73] Establishing a link between the the energetics of binding and structure is important when trying to understand biomolecular interactions and improving the binding affinity. However, there are some drawbacks: improving binding enthalpy does not necessarily lead to a higher binding affinity because of enthalpy-entropy compensation, where enthalpy gains countered by entropy losses, leading to no net increase in affinity. One major cause of this compensation mechanism is the nature of non-covalent interactions. For example, the enthalpic gain via hydrogen bond formation within a complex is often accompanied by entropic loss as these new bonds limit movement within the complex.

4 CHAPTER 4: Structural elucidation of lexitropsin-DNA complexes

4.1 Introduction

The rational design of new drugs with novel biological activities and enhanced affinities for defined DNA sequences require detailed structural knowledge of the interactions responsible for complex formation. NMR spectroscopy is an important technique for the study of ligand-DNA interactions. The principles of NMR spectroscopy have been discussed in Chapter 1. A review of the applicability of NMR to the study of ligand-DNA interactions has been compiled by Lane.[107] This technique was used by Parkinson *et al* [38] to study the binding of thiazotropsin A to the self-complementary DNA decamer d(5′-CGACTAGTCG-3′)$_2$. Here the application of this technique to related ligand interactions with DNA is reported.

In this chapter, the interaction of thiazotropsin B and AIK18-51 (Figure 4.1) with the decameric ODNs d(5′-CG**ACTAGT**CG-3′)$_2$ and d(5′-CG**ACGCGT**CG-3′)$_2$, respectively, were examined by two-dimensional NMR spectroscopy in solution to obtain detailed structural information and to compare them with the previously determined structures. DNase I footprinting,[147] ITC[148] and NMR spectroscopy[147] studies have shown that thiazotropsin A binds to the hexanucleotide 5′-ACTAGT-′3 sequence. Replacing one of the *N*-methylpyrrole rings in the structure of thiazotropsin A with *N*-methylimidazole (thiazotropsin B) changes the preferred binding sequence to (A/T)CGCG(T/A).[129] The nitrogen of the imidazole and thiazole rings is responsible for dictating the specificity of these ligands for GC base pairs by forming hydrogen bonds with the amino group of G.

The NMR NOE derived inter-proton distances were used to generate three dimensional structures for these complexes by using restrained molecular dynamic simulations. The assigned 2D [^{1}H,^{1}H] NOESY NMR spectra of both the free and ligand-bound DNA were used to determine the location of the binding site of thiazotropsin B and AIK18-51 by measuring the change in chemical shift of DNA protons upon ligand binding. NOE connectivities observed between the protons of *N*-methyl, amide, DMAP, acetyl, isopropyl groups and the heteroaromatic rings on these ligands with the protons on the DNA also provided more detailed information on the location of the binding site, in association with the inter-ligand and ligand-DNA NOESY cross-peaks.

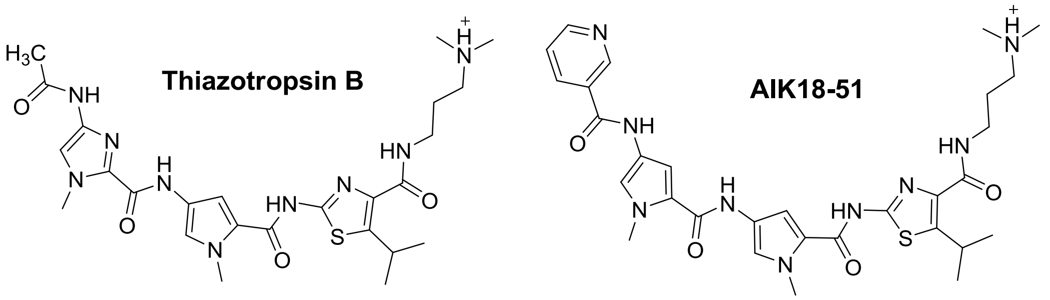


Figure 4.1 Structures of thiazotropsin B and AIK18-51.

4.2 NMR study of the thiazotropsin B-d(CGACGCGTCG)$_2$ complex

4.2.1 Titration of d(CGACGCGTCG)$_2$ with thiazotropsin B

The formation of the complex between **d(CGACGCGTCG)$_2$** and **thiazotropsin B** was carried out according to the experimental details described in section 2.3.5. 1D ^{1}H NMR spectra were acquired on the sample at various ligand/DNA ratios to allow titrations to be monitored and complex formation to be rapidly assessed. Upon addition of thiazotropsin B to the solution of d(CGACTAGTCG)$_2$, a number of new resonances appeared, most obviously in the imino proton resonance region. The intensity of these resonances continued to increase during the titration up to a 2:1 ratio of ligand to DNA, whilst the resonances of the free ODN completely disappeared at this point. In the presence of free DNA, the ^{1}H NMR spectrum simultaneously showed the presence of imino proton resonances from both free and ligand-bound DNA indicating a slow exchange between the free duplex and ligand-bound complex. The 1D ^{1}H NMR spectrum for the free duplex d(CGACGCGTCG)$_2$ is simplified, due to the self complementary nature of this sequence. The 1D ^{1}H NMR spectrum for the complex is also simplified when all the free duplex in solution is bound to ligand because two molecules of thiazotropsin B oriented themselves in opposite directions in the minor groove; the symmetry of the free duplex therefore was maintained in the complex which simplified the NMR spectrum. In the imino proton resonance region, the 1D ^{1}H NMR spectrum showed the presence of four imino proton resonances and three additional resonances from the ligand (Figure 4.2). This is expected for a symmetrical 10 base pair DNA sequence. The ligand resonances were anticipated to arise from N*H* protons associated with the ligands′ amide linkages (Figure 4.2d). Thus the complexity of the NMR spectrum alone was used as an indicator of the stoichiometry of interaction. The formation of a ^{1}H NMR spectrum, with only one peak corresponding to each ligand proton and the appearance of only four imino proton resonances for the DNA duplex, indicated that a 2:1 symmetric complex was being formed with no evidence for the presence of other complexes. An observation worthy of note is that some minor peaks remain present in the imino proton resonance region of the complex

(Figure 4.4). These are believed to originate from impurities in the free DNA sample as the 1D ^{1}H NMR spectrum of the free DNA sample showed traces of these impurities at 13.85 ppm. The source of these impurities may be short monomer DNA sequences arising from the solid phase synthesis of the oligonucleotide.

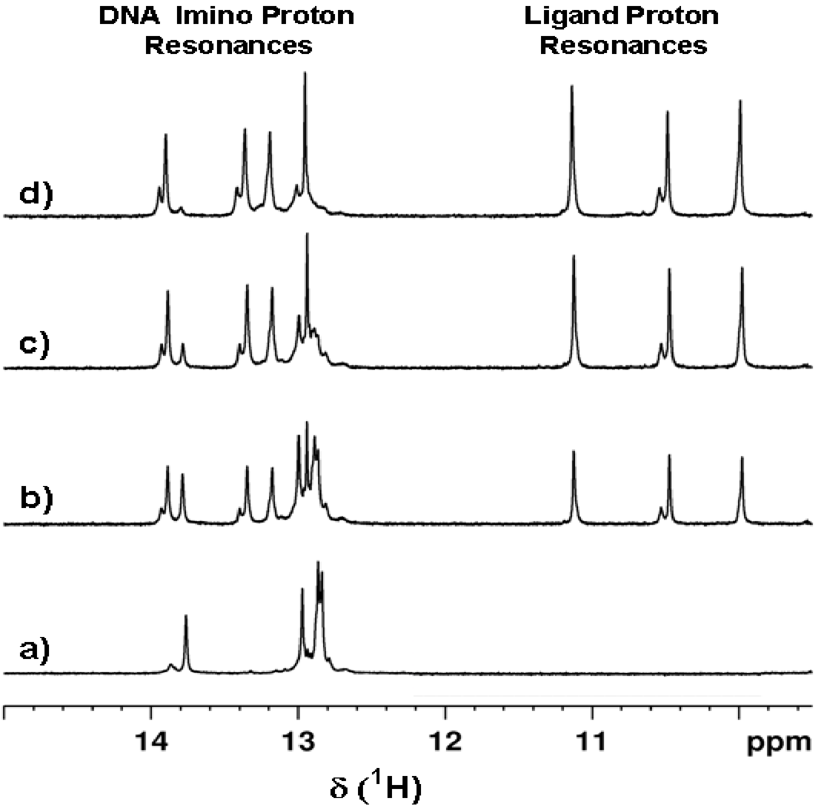


Figure 4.2 . Imino proton resonance regions of 1D ^{1}H NMR spectra acquired at 600 MHz using a dpfgse routine and showing the result of titrating a solution of thiazotropsin B into a sample of d(CGACGCGTCG)$_2$. The ^{1}H NMR resonances of the imino protons belonging to Watson−Crick base pairs are visible between 12.5 and 14 ppm. Resonances between 9.5 and 11.5 ppm are assigned to amide N*H* protons in thiazotropsin B. (a) Free DNA. (b) After addition of ca. 1 equiv of thiazotropsin B. (c) After addition of slightly less than 2 equiv of thiazotropsin B. (d) Exact 2:1 equivalence between thiazotropsin B and d(CGACGCGTCG)$_2$. Ligand residency time was relatively long as shown by the presence of free and bound forms of DNA at a ligand/DNA duplex ratio of ~1:1. Signal integration indicated that two molecules of thiazotropsin B occupied the DNA minor groove.

4.2.2 NMR data assignment

The NMR analysis of the complex between d(CGACGCGTCG)$_2$ and thiazotropsin B was undertaken by firstly characterising the free DNA duplex. This was followed by examination of the complex formed upon reaction with thiazotropsin B. The free and ligand bound d(CGACGCGTCG)$_2$ were fully assigned using a combination of through space and through bond connectivities from 2D [^{1}H, ^{1}H] NOESY, DQFCOSY, TOCSY, and [^{1}H,^{31}P] correlation NMR spectra.

Deoxyribose H1′ proton resonances of B-form DNA typically appear in the 5.5-6.5 ppm region of the ^{1}H NMR spectrum, while H6/H8 protons resonate in the range 7.0-8.5 ppm.[115] NOESY cross-peaks between H1′ and H6/H8 protons are usually well resolved. Therefore, the starting point for the assignment of the free and ligand bound duplex was the NOE assignment in this region of the NOESY spectrum. The assignments were made using the sequential walking procedure for right-handed B-form DNA. In this assignment, H6/H8 of the 5′- terminal residue interacts through space with H1′ of its own deoxyribose sugar only, whereas subsequent H6/H8 protons interact with both their own H1′ protons and with the sugar protons of the preceding residue. The H6/H8-H1′ assignment pathway can then be followed to its 3′-terminal H6/H8-H1′ cross-peak. H5 protons of cytosine bases resonate in the same frequency range as H1′ protons and show J-coupled cross-peaks with H6 in this region of the DQF-COSY NMR spectrum. These are easy to identify and give an indication of whether the initial assignment is correct or not. A portion of the 600 MHz, 100 ms, 2D [^{1}H, ^{1}H] NOESY NMR spectrum of d(CGACGCGTCG)$_2$ is shown with assignments in Figure 4.3 and the atom notations used in the NMR assignment of DNA protons are shown in Figure 4.4. The sugar protons H2′/H2″ interact through space with the aromatic protons H6/H8 and can also be assigned using a similar sequential walk to that used in the assignment of H1′ protons. H3′, H4′ and H5′/H5″ proton resonances are congested in the data and not well resolved. Assignment of these protons required the use of a combination of through space

(NOESY) and through bond (DQFCOSY, TOCSY, 2D [^{1}H, ^{13}C] and 2D [^{1}H, ^{31}P] HSQC) data.

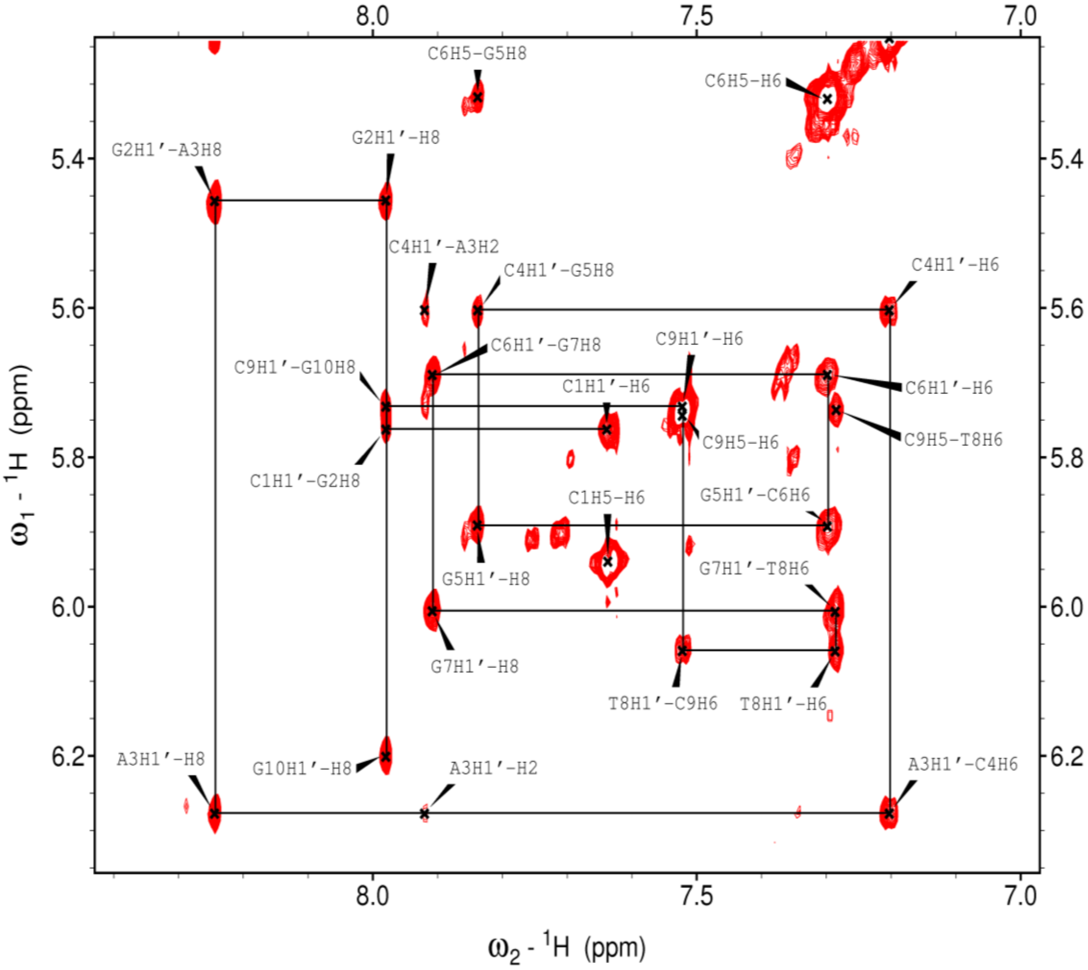


Figure 4.3. Fingerprint region of the 100 ms 2D [^{1}H, ^{1}H] NOESY NMR spectrum of d(CGACGCGTCG)$_2$ at 600 MHz in the absence of the ligand. Resonance assignments are shown and the assignment 'walk' indicated by horizontal and vertical lines, which join the NOEs between aromatic and sugar H1' protons.

Imino proton ^{1}H NMR signal assignments were initially made on the basis of saturation transfer effects (Figure 4.5), which were observed following a presaturation approach to solvent suppression. This provided an initial clue as to the assignment of these signals. A detailed NOE assignment was then made and used to confirm these speculative assignments. The exchangeable H1 guanine imino proton

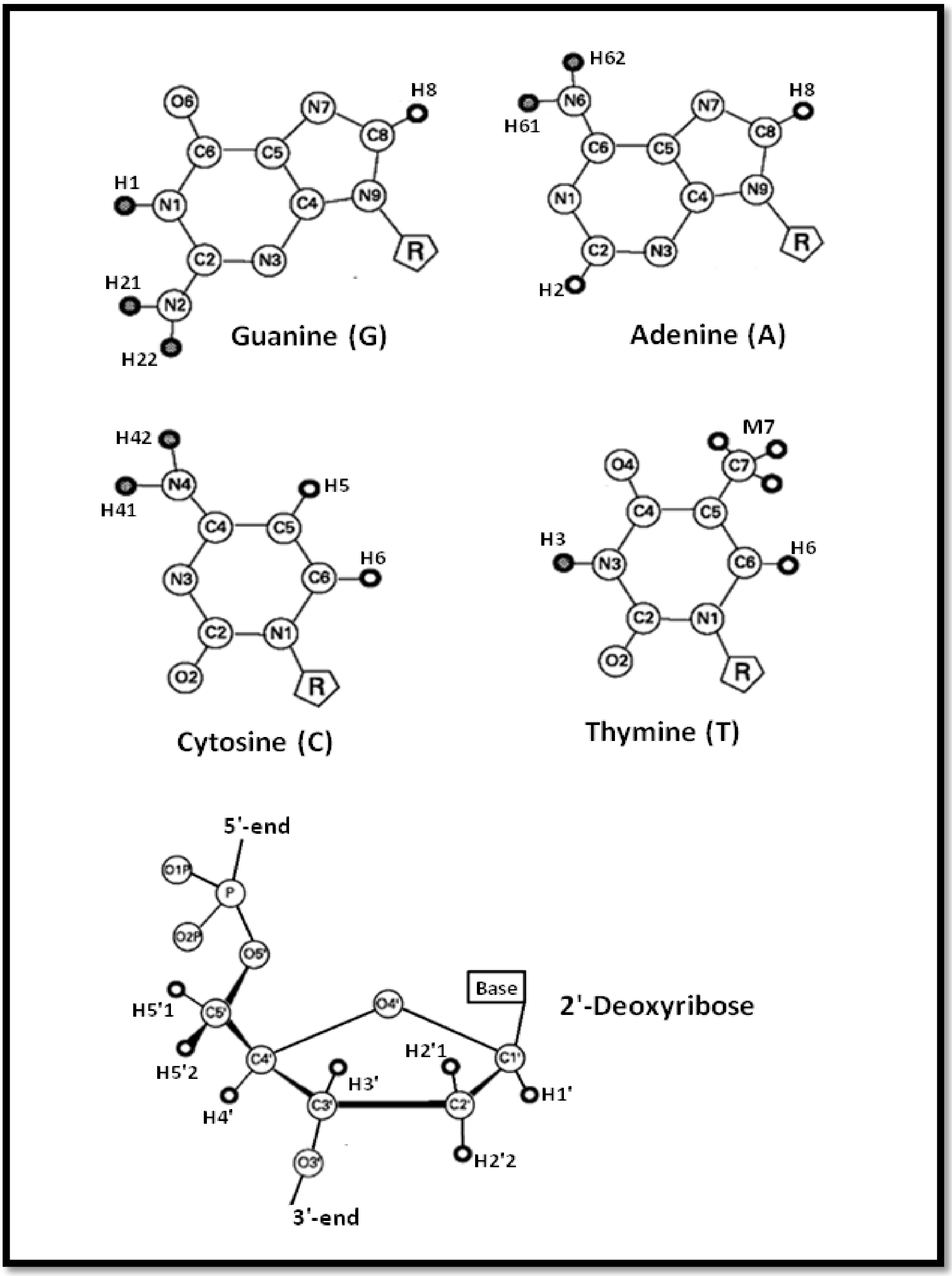


Figure 4.4 Atom notations for deoxyribose sugars and DNA bases used in both the NMR and the molecular modelling studies.[149]

resonances were identified from inter-strand NOE contacts with cytosine amino groups and intra-strand NOEs with its own amino group. H3 thymine imino protons were identified from inter-strand NOE contacts with adenine H2. The chemical shifts of the proton NMR resonances assigned for free d(CGACGCGTCG)$_2$ are listed in Table 4.1.

In the ligand-DNA complex, completion of the assignment of NOE cross-peaks between d(CGACGCGTCG)$_2$ and thiazotropsin B required the identification of intra-ligand, inter-ligand, intra-strand, inter-strand, and ligand−DNA NOEs from one another. Initially, various DQFCOSY correlations were used in order to distinguish key ligand resonances. DQFCOSY cross-peaks were observed between protons resonating at the following chemical shifts (Figure 4.6): 7.922 and 6.359 ppm (cross peak a, L H10 and H13, see numbering scheme on Figure 4.7); 3.30 and 8.40 ppm (b, L H231 & H22); 3.50 and 8.40 ppm (c, L H232 & H22); 2.99 and 8.92 ppm (d, L H27m & L H26); 3.19 and 8.92 ppm (e, L H25 and H26). These cross-peaks are uncommon in a DNA setting and were consequently assigned to thiazotropsin B. The DQFCOSY cross-peak at 3.30/8.40 (b) and 3.50/8.4 (c) ppm are mirrored by a strong NOE in the H_2O NOESY spectrum. On the basis of the peak shape, labile proton exchange characteristics, chemical shift location, and line width compared with its partnering resonances at 3.30 and 3.50 ppm, the signal at 8.40 ppm was assigned to an amide NH signal (namely H22 in thiazotropsin B). Hence, the signals at 3.30 and 3.50 ppm were assigned to H231 and H232 in the ligand, respectively. Similarly, strong DQFCOSY cross peaks were mirrored by strong NOE cross-peaks in the H_2O NOESY spectrum at 2.99/8.92 and 3.19 /8.92 ppm. The signal at 8.92 ppm was assigned to H26 (NH), and partnering signals at 2.99 and 3.19 were assigned to H27m (representing the equivalent methyl groups HM2 and HM3 of the DMAP tail) and H25 (representing both H251 & H252), respectively. The remaining DMAP methyl ^{1}H NMR resonances (H241 & H242) were assigned through correlation to H231 and H232 via DQFCOSY data and NOEs. Methyl group ^{1}H NMR methyl resonances from thiazotropsin B were visible in three regions of the ^{1}H NMR spectrum (Figure 4.8), namely, 3.96−4.0 ppm (*N*-methyl pyrrole), 2.99 ppm (*N*-dimethyl), and 1.44−1.63 ppm (isopropyl-CH_3 & acetyl-CH_3). The isopropyl methyl resonances were assigned from the DQFCOSY NMR spectrum, in which a strong

Table 4.1. ^{1}H Chemical shift assignments for the DNA duplex d(CGACGCGTCG)$_2$.

	Chemical shift : δ ^{1}H (ppm)														
	H8	**H6**	**H2**	**H5**	**CH$_3$**	**H1′**	**H2′**	**H2″**	**H3′**	**H4′**	**H5′**	**H5″**	**GH1/ TH3**	**H41**	**H42**
Base															
C1		7.635		5.941		5.763	1.893	2.384	4.711	4.078	3.728	3.728			
G2	7.980					5.456	2.742	2.811	5.015	4.322	3.983	3.983	13.01		
A3	8.244		7.920			6.275	2.729	2.928	5.08	4.497	4.242	4.176			
C4		7.203		5.237		5.603	1.939	2.339	5.081	4.281	4.162	4.162		8.196	6.509
G5	7.839					5.889	2.618	2.738	4.983	4.367	4.133	4.051	12.90		
C6		7.297		5.318		5.691	2.075	2.415	4.856	4.194	4.145	4.145		8.324	6.509
G7	7.908					6.005	2.655	2.815	4.978	4.394	4.149	4.088	12.88		
T8		7.286			1.458	6.058	2.093	2.475	4.88	4.222	4.141	4.141	13.79		
C9		7.521		5.739		5.731	2.049	2.397	4.871	4.141	4.081	4.081		8.690	7.077
G10	7.980					6.199	2.441	2.766	4.708	4.204	4.085	4.085			

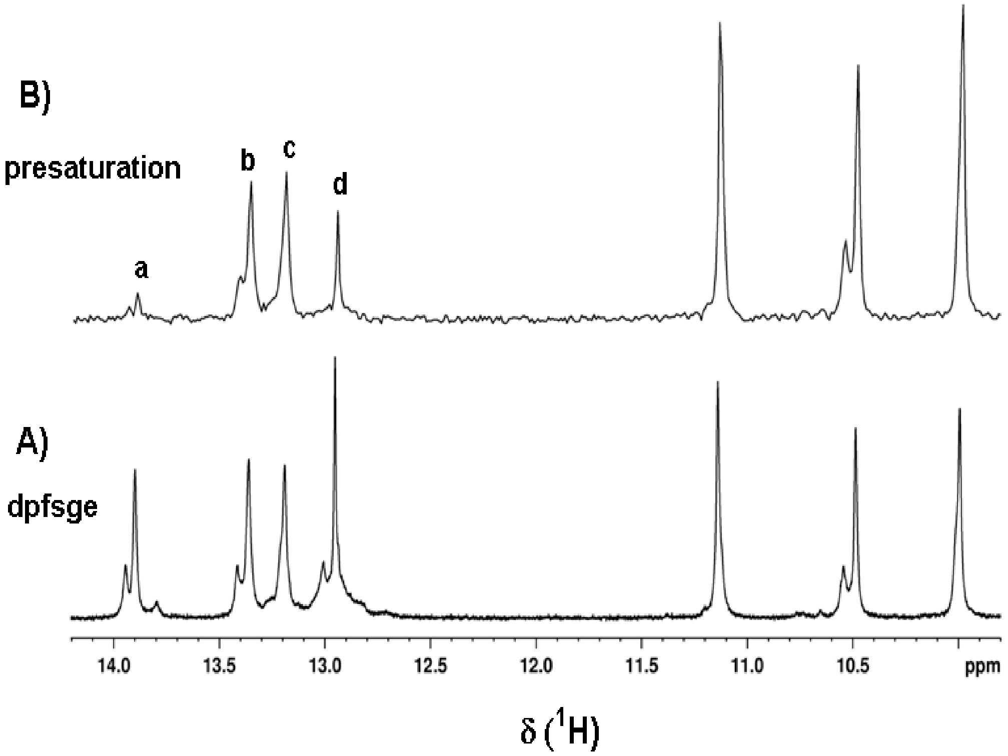


Figure 4.5 Imino proton resonance region of the ^{1}H NMR spectrum of the 2:1 complex between thiazotropsin B d(CGACGCGTCG)$_2$ **A)** using a dpfgse routine for solvent suppression and **B)** using solvent presaturation. Saturation transfer effects (shown by loss of intensity for signals **a** and **d** in **B**) enabled imino proton ^{1}H NMR resonance assignments to be made under the assumption that a greater degree of chemical exchange occurs with the solvent for protons nearest each end of the DNA duplex :**a**-T^{8}H3; **b**-G^{7}H1; **c**-G^{5}H1; **d**-G^{2}H1. These assignments were later confirmed by detailed analysis of NOESY NMR data. The peptide NH resonances of thiazotropsin B between 9.5-11.5 ppm are not influenced by saturation transfer effects indicating that they are protected from solvent exchange when the ligand is DNA bound since the atoms are buried on the minor groove floor, and thereby protected from solvent exposure.

correlation occurred between the methyl resonances at 1.448/1.547 ppm and a multiplet at 4.099 ppm (H20). A methyl resonance was observed in the δ^1H = 1.63 ppm region of the 1D ^{1}H NMR spectrum of the ligand bound d(CGACGCGTCG)$_2$, corresponding to the methyl proton resonance of DNA residues T8 and T18. The appearance of only one signal representing two thymidine units per duplex indicates that the symmetry for the ligand complex is maintained. The resonances at 7.922, 6.359, and 7.185 were assigned to pyrrole ″H10″, ″H13″ and ″H4″ protons in thiazotropsin B respectively. The assignments of these signals relied upon inspection of the relative sizes of NOEs associated with these resonances, together with NOE assignments of methyl and amide resonances from the ligands in addition to the observed DQFCOSY and TOCSY cross-peaks. A NOESY cross-peak was observed at 6.359/7.922 ppm and was associated with DQFCOSY (cross-peak **a**, Figure 4.6) and a TOCSY cross-peak. The signals at 11.140 and 9.995 ppm were associated with the amide NH protons, namely H14 and H8 of thiazotropsin B (Figure 4.9). The signal at 6.359 ppm shared strong NOEs with the signals at 11.140 ppm (H14) and 9.995 ppm (H8) (Figure 4.9). These NOEs were explained by assigning the signals at 6.359 ppm to H13. Consequently, the signal at 7.922 ppm was assigned to H10. The pyrrole methyl resonances were assigned through strong NOEs to their associated ring protons. Two large NOEs related to the pyrrole methyl of the ligand were observed at 4.001/7.922 and 3.961/7.185 ppm (Figure 4.11). The signal at 7.922 ppm was previously assigned to H10 and the signal at 4.001 was therefore assigned to HM6 (*N*-methyl). The second NOE at 3.961/7.185 ppm was then explained as a close contact between HM5 and H4. The acetamide head of the ligand was assigned through a very strong NOE between HM1 and H2 and other NOEs between H2 and both H4 and H8 (Figure 4.9). It was therefore possible to make a complete assignment of the ^{1}H NMR resonances of thiazotropsin B bound to DNA, full details of which are listed in Table 4.2.

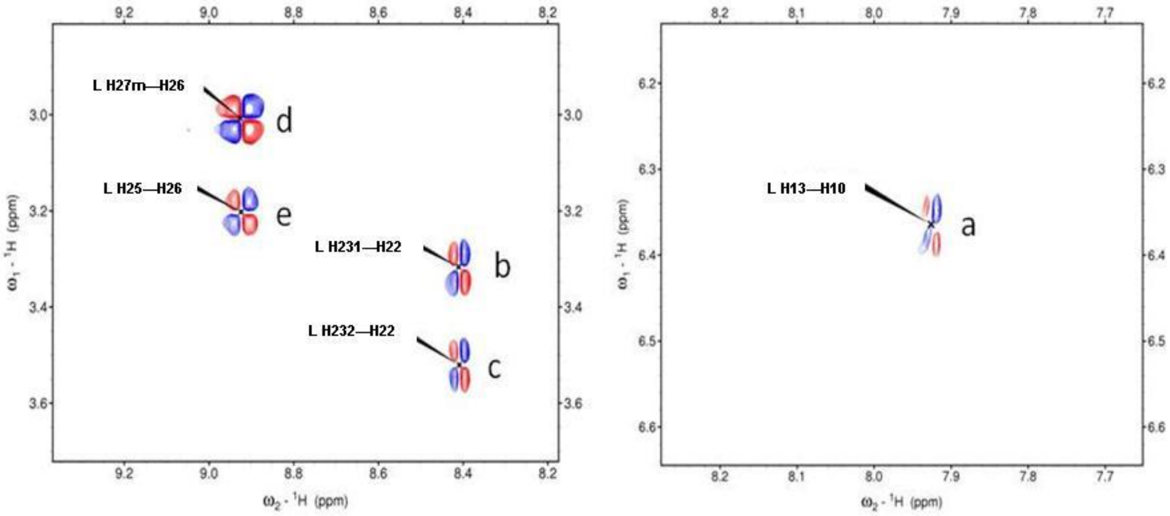


Figure 4.6 Regions of the 600 MHz 2D [^{1}H, ^{1}H] DQFCOSY NMR spectrum of the 2:1 complex of thiazotropsin B with d(CGACGCGTCG)$_2$ used for assigning ^{1}H NMR resonances to protons in thiazotropsin B when bound.

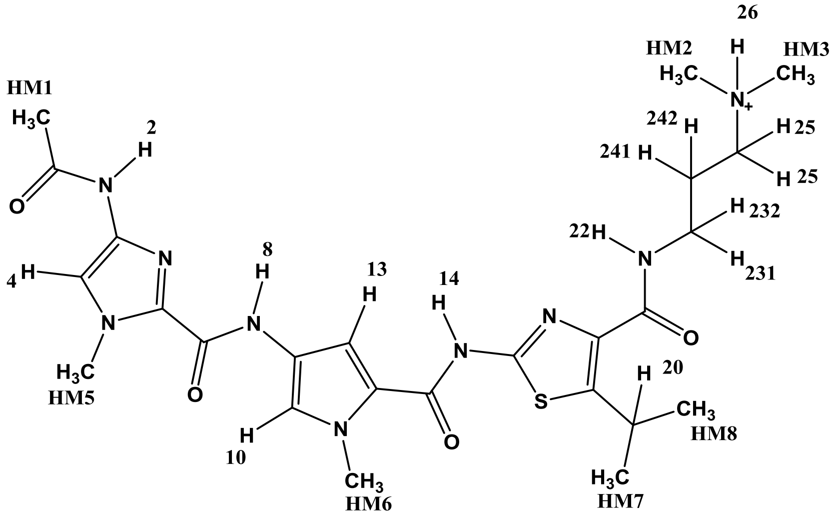


Figure 4.7 The numbering scheme for the ^{1}H-NMR assignment of thiazotropsin B.

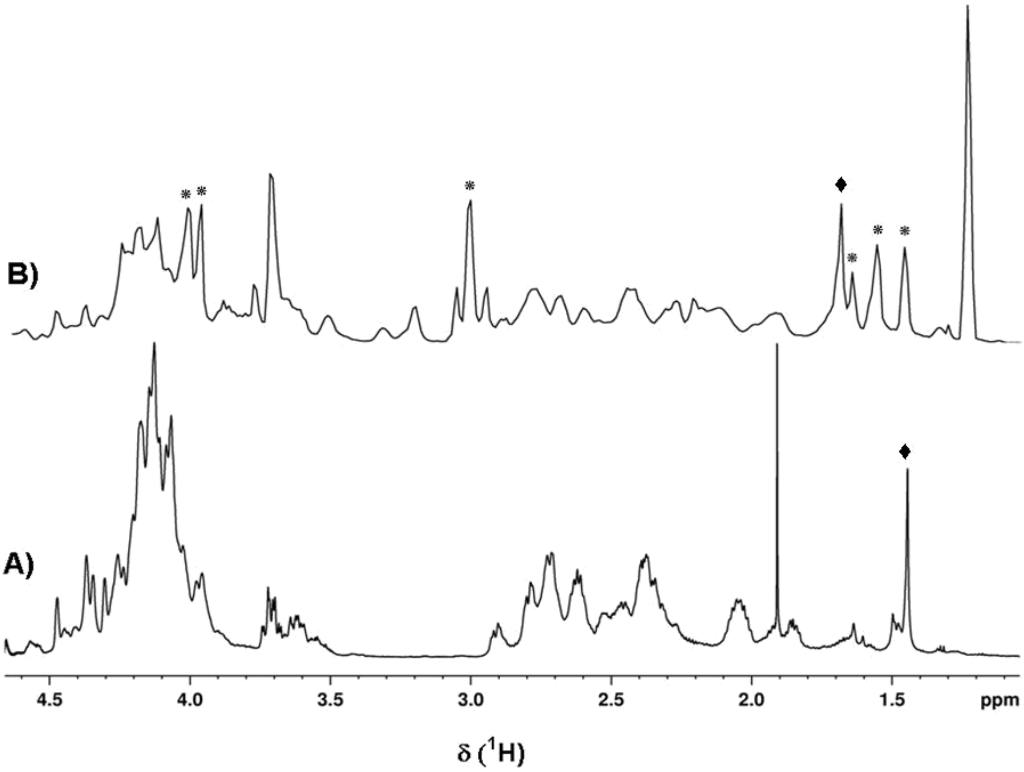


Figure 4.8 Comparison of the aliphatic region of the ^{1}H NMR spectrum of **A)** free DNA duplex d(CGACGCGTCG)$_2$ and **B)** the 2:1 complex between thiazotropsin B and d(CGACGCGTCG)$_2$ at 600 MHz using a dpfgse routine for solvent suppression. Methyl resonances from the ligand in the complex are indicated by *. Methyl resonances from thymine residues are indicated by ♦.

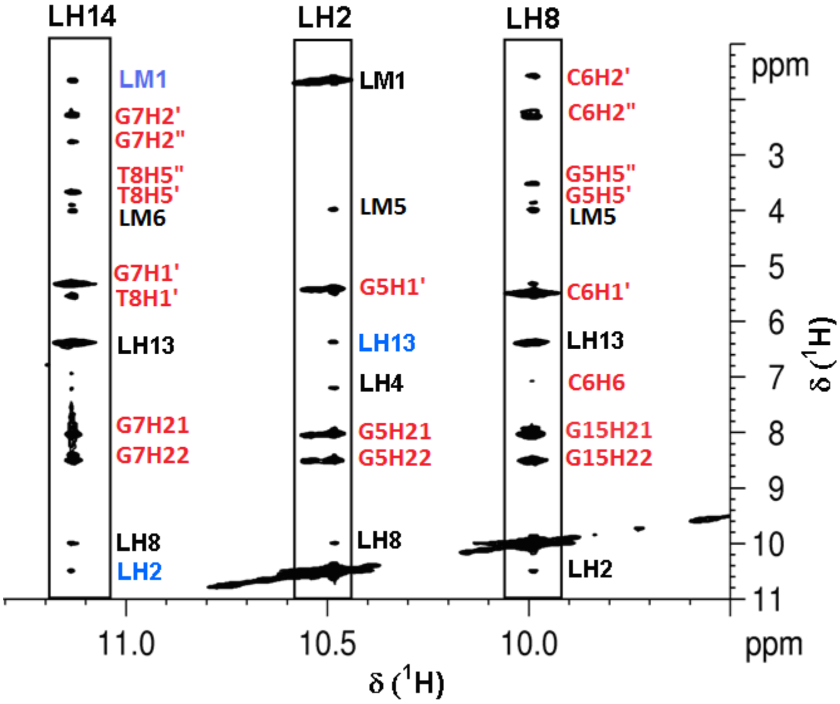


Figure 4.9 Strip plots of data taken from the 100 ms 2D [^{1}H, ^{1}H] NOESY NMR spectrum acquired on the 2:1 complex between thiazotropsin B and d(CGACGCGTCG)$_2$ at 600 MHz. Data are shown at the ligand resonance chemical shifts of H2, H8 and H14. Data labeling scheme: DNA resonance assignments, red labels; ligand resonance assignments, black labels; interligand NOEs, blue labels.

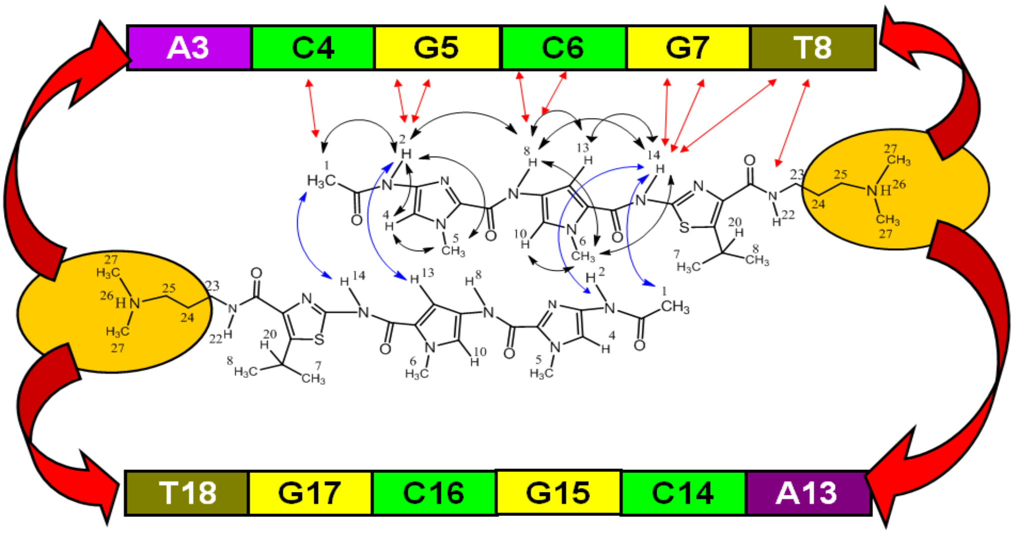


Figure 4.10 A schematic representation of how some of the NOEs relate to the structure of the complex. Ligand-DNA contacts, red arrows; intra-ligand contacts, black arrows; inter-ligand contacts, blue arrows.

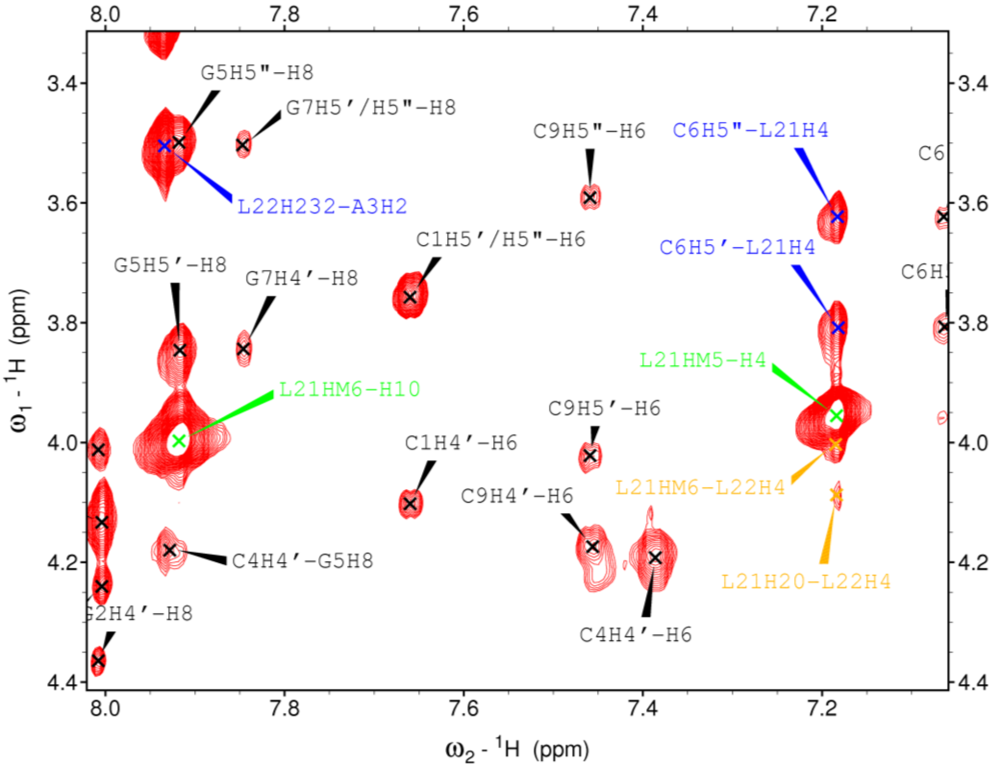


Figure 4.11 Part of the 2D [^{1}H, ^{1}H] NOESY NMR spectrum of the 2:1 complex of thiazotropsin B with d(CGACGCGTCG)$_2$ showing large intra-ligand NOE cross-peaks between the pyrrole methyl and their associated ring protons (green labels). Data labeling scheme: DNA resonance assignments, black labels; ,ligand-DNA resonance assignments, blue labels; intra-ligand NOEs, green labels; inter-ligand NOEs, orange labels.

Table 4.2 ^{1}H NMR chemical shift assignments of thiazotropsin B complexed with d(CGACGCGTCG)$_2$.

Chemical shift assignment: δ ^{1}H (ppm)									
HM1	H2	H4	HM5	H8	H10	HM6	H13	H14	H20
1.634	10.49	7.185	3.961	9.995	7.922	4.001	6.359	11.14	4.099
Chemical shift assignment: δ ^{1}H (ppm)									
HM7	HM8	H22	H231	H232	H241	H242	H25	H26	HM (2&3)
1.448	1.547	8.406	3.305	3.509	2.111	1.987	3.195	8.922	2.999

^{1}H NMR resonance assignments of the oligonucleotide to which thiazotropsin B was bound were made from the D_2O and H_2O 2D [^{1}H, ^{1}H] NOESY NMR spectra, using the same procedure as followed for d(CGACGCGTCG)$_2$ alone and by direct comparison with the original d(CGACGCGTCG)$_2$ ^{1}H NMR spectrum. Part of the NOESY NMR spectrum of the thiazotropsin B-DNA complex in D_2O is shown in Figure 4.12. All the assigned chemical shifts of the ligand bound d(CGACGCGTCG)$_2$ are listed in Table 4.3.

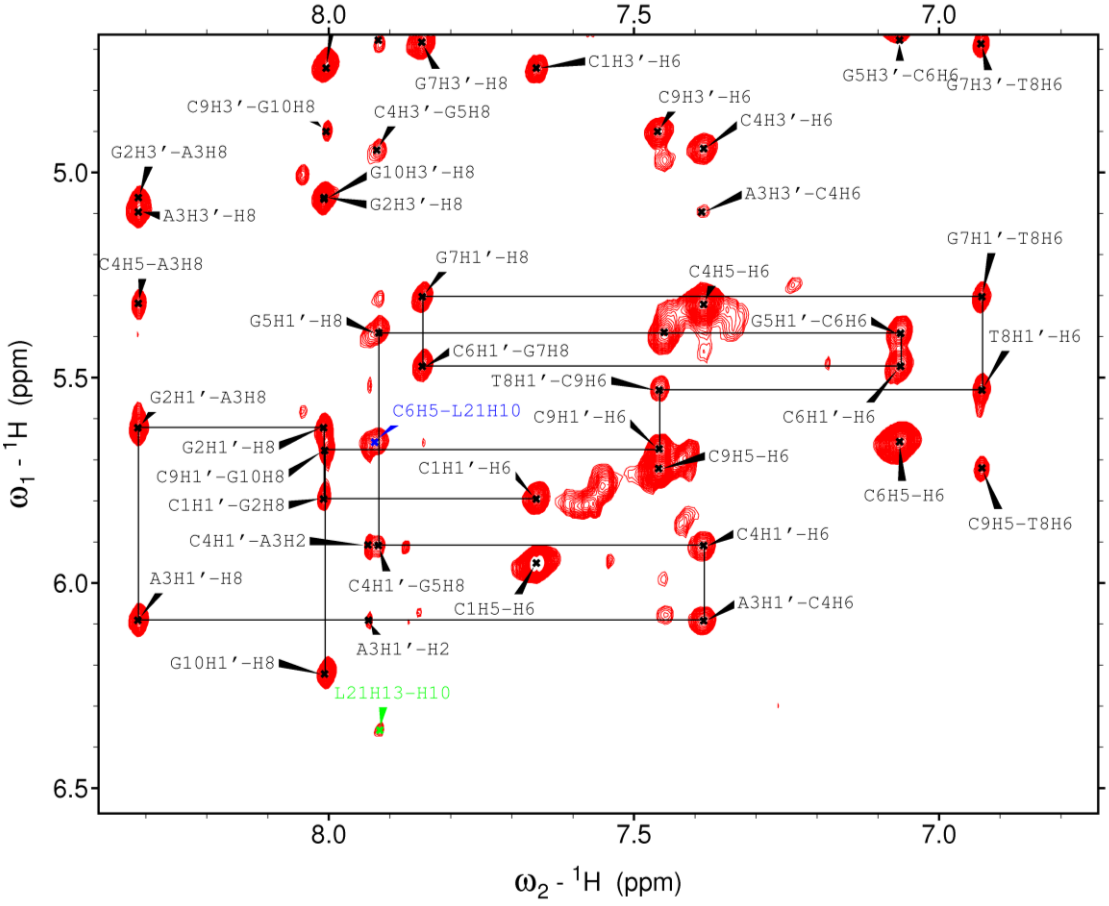


Figure 4.12 Fingerprint region of the 100 ms 2D [^{1}H, ^{1}H] NOESY NMR spectrum of d(CGACGCGTCG)$_2$, at 600 MHz in a ligand-duplex ratio of 2:1. Resonance assignments are shown and the assignment 'walk' indicated by horizontal and vertical lines, which join the NOEs between aromatic and sugar H1' protons.

Large chemical shift changes were expected for resonances of the sugar protons H1′, H4′, H5′, and H5′′ based on the assumption that thiazotropsin B would bind in the minor groove of DNA. The shielding influence of aromatic peptide *N*-methyl imidazole, *N*-methyl pyrrole, and isopropyl thiazole rings lodged in the DNA minor groove has a profound effect on the chemical shifts of the nearby sugar protons.[38] The differences in chemical shift ($\Delta\delta$ ^{1}H) between the free and ligand bound DNA are listed in Table 4.4.

Some of cross-peaks from H4′ protons, which usually resonate at chemical shifts > δ^1H = 3.9 ppm for unmodified ODN, appeared between δ^1H =2.1-2.3 ppm when the complex was formed (Table 4.3). This change in chemical shift upon binding was used to determine the location of the binding site. By plotting the difference ($\Delta\delta$) between the chemical shift of protons in the free ODN and the chemical shifts of the same protons in the ligand-bound ODN, it was possible to determine the location of the ligand against the DNA sequence. A plot of ($\Delta\delta$) for the H4′ resonances of d(CGACGCGTCG)$_2$ (Figure 4.13) showed that the aromatic rings of the ligand were lying in the minor groove juxtaposed against the sugar rings of the sequence 5′-ACGCGT-3′ and the ligand was arranged in a 2:1, head-to-tail, side-by-side manner within the minor groove of DNA with the retention of duplex symmetry.

The H4′ proton resonances showed the biggest change in their chemical shifts when the ligand was bound. H1′ proton resonances also showed significant shift changes (Figure 4.14) relative to the DNA sequence. The dashed line shows the shift changes for the opposing DNA strand.

Table 4.3 ^{1}H chemical shift assignments for DNA d(CGACGCGTCG)$_2$ in the presence of 2 equiv. of thiazotropsin B per duplex.

	Chemical shifts : δ ^{1}H (ppm)																
	H8	**H6**	**H2**	**H5**	**CH_3**	**H1′**	**H2′**	**H2″**	**H3′**	**H4′**	**H5′**	**H5″**	**GH1/ TH3**	**H41**	**H42**	**H21**	**H22**
Base																	
C1		7.660		5.951		5.795	1.926	2.424	4.745	4.103	3.755	3.755					
G2	8.008					5.624	2.769	2.879	5.063	4.366	4.012	4.012	12.95				
A3	8.312		7.937			6.093	2.684	2.794	5.094	4.465	4.231	4.231					
C4		7.387		5.322		5.909	1.893	2.087	4.944	4.187	4.190	4.190		7.725	6.119		
G5	7.920					5.390	2.458	2.604	4.677	2.247	3.847	3.501	13.20			8.022	8.497
C6		7.065		5.659		5.469	1.555	2.304	4.642	2.214	3.810	3.625		8.465	6.295		
G7	7.847					5.304	2.180	2.745	4.688	3.845	3.502	3.502	13.36			8.031	8.499
T8		6.931			1.678	5.528	1.721	1.900	4.579	2.262	3.886	3.652	13.90				
C9		7.459		5.726		5.676	2.165	2.393	4.899	4.174	4.021	3.591		8.808	6.962		
G10	8.006					6.223	2.439	2.675	5.064	4.239	4.136	4.136					

Table 4.4 ^{1}H chemical shift differences ($\Delta\delta$) defined as ($\Delta\delta$)= δ [(ligand-bound DNA)- δ(ligand free DNA)] for duplex d(CGACGCGTCG)$_2$.

	Chemical shift differences : $\Delta\delta$ ^{1}H (ppm)																
	H8	**H6**	**H2**	**H5**	**CH_3**	**H1′**	**H2′**	**H2″**	**H3′**	**H4′**	**H5′**	**H5″**	**GH1/ TH3**	**H41**	**H42**	**H21**	**H22**
Base																	
C1		0.025		0.010		0.032	0.033	0.040	0.034	0.025	0.027	0.027					
G2	0.028					0.168	0.027	0.068	0.048	0.044	0.029	0.029	-0.060				
A3	0.068		0.017			-0.182	-0.045	-0.134	0.014	-0.032	-0.011	0.055					
C4		0.184		0.085		0.306	-0.046	-0.252	-0.137	-0.421	0.027	0.027		-0.471	-0.390		
G5	0.081					-0.499	-0.160	-0.134	-0.306	**-2.120**	-0.286	**-0.550**	0.300			n/a	n/a
C6		-0.232		0.341		-0.222	**-0.520**	-0.111	-0.214	**-1.980**	-0.335	**-0.520**		0.141	-0.214		
G7	-0.061					**-0.701**	-0.475	-0.070	-0.290	**-0.892**	-0.304	**-0.586**	0.480				
T8		-0.355			0.220	**-0.530**	-0.372	**-0.575**	-0.301	**-1.960**	-0.255	-0.489	0.110				
C9		-0.062		-0.013		-0.055	0.116	-0.004	0.028	0.033	-0.060	-0.490		0.118	-0.115		
G10	0.026					0.024	-0.002	-0.091	0.356	0.035	0.051	0.051					

n/a- not applicable due to absence of protons in free DNA. Colour coding: Figures are colour coded blue or red for $|\Delta\delta|$ > 0.1 ppm: red-base protons; blue-sugar protons. Bold typeface entries are for the most significant changes where $|\Delta\delta|$ > 0.5 ppm

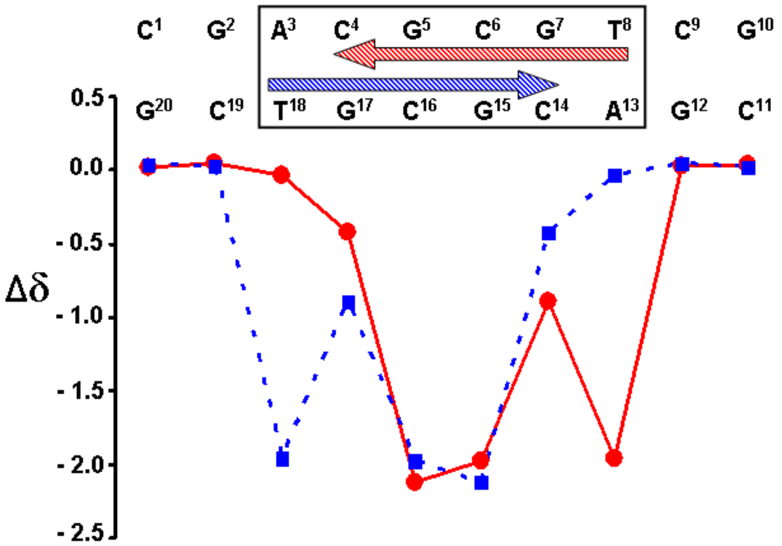


Figure 4.13 Chemical shift differences for H4' resonances of ligand-bound and ligand-free DNA duplex [bound-free]. Shaded arrows represent the location of the ligand relative to the DNA sequence. The dashed line shows the shift changes for the opposing DNA strand.

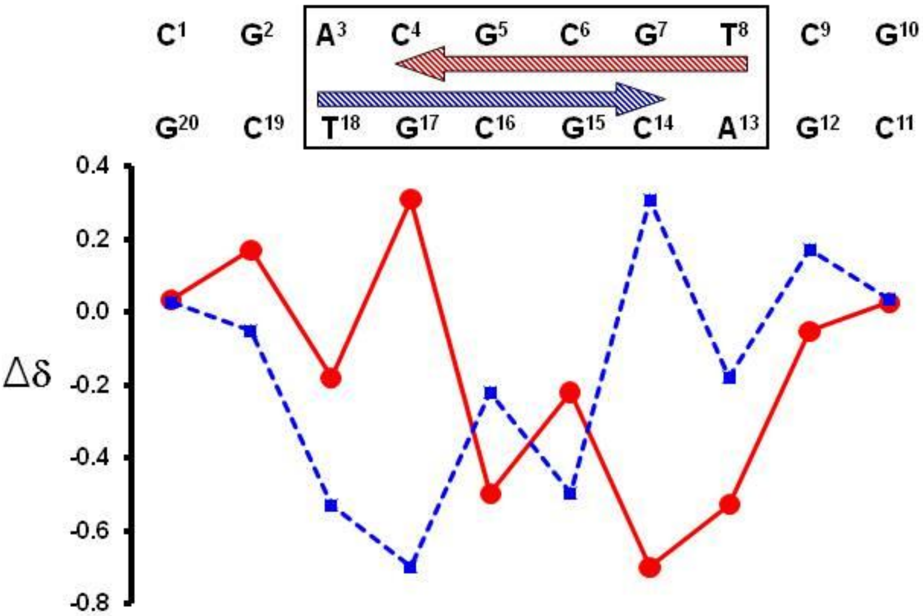


Figure 4.14 Chemical shift differences for H1′ resonances of ligand-bound and ligand-free DNA duplex [bound-free]. Shaded arrows represent the location of the ligand relative to the DNA sequence. The dashed line shows the shift changes for the opposing DNA strand.

Other changes in the chemical shifts (around $|\Delta\delta| = 0.5$ ppm) were observed with H5″ (G5, C6 & G7), H2′ (C6), and H2″ (T8). These changes were believed to be caused not only by the influence of nearby aromatic rings of the ligand (*via* van der Waals contacts) but also by changes in the DNA backbone conformation at the G5pC6pG7pT8 steps. The change in the chemical shift is attributed to an alteration in the geometry of the DNA backbone at these positions in the sequence. This was suspected from the changes in the ^{31}P NMR data of the DNA upon ligand binding (Figure 4.15). However, this change in the DNA backbone conformation was not as profound as that found in the previously determined structure of the thiazotropsin A/ACTAGT complex,[38] where an unusually large change in the chemical shift of A6H5″ (1.173 ppm) was observed. This may be explained by the fact that GC rich DNA sequences have a wider minor groove compared with AT rich sequences, and perturbations in the backbones are less pronounced. Resonance assignments for 2D [^{31}P, ^{1}H] COSY NMR data are listed for both free and ligand bound DNA in Table 4.5.

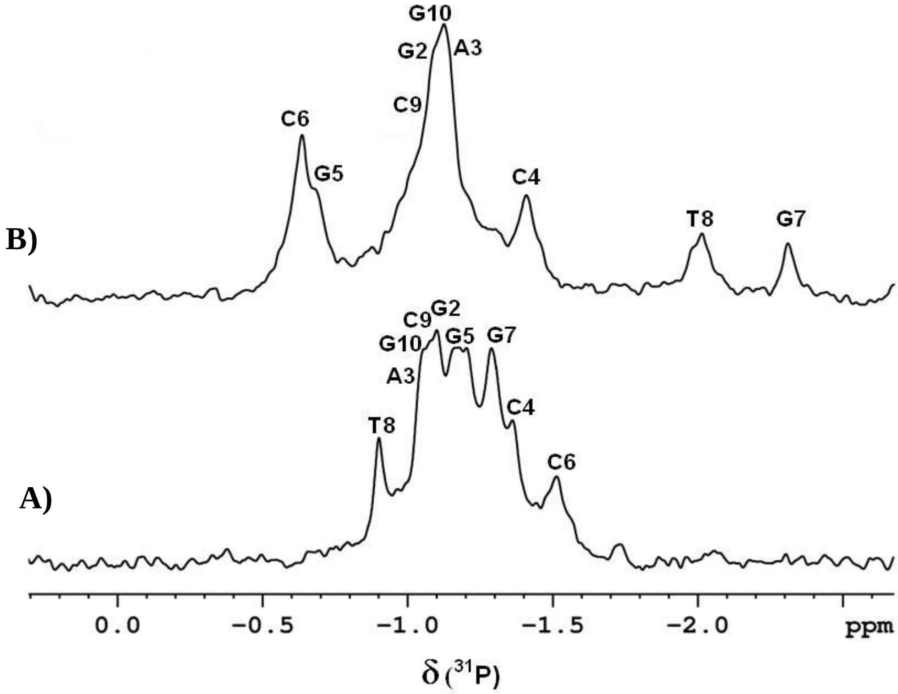


Figure 4.15 ^{31}P-{^{1}H} NMR spectra of ODN1 (A) and of the complex between d(CGACGCGTCG)$_2$ and thiazotropsin B (B) acquired at 9.4 T. The effect of thiazotropsin B binding to the DNA duplex was clear from the dispersion of signals that occurred for the complex, indicative of DNA backbone alteration.

Table 4.5 Comparison of ^{31}P chemical shift assignments for d(CGACGCGTCG)$_2$ in the absence (free) and presence (bound) of 2 equiv of ligand per duplex.

	Chemical shift assignment[a]: δ ^{31}P (ppm)		**Δδ ^{31}P(δ $_{bound}$ - δ$_{free}$) (ppm)**
Bases	**Bound**	**Free**	
C1	-	-	-
G2	-1.06	-1.10	0.04
A3	-1.12	-1.03	-0.09
C4	-1.41	-1.37	-0.04
G5	-0.68	-1.15	0.47
C6	-0.61	-1.51	0.90
G7	-2.31	-1.28	-1.03
T8	-2.02	-0.89	-1.13
C9	-1.11	-1.08	-0.03
G10	-1.11	-1.06	-0.05

[a] Assignments are for the phosphates that are 5′ with respect to the base.

4.3 Molecular modeling

Proton-proton NOEs assigned from 2D [^{1}H, ^{1}H] NOESY NMR data acquired for d(CGACGCGTCG)$_2$ and for the 2:1 complex between thiazotropsin B and d(CGACGCGTCG)$_2$ were used for model building. Intermolecular ligand-DNA and ligand-ligand NOE contacts unambiguously confirm the formation of the 2:1 thiazotropsin B-d(CGACGCGTCG)$_2$ complex at a six base pair binding site (5′-ACGCGT-3′) in the minor groove of DNA. The NOEs of the complex were consistent with the previously determined structure of the thiazotropsin A-ACTAGT complex, in which the ligand is arranged in a 2:1, head-to-tail, side-by-side manner within the minor groove of DNA.

4.3.1 Solution structure of the free d(CGACGCGTCG)$_2$

The solution structure of d(CGACGCGTCG)$_2$ was determined in order to assess the effect of ligand binding on the structure of the DNA duplex. The Watson-Crick base pairing of the free duplex was intact as indicated by the presence of imino proton ^{1}H NMR resonances in the NMR data. The starting model of the free d(CGACGCGTCG)$_2$ was constructed in the canonical B-form as the NMR data were more consistent with the right-handed B-form of a DNA duplex than other DNA structures. A total of 207 NOEs were incorporated into the starting structure as distance restraints with a force constant of 1 kcal mol^{-1} Å^{-2}. Representations of the resulting calculated average structure are shown in Figure 4.16. The calculated structure of the free d(CGACGCGTCG)$_2$ was analysed using CURVES.[134] Figure 4.17 illustrates the pictorial definitions of the parameters that relate the DNA bases. The overall nature of d(CGACGCGTCG)$_2$ did not deviate excessively from that of canonical B-form DNA. The average minor groove width increased from 5.9 Å in canonical B-DNA to 7.9 Å in the calculated structure, which may reflect the higher G/C base pair content of the central region of the free DNA. G/C-bearing regions of DNA generally have wider minor grooves with smaller negative electrostatic potentials compared with A/T bearing regions. Moreover, the structure of nucleic acids is hugely affected by the salt concentration in solution, and the observed wide minor groove of the free duplex may be induced by the buffer salts (phosphate buffer) used to stabilize the duplex.

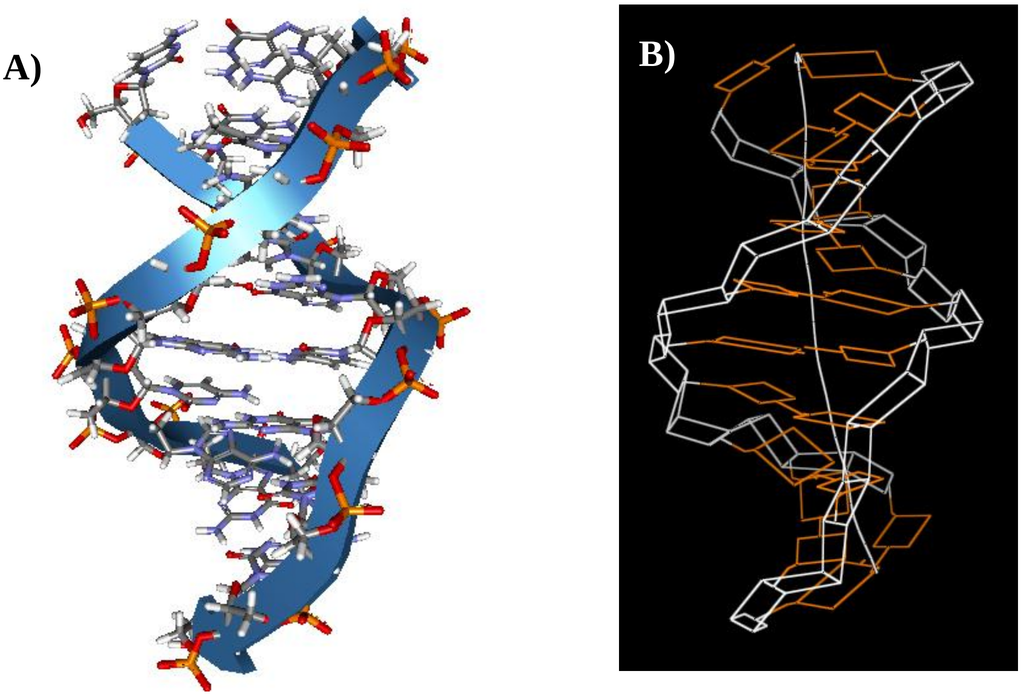
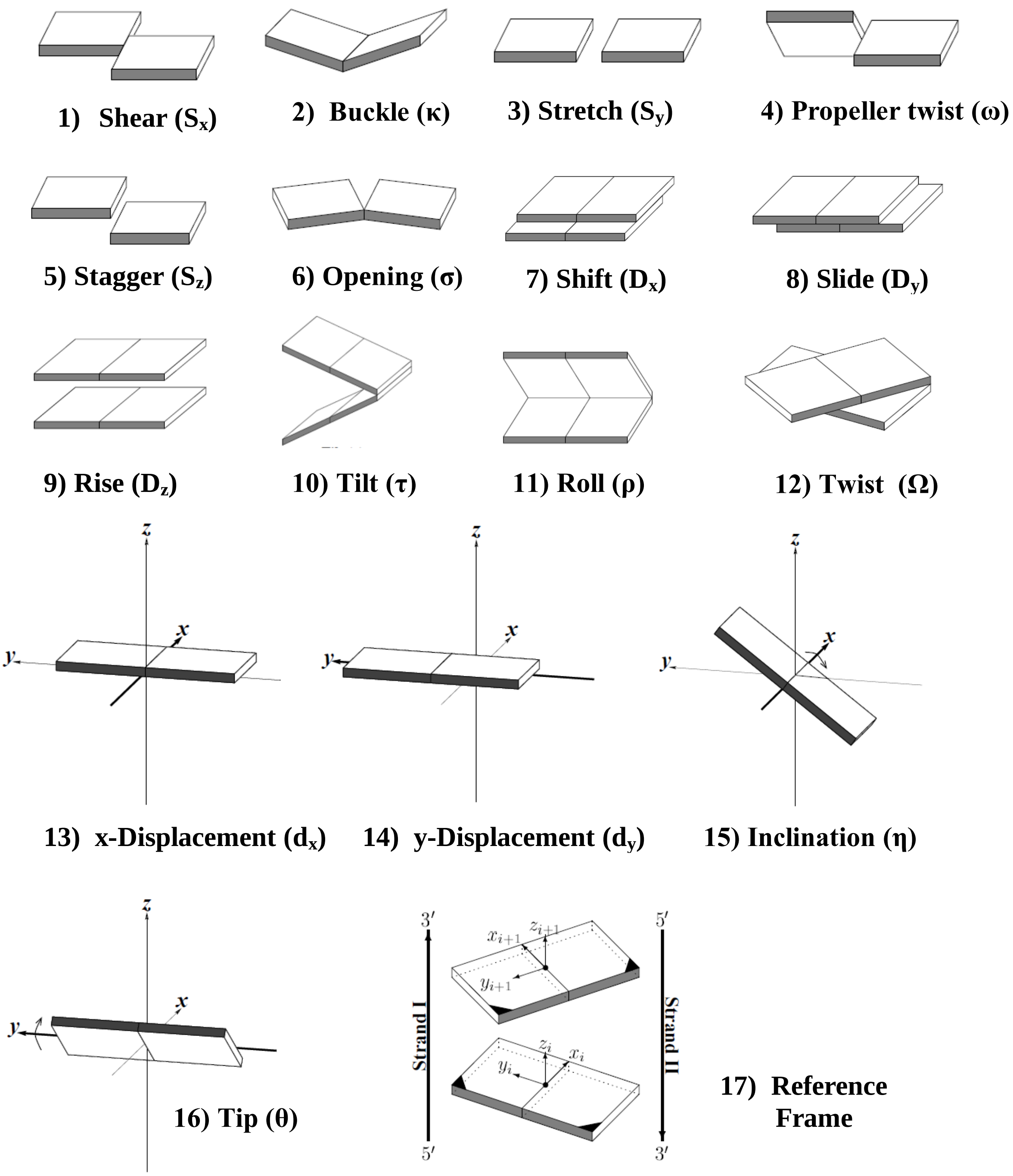


Figure 4.16 Representation of the solution structure of d(CGACGCGTCG)$_2$ alone based on restrained molecular dynamics simulations **A)** Stick vs. arrows representation of the average structure of d(CGACGCGTCG)$_2$ taken from 100 ps of restrained molecular dynamics simulations **B)** CURVES cartoon representation of the average structure of d(CGACGCGTCG)$_2$.

The structure displayed an overall axis curvature of 17.5°. Tip angle (θ), x- and y-displacement of base pairs was negligible (average θ = -1.05°, average dx = -1.78 Å, average dy = 0.25 Å) while inclination of the base pairs was significant (average η =12.9°). Average values of base pair shear (S_x), stretch (S_y), stagger (S_z) and buckle (κ) were similar to those of canonical B-DNA. Some variations in the propeller twist (ω) and opening (σ) of base pairs were noted (average ω =3.98, average σ= -3.16). Average global inter-base pair parameters shift (Dx), slide (Dy), rise (Dz), tilt (τ), roll (ρ) and twist (Ω) were similar to those of canonical B-DNA. Sugar puckers fell into either C1′-exo or C2′-endo configurations.



1) Shear (S_x)

2) Buckle (κ)

3) Stretch (S_y)

4) Propeller twist (ω)

5) Stagger (S_z)

6) Opening (σ)

7) Shift (D_x)

8) Slide (D_y)

9) Rise (D_z)

10) Tilt (τ)

11) Roll (ρ)

12) Twist (Ω)

13) x-Displacement (d_x)

14) y-Displacement (d_y)

15) Inclination (η)

16) Tip (θ)

17) Reference Frame

Figure 4.17 Pictorial definitions of parameters that relate complementary base pairs **(1-6)**, sequential base-pair steps **(7-12)** and base pair to its helical frame **(13-16)**. The base pair reference frame **(17)** is constructed such that the *x*-axis points away from the (shaded) minor groove edge.[2]

4.3.2 Solution structure of the 2:1 thiazotropsin B/ d(CGACGCG TCG)$_2$ complex.

A total of 740 intermolecular ligand-DNA and ligand-ligand and intramolecular DNA-DNA and ligand-ligand distance restraints derived from NOE data were used to obtain the energy minimised model of the d(CGACGCGTCG)$_2$ -thiazotropsin B complex (See Appendix 1 for full details of NOE assignments). The two ligands fit snugly into the minor groove, stacked next to each other with the positively charged DMAP tails pointing in opposite directions.

As noted previously, 1D ^{1}H NMR data revealed that complex formation between thiazotropsin B and d(CGACGCGTCG)$_2$ was symmetrical. This occurs only if the ligand is arranged in an anti-parallel side-by side mode within the minor groove of a DNA duplex. The NOE data also confirmed this mode of binding via the observed NOEs between the ″tail″ of one ligand and the ″head″ of the partnering ligand and *vice versa*. The change in the chemical shift of the DNA proton resonances firmly placed thiazotropsin B in the minor groove of the duplex as expected and ligand/DNA NOEs supported this finding. A representation of the calculated solution structure of the complex is shown in Figure 4.18 A. The structure shown represents the average structure of the complex taken from the final 100 ps of 300 ps of restrained molecular dynamic simulation in explicit solvent.

The hydrogen- bonding scheme for the complex was deduced indirectly on the basis of the calculated solution structure (Figure 4.19 and Table 4.6). The amide protons and the nitrogen of imidazole and thiazole rings of thiazotropsin B face the floor of the minor groove and form hydrogen bonds with the nearby DNA bases. Hydrogen bonds were assigned between the thiazotropsin B amide protons NH-2, NH-8, NH-14, and NH-22 with G^5 N3, C^6 O2, G^7 N3, and T^8 O2, respectively. The imidazole and thiazole nitrogens of thiazotropsin B, which are responsible for the specificity of the complex, formed hydrogen bonds with one amino proton each of G5 and G7, respectively.

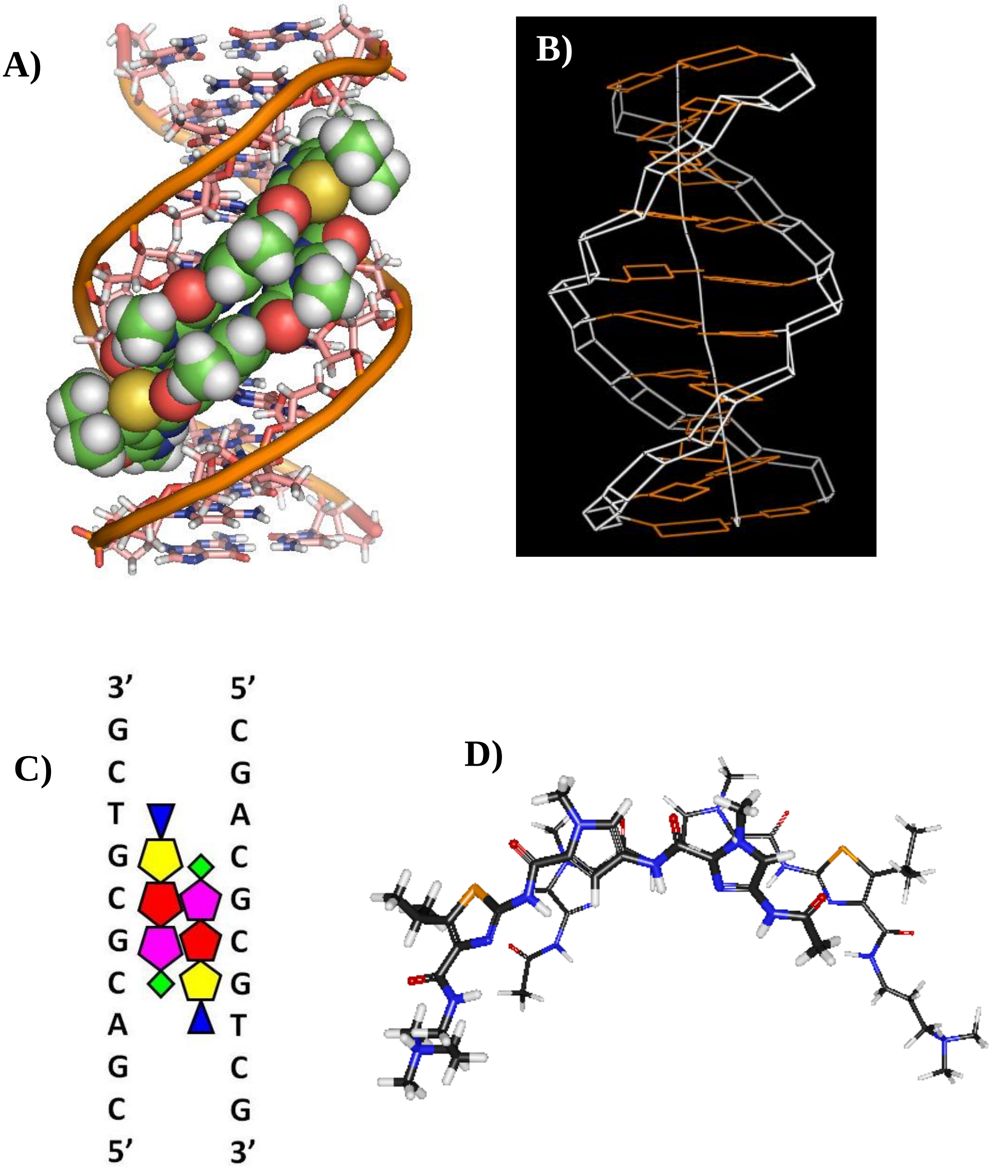


Figure 4.18 Cartoon and schematic representation of the complex between thiazotropsin B and d(CGACGCGTCG)$_2$ showing the location of ligand with respect to the DNA sequence **(A)** Representation of the refined solution structure of the complex between thiazotropsin B (CPK drawing) and d(CGACGCGTCG)$_2$ (stick and tubes) **B)** CURVES cartoon representation of the average structure of the ligand-bound d(CGACGCGTCG)$_2$ **C)** Schematic indicating the ligand alignment relative to the DNA sequence. Colour coding: green diamond = formyl "head"; magenta pentagon =*N*-methylimidazole; red pentagon =*N*-methylpyrrole; yellow pentagon = isopropylthiazole; blue triangle = DMAP "tail" **D)** Relationship between associated ligands in the complex.Thicker lines are shown for one ligand compared with its partner.

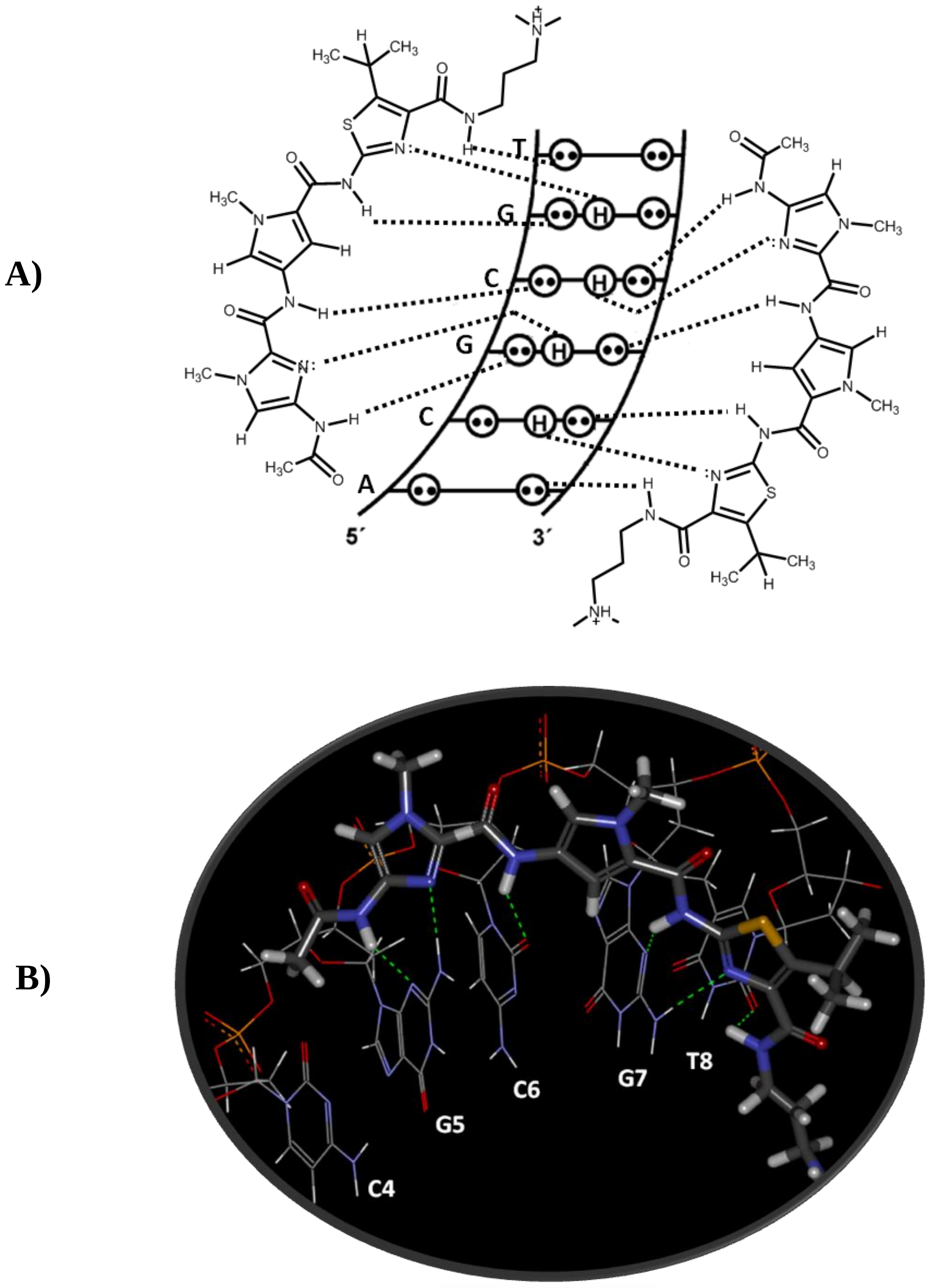


Figure 4.19 A) Deduced arrangement of hydrogen bonding between thiazotropsin B and d(CGACGCGTCG)$_2$. **B)** Part of the calculated average structure showing the hydrogen bonds (green dashed lines) formed between thiazotropsin B (thick lines) and one strand of the DNA duplex. Hydrogen bonds were assigned using the Discovery Studio program.

Table 4.6 Summary of the hydrogen bonds in the 2:1 complex formed between thiazotropsin B and d(CGACGCGTCG)$_2$ based on the labile proton exchange characteristics and solution structure information.

Ligand atom	**DNA atom**	**Distance (Å)**
NH-2,	G5 N3	1.9
NH-8,	C6 O2	2.1
NH-14,	G7 N3	2.4
NH-22	T8 O2	2.3
Imidazole N	G5 H22	1.9
Thiazole N	G7 H22	2.1

The labile protons of the ligand (amide protons) which form hydrogen bonds with DNA bases appeared to be inaccessible to solvent (See Figure 4.5). 2D [^{1}H, ^{1}H] NOESY NMR data acquired for the complex did not show cross-peaks between the ligands′ peptide NH protons and the solvent resonance. The absence of an exchange correlation between the solvent resonance and that of the labile protons indicates that these protons are solvent inaccessible. This occurs as a consequence of the complex formation which makes the ligand-DNA binding interface less accessible to solvent when the ligand occupies the minor groove of DNA.

CURVES analysis of the ligand bound DNA duplex did not show significant change in its conformation upon ligand binding compared with the calculated structure of the free duplex. This could be related to the structure of the free duplex which inherently has a wide minor groove (7.9 Å) sufficient to accommodate the ligand without excessively disrupting the DNA structure. The overall curvature of the ligand bound DNA duplex was 17.8°, in close agreement with the value for the calculated ligand-free DNA structure. The global shape of the DNA structure in the complex agreed well with that of the free DNA. Average global base pair axes parameters generally showed minor deviations from the values calculated for canonical B-DNA- (dx = -0.97 Å, dy = 0.16 Å, tip (θ) =0.72°) although inclination (η) showed a higher average value (η = 5.1°). Remarkably, the central G/C base pairs showed significant inclination (10°) compared with the other base pairs. Global base-base and inter-base pair parameters were in close agreement with the value for the calculated ligand-free DNA structure. As described earlier, such relatively minor perturbations to the free DNA structure upon binding may explain the absence of an entropic penalty noted when binding was measured by ITC. Some changes were observed in the width and depth of minor groove upon ligand binding. Analysis of the complex structure generated average DNA minor groove parameters of 8.4 Å wide and 4.9 Å deep in the-ACGCGT- region which represent 0.5 and 1.3 Å increase in the width and depth of the minor groove of ligand bound duplex, respectively. This change in the minor groove dimensions is caused by the presence of side-by-side binding of thiazotropsin B within the minor groove of the DNA duplex. The acetyl head group was orientated toward the 5′-end of the 5′- ACGCGT-3′ segment of DNA with the positively charged DMAP tail located toward the 3′ end. The dimensions of the minor groove for both canonical B-DNA and the NMR solution structures calculated for the ligand-free and ligand-bound DNA structures are listed in Table 4.7. Although NMR data poorly define the dimensional features of DNA duplex such as curvature and groove width, these features can be obtained from modelling methods used in the determination of the three-dimensional structure of nucleic acids. Examination of these features can provide insight into the changes of the DNA structure upon ligand binding.

Table 4.7 Comparison of key global parameters for three DNA structures: canonical B-DNA, the calculated ligand-free B-DNA structure, and the calculated DNA structure in the presence of two molar equiv of ligand bound in the DNA minor groove.

	Minor groove[1]		
	width (Å)	depth (Å)	overall axis bend (deg)
canonical B-DNA[2]	5.9	4.6	0
ligand-Free DNA[3]	7.9	3.5	17.5
DNA with 2 equiv of thzB[3]	8.4	4.8	17.8

[1] Measured for the region -ACGCGT-. [2] Software-generated structure. [3] NMR structure

4.4 NMR study of the AIK18-51-d(CGACTAGTCG)$_2$ complex

4.4.1 Titration of d(CGACTAGTCG)$_2$ with AIK18-51

The oligonucleotide d(CG**ACTAGT**CG)$_2$ was titrated with AIK18-51 as described in Chapter 2. 1D ^{1}H NMR spectra were acquired on the sample at various ligand/DNA ratios to allow titrations to be monitored and complex formation to be rapidly assessed. As the titration continued, the amount of free DNA decreased continuously, while the amount of 2:1 complex (ligand to DNA) increased. The resonances of the free ODN completely disappeared at a 2:1 ratio of ligand to DNA. There was no evidence of other complexes (such as 1:1 complex) being formed throughout the titration. In the aliphatic proton resonance region of the free DNA the ^{1}H NMR spectrum showed the presence of two methyl proton resonances at δ^1H = 1.262 and δ^1H = 1.593 ppm for T^8 and T^5, respectively (Figure 4.20 A). This is expected for a symmetrical 10 base pair sequence containing four thymine bases. These resonances were replaced as AIK18-51 was titrated against d(CGACTAGTCG)$_2$ by the simultaneous appearance of new thymine methyl signals (δ^1H = 1.566 and δ^1H= 1.725 ppm) and another two resonances from AIK18-51. The ligand resonances were anticipated to arise from methyl protons of the ligand isopropyl group (HM3 and HM4, see Figure 4.21 for the numbering scheme of AIK18-51). These resonances appeared as doublets because both HM3 and HM4 couple with H25 (J=7.2). Titration of d(CGACTAGTCG)$_2$ with AIK18-51 was associated with slow chemical exchange on the NMR timescale. All the 1D ^{1}H NMR spectra acquired at a ratio less than 2:1 ligand to DNA showed proton resonances from both free and ligand-bound DNA. Replacement of the free DNA signals with another set of signals for the ligand bound DNA without any increase in spectral complexity provided firm evidence that two ligand molecules bind to the self-complementary ODN in an anti-parallel, side-by-side fashion with retention of duplex symmetry.

ITC analysis of AIK18-51 showed the presence of two apparent phases indicative of two distinct binding events; an initial exothermic specific minor groove binding event, and a second endothermic binding event before saturation of the first site.

NMR titration of the AIK-18-51 with d(CGACTGTCG)$_2$ showed only the presence of the dimeric ligand binding into the minor grooves with no evidence of other complexes being formed. This suggests that the second binding event does not recognise a defined binding site as the chemical shifts of DNA protons were not affected and supports our conclusion that the second binding event is due to a non-specific backbone binding.

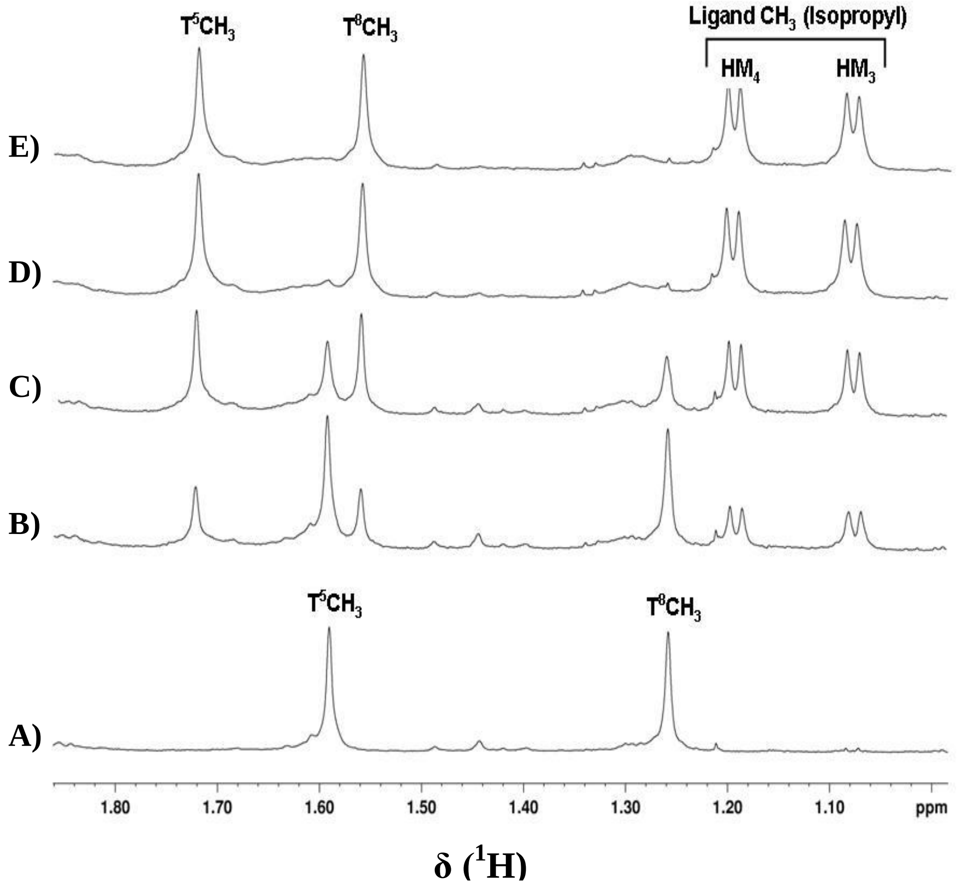


Figure 4.20 Aliphatic region of the 1D ^{1}H NMR data acquired at 600 MHz using a one-dimensional noesypresat routine for solvent suppression and showing the result of titrating AIK18-51 into a sample of d(CGACTAGTCG)$_2$. The ^{1}H NMR resonance of the T^{5}CH$_3$ and T^{8}CH$_3$ groups are visible at δ^{1}H = 1.593 and 1.262 ppm in the free ODN and at 1.725 and 1.566 ppm in the ligand:ODN complex respectively. Resonances at δ^{1}H =1.082 and 1.197 ppm are assigned to CH$_3$ protons HM3 and HM4 of AIK18-51. **A)** Free DNA**; B)** with 0.6 equiv.AIK18-51; **C)** with 1.2 equiv. AIK18-51; **D**) with 1.8 equiv. AIK18-51; **E)** with 2.4 equiv.AIK18-51.

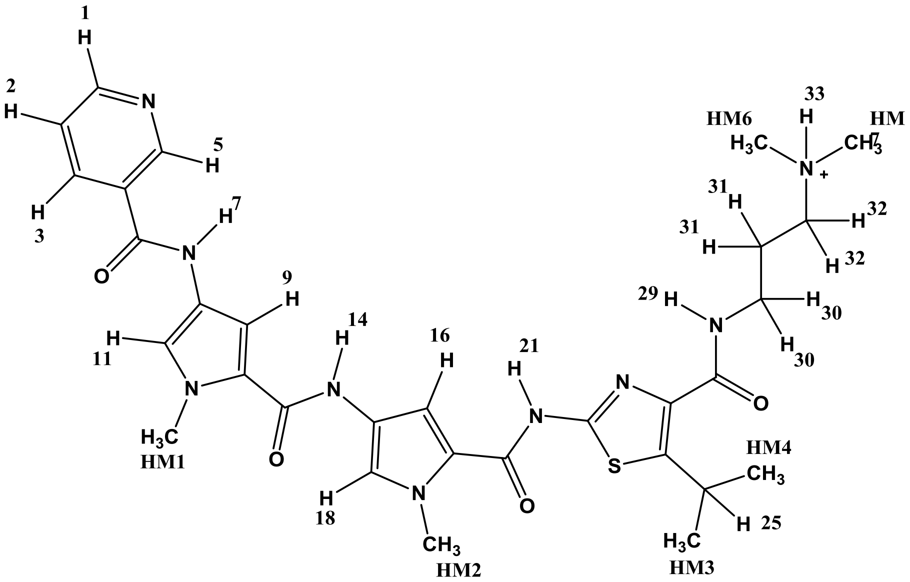


Figure 4.21 The numbering scheme for the ^{1}H-NMR assignment of AIK18-51. Equivalent protons that resonate at the same chemical shift were given the same atom number.

4.4.2 NMR data assignment

The NMR assignments of both the free and ligand-bound d(CG**ACTA GT**CG)$_2$ were carried out according to standard methods described previously. The appearance of imino proton ^{1}H NMR resonances in the ^{1}HNMR data of the free and ligand-bound d(CGACTAGTCG)$_2$ indicated that the Watson-Crick base pairings were intact (Figure 4.22). The ^{1}H NMR spectrum also showed three additional resonances from the ligand (Figure 4.22 B). These resonances were expected to arise from N*H* protons associated with the ligand amide linkages. The assignments of imino proton ^{1}H NMR resonances were initially made based on the saturation transfer effects (Figure 4.23), which were observed following a presaturation approach to solvent suppression. This provided an initial clue to the assignment of these signals. A detailed NOE assignment was then made and used to confirm these speculative assignments. The exchangeable H1 guanine imino protons were identified from inter-strand NOE contacts with cytosine amino groups and intra-strand NOE with its own amino group. H3 thymine imino protons were identified from inter-strand NOE contacts with adenine H2.

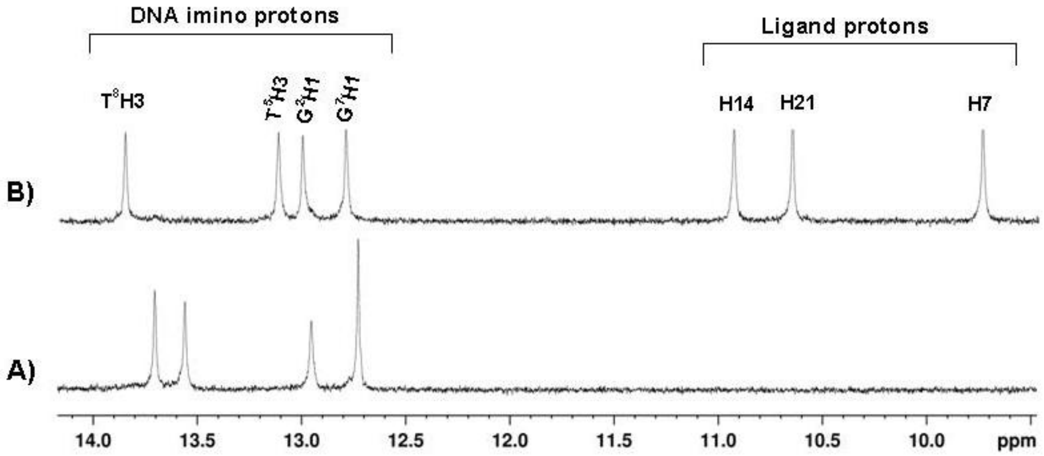


Figure 4.22 1D ^{1}H NMR resonances associated with the DNA imino protons (data acquired at 600 MHz using a dpfgse routine) showing the changes in chemical shifts of DNA imino protons and the appearance of ligand amide *NH* protons upon the addition of 2 equivalents of AIK18-51 to a sample of d(CGACTAGTCG)$_2$. **A)** Free DNA. **B)** Complex with AIK18-51. The ^{1}H NMR resonances of the imino protons belonging to Watson-Crick base pairs were visible between 12.5 and 14.0 ppm. Each one of these resonances represents two equivalent protons of the self complementary ODN (e.g. the equivalent G^{7}H1 and G^{17}H1 protons have the same resonance at 12.79 ppm). Resonances between 9.5 and 12.0 ppm were assigned to amide N*H* protons in AIK 18-51.

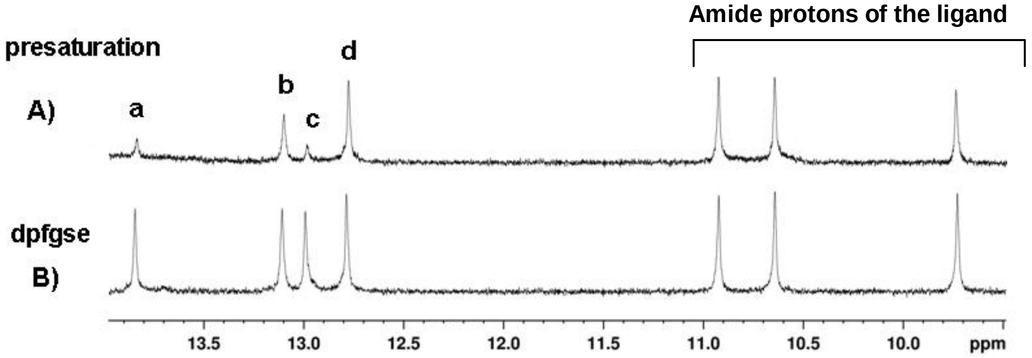


Figure 4.23 Imino proton resonance region of the ^{1}H NMR spectrum of the 2:1 complex AIK18-51 and d(CGACTAGTCG)$_2$ **A)** Using solvent presaturation; and **B)** using a dpfgse routine for solvent suppression. Saturation transfer effects (shown by the loss of intensity for signals **a** and **c** in **A)** enabled imino ^{1}H NMR resonance assignment to be made under the assumption that a greater degree of chemical exchange occurs with the solvent for protons nearest each end of the DNA duplex. **a-** T^{8}H3; **b-** T^{5}H3; **c-**G^{2}H1; **d-**G^{7}H1.

Part of a 2D [^{1}H, ^{1}H] NOESY NMR spectrum of the 2:1 complex between AIK18-51 and d(CGACTAGTCG)$_2$ is shown in Figure 4.24. All the assigned chemical shifts of the ligand bound d(CGACTAGTCG)$_2$ are listed in Table 4.8. NOESY cross-peak assignments were made by both comparisons with the previously determined complex between thiazotropsin A and d(CGACTAGTCG)$_2$ and through detailed analysis of 2D homonuclear [^{1}H, ^{1}H] correlation data. This involves identifying key ligand resonances via assessment of 2D [^{1}H, ^{1}H] DQFCOSY and TOCSY NMR data. Comparisons with 2D [^{1}H, ^{1}H] NOESY NMR data confirm ligand resonance identities. Peak shape and linewidths are considered when determining resonance assignments, particularly in the case of *NH* signals of the ligand.

Methyl group ^{1}H NMR resonances from AIK18-51 were visible in three regions of the ^{1}H NMR spectrum (Figure 4.25): 3.872−3.971 ppm (*N*-methyl pyrrole), 3.029 ppm (*N*-dimethyl), and 1.082−1.197 ppm (isopropyl-CH_3). Isopropyl methyl resonances, assigned from the 2D [^{1}H, ^{1}H] DQFCOSY NMR spectra, show strong correlations between methyl resonances at δ^1H = 1.082 (HM$_3$) and 1.197 ppm (HM$_4$)) and a multiplet at δ^1H = 3.583ppm, corresponding to H25 of AIK18-51 (Figure 4.26). ^{1}H NMR proton resonances associated with pyrrole *N*-methyl groups were assigned through strong NOEs to their associated ring protons (Figure 4.27) Combined use of 2D [^{1}H, ^{1}H] TOCSY and DQFCOSY NMR data enables assignment of the resonances associated with the ligand DMAP tail to be made.

The resonances of the aromatic protons H1, H2, H3, H5, H9, H11, H16, and H18 were assigned based on inspection of the relative sizes of NOEs associated with these resonances, together with NOE assignments of methyl and amide resonances from the ligands in addition to the observed DQFCOSY and TOCSY cross-peaks.

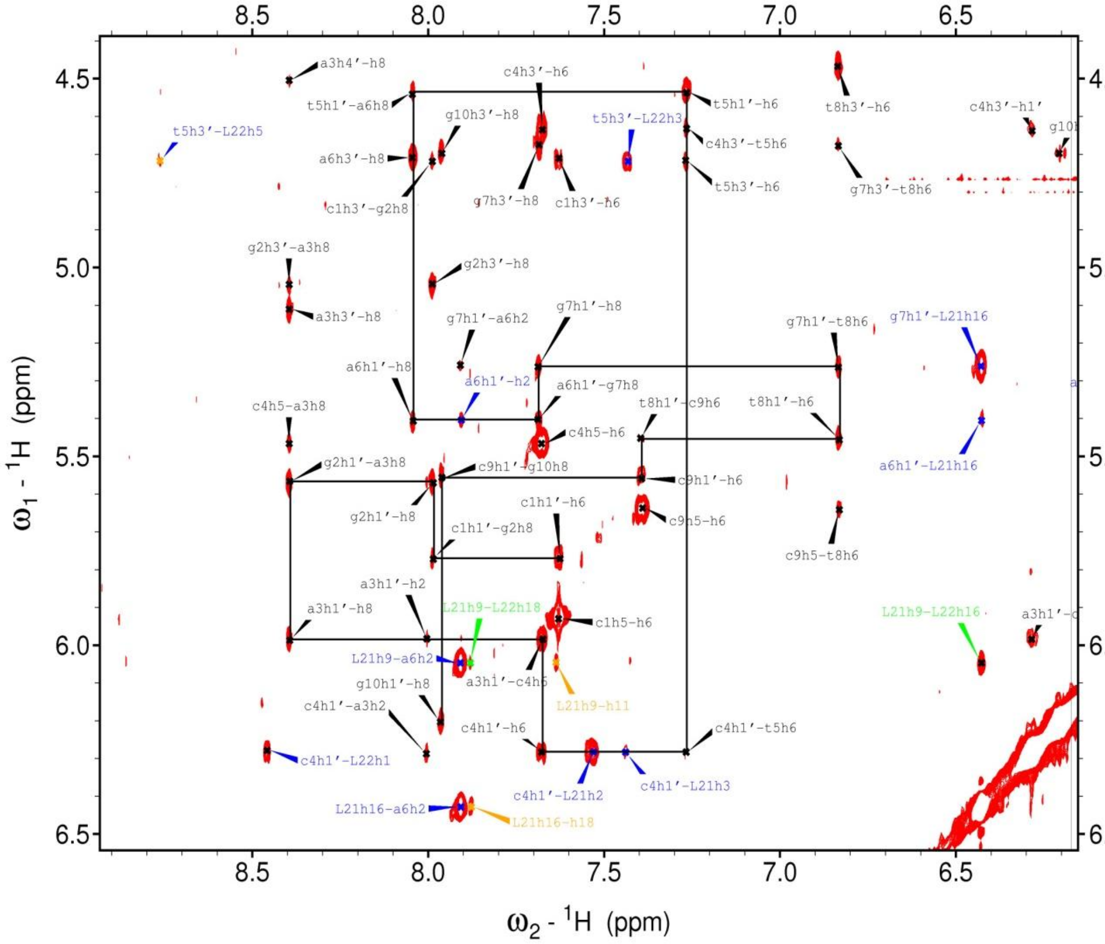


Figure 4.24 Fingerprint region of the 100 ms 2D [^{1}H, ^{1}H] NOESY NMR spectrum of the complex between d(CGACTAGTCG)$_2$ and AIK1851 acquired at 600 MHz in a ligand:DNA duplex ratio of 2:1. The assignment pathway between the 5′ and 3′ ends of the molecule is indicated by a continuous trace for H1′ -aromatic H6/H8 NOEs. Ligand-DNA NOE assignments are indicated in blue; intra-ligand NOE assignments are indicated in orange; inter-ligand NOE assignments are indicated in green intra-strand DNA NOE assignments are indicated in black.

Table 4.8 ^{1}H NMR chemical shift assignments for d(CGACTAGTCG)$_2$ in the presence of 2 equiv of AIK18-51 per duplex.

	Chemical shifts : δ ^{1}H (ppm)																
	H8	**H6**	**H2**	**H5**	**CH_3**	**H1′**	**H2′**	**H2″**	**H3′**	**H4′**	**H5′**	**H5″**	**GH1/ TH3**	**H41**	**H42**	**H21**	**H22**
Base																	
C1		7.63		5.927		5.772	1.887	2.386	4.724	4.075	3.275	3.25					
G2	7.987					5.567	2.74	2.816	5.047	4.34	4.104	3.984	13.001				
A3	8.395		8.005			5.983	2.617	2.927	5.111	4.505	4.26	4.178					
C4		7.677		5.466		6.283	1.868	2.249	4.635	4.268	3.319	3.853		7.498	6.5		
T5		7.268			1.725	4.535	1.902	2.198	4.721	2.771	3.711	3.558	13.12				
A6	8.045		7.906			5.406	2.257	2.752	4.707	2.692	3.848	3.259					
G7	7.686					5.26	2.105	2.612	4.678	2.58	3.882	3.552	12.79			8.515	6.172
T8		6.834			1.566	5.452	1.63	1.716	4.469	2.166	3.661	3.529	13.85				
C9		7.394		5.638		5.553	2.12	2.308		4.014				8.712	6.883		
G10	7.964					6.203	2.64	2.41	4.699	4.216	4.069	4.069					

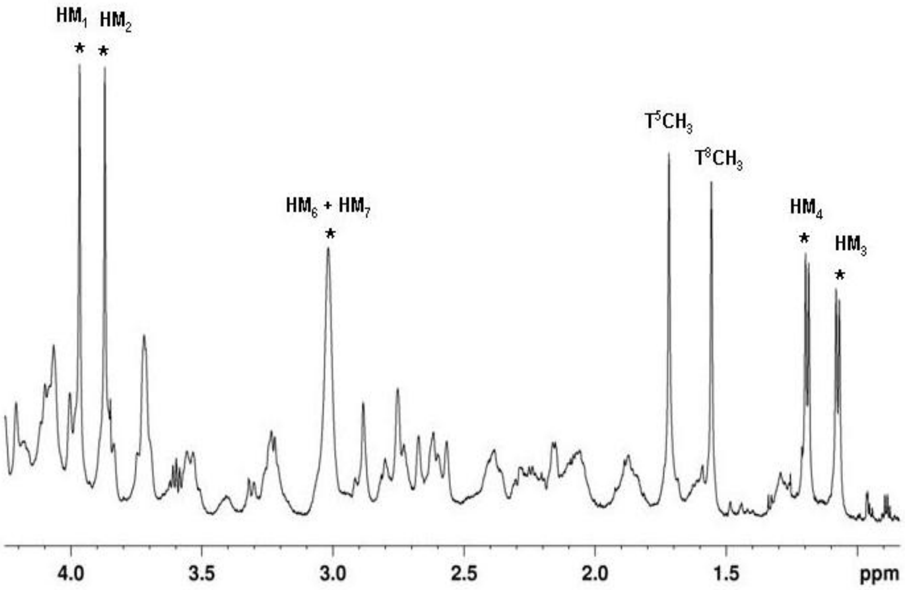


Figure 4.25 1D ^{1}H NMR data in the aliphatic resonance region for the binding of 2 equiv. AIK18-51 at the ACTAGT sequence. Ligand methyl resonances are indicated by *.

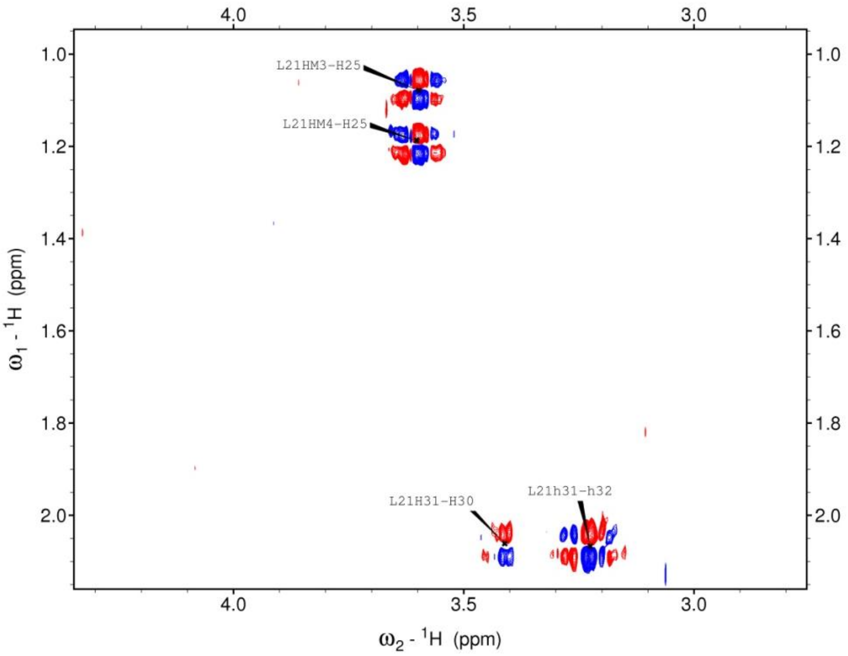


Figure 4.26 Part of the 2D DQFCOSY NMR spectrum of the 2:1 complex of AIK18-51 with d(CGACTAGTCG)$_2$ showing the observed COSY cross peaks between the protons of the isopropyl group and between some protons in the DMAP tail.

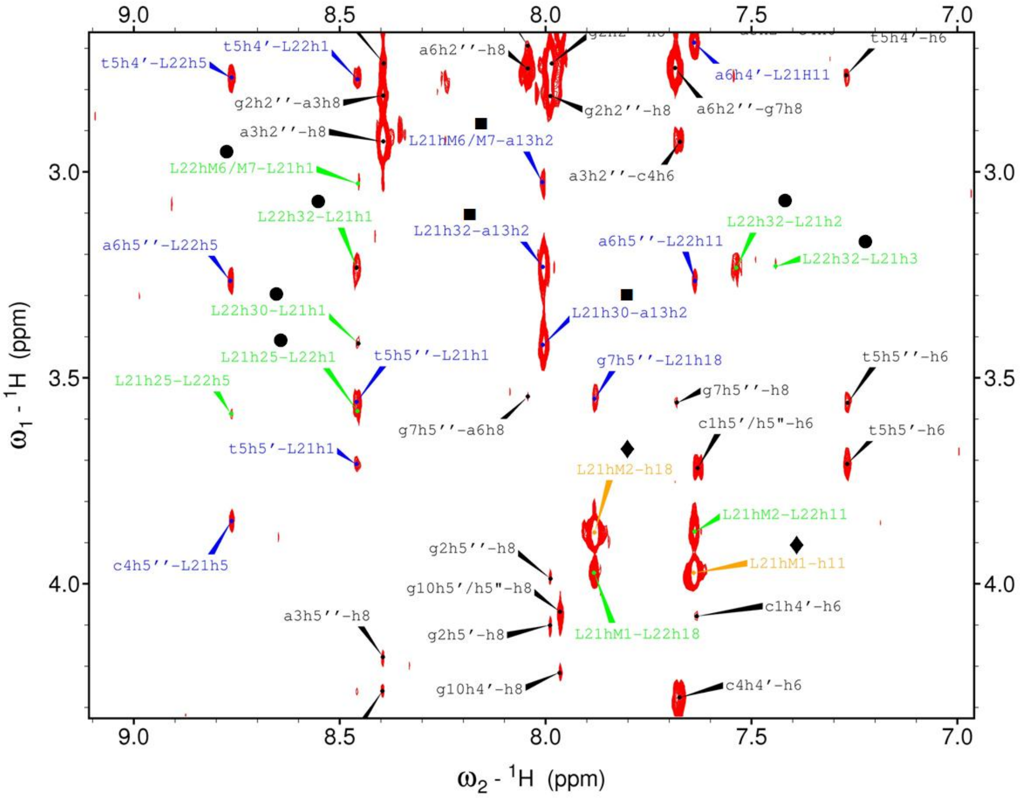


Figure 4.27 Part of the 2D [^{1}H, ^{1}H] NOESY NMR spectrum of the 2:1 complex of AIK18-51 with d(CGACTAGTCG)$_2$ showing: **1)** strong intra-ligand NOE cross-peaks between the pyrrole *N*-methyl groups and their associated ring protons (indicated by ♦); **2)** inter-ligand NOE cross-peaks between the pyridine head and the DMAP tail protons (indicated by ●); **3)** inter-ligand-DNA NOE cross-peaks between the DMAP tail proton resonances of the first ligand (L21) and A^{13}H2 of the second strand of the duplex (indicated by ■). Data labeling scheme: DNA resonance assignments, black labels; ligand-DNA resonance assignments, blue labels; intra-ligand NOEs, orange labels; inter-ligand NOEs, green labels.

The NOE assignment of the aromatic protons associated with the pyridine ring relied upon the assignment of both DQFCOSY and TOCSY spectra associated with these protons (Figure 4.28). Strong DQFCOSY cross-peaks were observed between H1 and H2, and between H2 and H3, whilst a weak cross-peak was observed between H3 and H5 (Figure 4.28 B). TOCSY cross-peaks were also observed between the following pairs of protons: H1/H2, H2/H3, H3/H5, H1/H3 and H2/H5 (Figure 4.28 A). These cross-peaks were mirrored by NOE cross-peaks in the 2D [^{1}H,^{1}H] NOESY NMR spectrum.

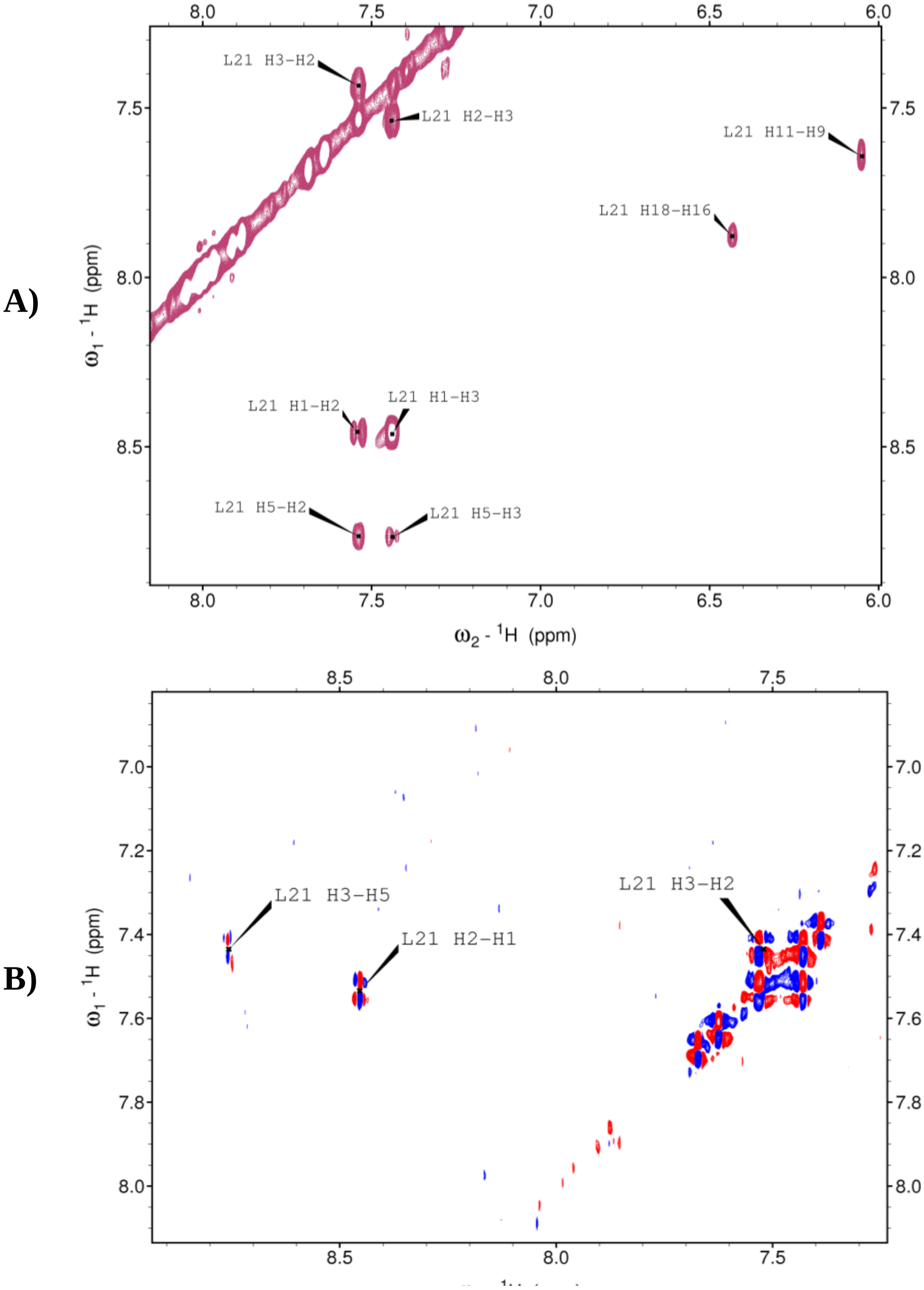


Figure 4.28 A) Part of the 2D [^{1}H, ^{1}H] TOCSY NMR spectrum of the 2:1 complex of AIK18-51 with d(CGACTAGTCG)$_2$ showing the observed TOCSY cross-peaks between the aromatic protons of AIK18-51. **B)** Part of the 2D [^{1}H, ^{1}H] DQFCOSY NMR spectrum of the 2:1 complex of AIK18-51 with d(CGACTAGTCG)$_2$ showing the observed COSY cross-peaks between the protons of the pyridine ring.

The signals at 9.738, 10.930, 10.650, and 8.279 ppm were associated with the amide *NH* protons: H7, H14, H21, and H29 of AIK18-51, respectively (Figure 4.29). Assignment of these signals was made on the basis of chemical shift location, labile proton exchange characteristics, peak shape, and detailed analysis of the observed intra-ligand, inter-ligand, and ligand-DNA NOE contacts.

The resonances at 6.045, 7.637, 6.426, and 7.882 ppm were assigned to pyrrole ″H9″, ″H11″, ″H16″, and ″H18″ protons in AIK18-51 respectively. NOESY cross-peaks were observed at 6.045/7.637 ppm and 6.426/7.882 ppm and were mirrored by TOCSY cross-peaks (Figure 4.28 A, H11/H9 and H18/H16). The signal at 6.045 ppm showed a strong NOE cross-peak with both H7 and H14 (Figure 4.29), whist the signal at 7.637 ppm showed a strong NOE cross-peak with HM1 (Figure 4.27). The signals at 6.045 and 7.637 ppm were therefore assigned to H9 and H11, respectively. Similarly, the signal at 6.426 ppm showed strong NOE cross-peaks to both H14 and H21 (Figure 4.29) whist the signal at 7.637 ppm showed a strong NOE cross-peak with HM2 (Figure 4.27). Hence the signals at 6.426 and 7.882 ppm were assigned to H16 and H18, respectively. It was therefore possible to make a complete assignment of the ^{1}H NMR resonances of AIK18-51 bound to d(CGACTAGTCG)$_2$, full details of which are listed in Table 4.9

^{1}H NMR resonance assignments of both the free and ligand-bound ODN (d(CGACTAGTCG)$_2$) enabled a direct comparison between the chemical shifts of free and ligand-bound ODN to be made. Such a comparison should reveal the location of the binding site as the chemical shifts of NMR resonances associated with DNA atoms which are in close contact with the binding ligand would be largely affected. Large chemical shift changes were expected for resonances of the sugar protons H1′, H4′, H5′, and H5″ as the ligand binds in the minor groove of DNA. The difference in chemical shift ($\Delta\delta$ ^{1}H) between the free and ligand bound DNA is listed in Table 4.10.

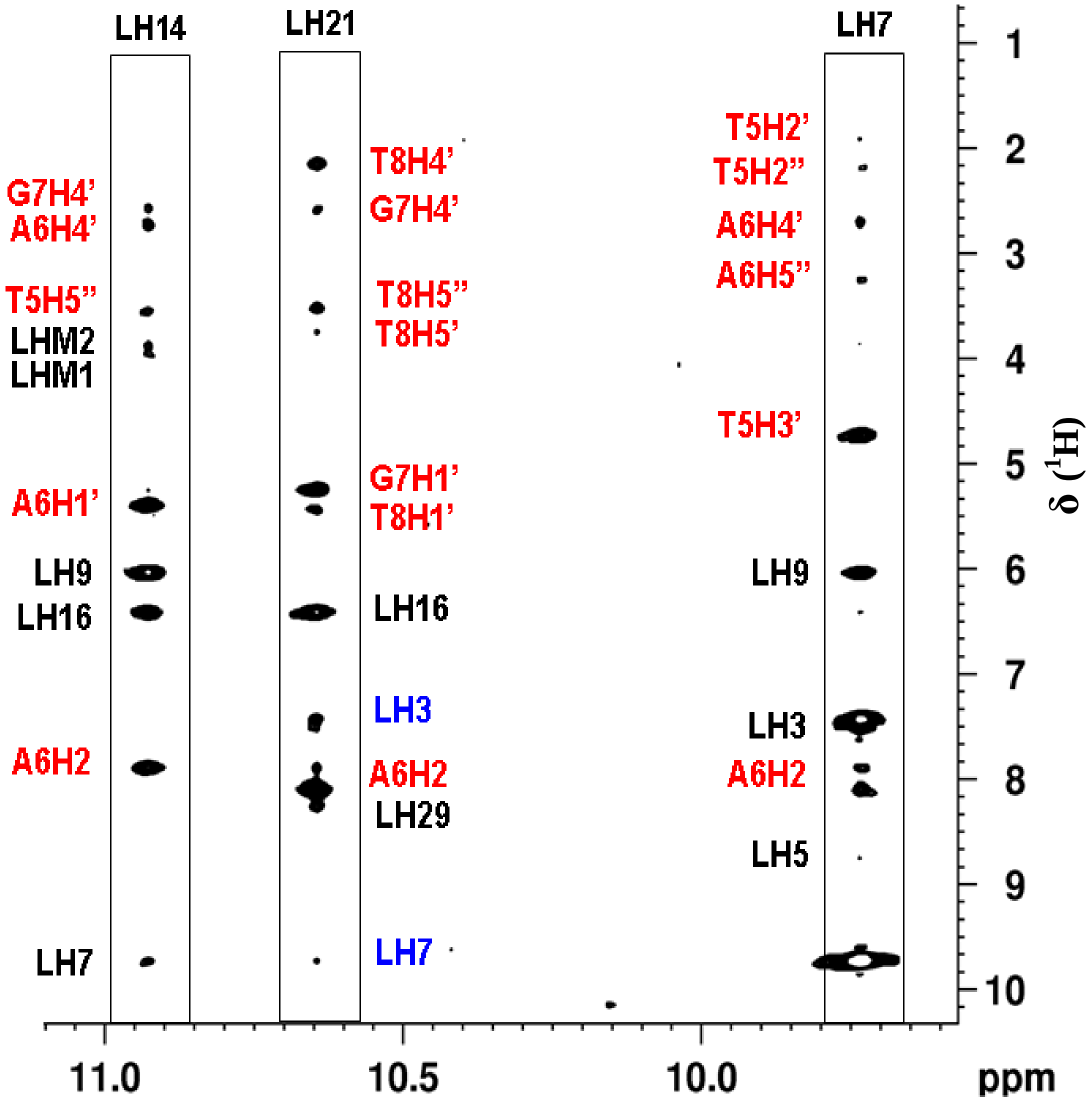


Figure 4.29 Strip plots of data taken from the 100 ms 2D [^{1}H, ^{1}H] NOESY NMR spectrum acquired on the 2:1 complex between AIK18-51 and d(CGACTAGTCG)$_2$ at 600 MHz. Data are shown at the ligand resonance chemical shifts of H7, H14 and H21. Data labeling scheme: DNA resonance assignments, red labels; ligand resonance assignments, black labels; inter-ligand NOEs, blue labels.

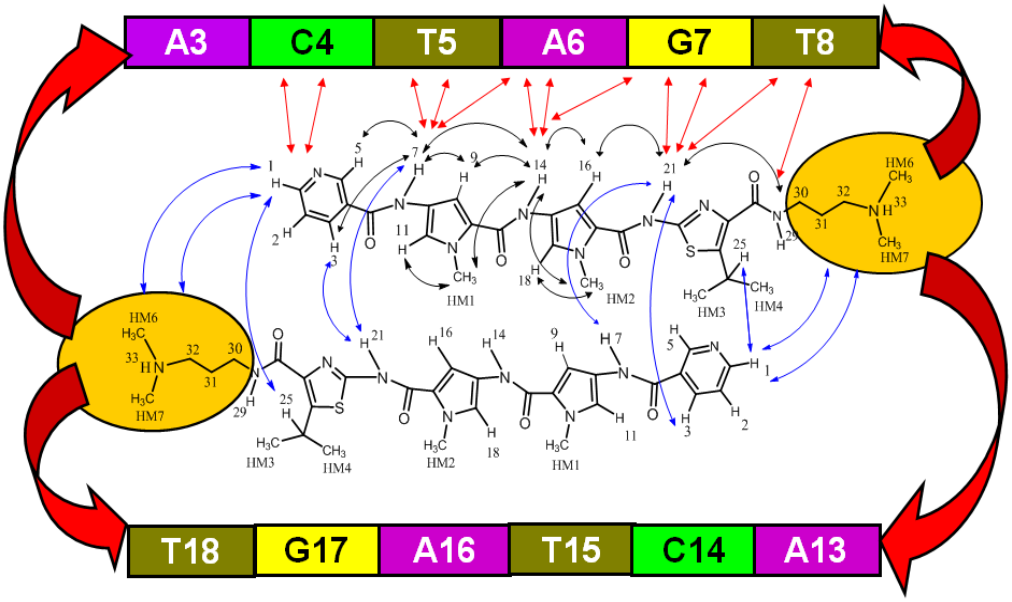


Figure 4.30 A schematic representation of how some of the NOEs relate to the structure of the complex. Ligand-DNA contacts, red arrows; intra-ligand contacts, black arrows ; inter-ligand contacts, blue arrows.

Table 4.9 ^{1}H NMR chemical shift assignments of AIK18-51 in a complex with d(CGACTAGTCG)$_2$.

Chemical shift assignment: δ 1H (ppm)										
H1	H2	H3	H5	H7	H9	H11	H14	H16	H18	H21
8.455	7.532	7.436	8.764	9.738	6.045	7.637	10.930	6.426	7.882	10.650
Chemical shift assignment: δ 1H (ppm)										
H25	H29	H30	H31	H32	H33	HM1	HM2	HM3	HM4	HM (6&7)
3.583	8.279	3.421	2.068	3.231	4.820	3.971	3.872	1.082	1.197	3.029

Table 4.10 ^{1}H chemical shift differences (Δδ) defined as (Δδ)= δ [(ligand-bound DNA)- δ(ligand free DNA)] for duplex d(CGACTAGTCG)$_2$.

	Chemical shift differences : Δδ ^{1}H (ppm)																
	H8	**H6**	**H2**	**H5**	**CH$_3$**	**H1′**	**H2′**	**H2″**	**H3′**	**H4′**	**H5′**	**H5″**	**GH1/ TH3**	**H41**	**H42**	**H21**	**H22**
Base																	
C1		0.011		0.011		0.027	0.016	0.014	0.026	0.008	n/a	n/a					
G2	0.021					0.081	0.001	-0.007	0.033	0.01	0.006	0.007	0.051				
A3	0.155					-0.301	-0.148	-0.003	0.043	-0.006	0.023	-0.017					
C4		0.409		0.26		0.484	-0.072	-0.219	-0.058	0.055		-0.482		-0.505	-0.15		
T5		-0.065			0.132	**-1.149**	-0.20	-0.277	-0.156	**-1.354**			-0.44			n/a	n/a
A6	-0.187		0.72			**-0.652**	**-0.515**	-0.152	-0.346	**-1.743**	-0.242	**-0.916**					
G7	0.092					**-0.607**	-0.367	-0.101	-0.189	**-1.800**	-0.373	**-0.729**	0.06				
T8		-0.426	n/a		0.304	**-0.586**	-0.436	**-0.727**	-0.392	**-2.057**		**-0.695**	0.15				
C9		-0.118		-0.285		-0.182	0.061	-0.082		-0.085				0.08	-0.14		
G10	0.011					0.035	0.003	0.02	0.005	0.015							

n/a- not applicable due to absence of protons in free DNA. Colour coding: figures are colour coded blue or red for ∣Δδ∣ > 0.1 ppm: red-base protons; blue-sugar protons. Bold typeface entries are for the most significant changes where ∣Δδ∣ > 0.5 ppm.

This study confirmed the position of the dimer formed by two molecules of AIK18-51 in the d(CGACTAGTCG)$_2$ sequence. By plotting the difference (Δδ) between the chemical shift of protons in the free DNA and the chemical shifts of the same protons in the ligand-bound DNA, it was possible to show the location of the ligand in the DNA sequence. A plot of Δδ for the H4′ resonances (Figure 4.31) indicated that the aromatic rings of the ligand were lying in the minor groove juxtaposed against the sugar rings of the sequence 5′-A^{3}C^{4}T^{5}A^{6}G^{7}T^8-3′ and the ligand was arranged in a 2:1, head-to-tail, side-by-side manner within the minor groove of DNA with the retention of duplex symmetry. The presence of these rings resulted in strong local magnetic shielding effects, and meant the location of the dimer could be determined by these effects on H4′ chemical shifts, which showed the biggest change when the ligand was bound, in association with H1′ significant shift changes (Figure 4.32).

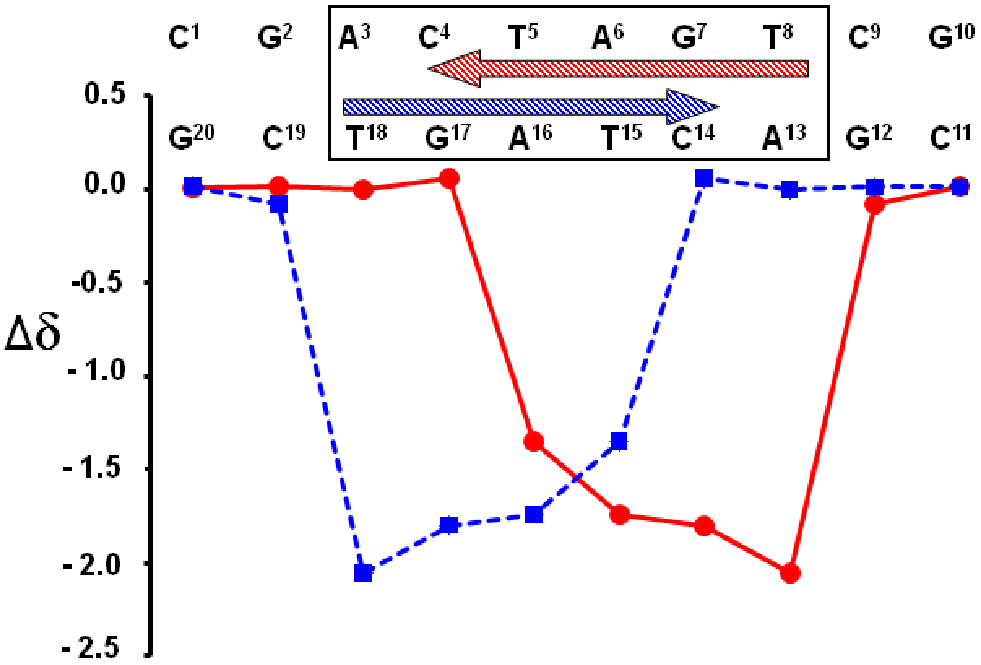


Figure 4.31 Chemical shift differences for H4′ resonances of ligand-bound and ligand-free DNA duplex [bound-free] for the complex between AIK18-51 and d(CGACTAGTCG)$_2$. Shaded arrows represent the location of the ligand relative to the DNA sequence. The dashed line shows the chemical shift changes for the opposing DNA strand.

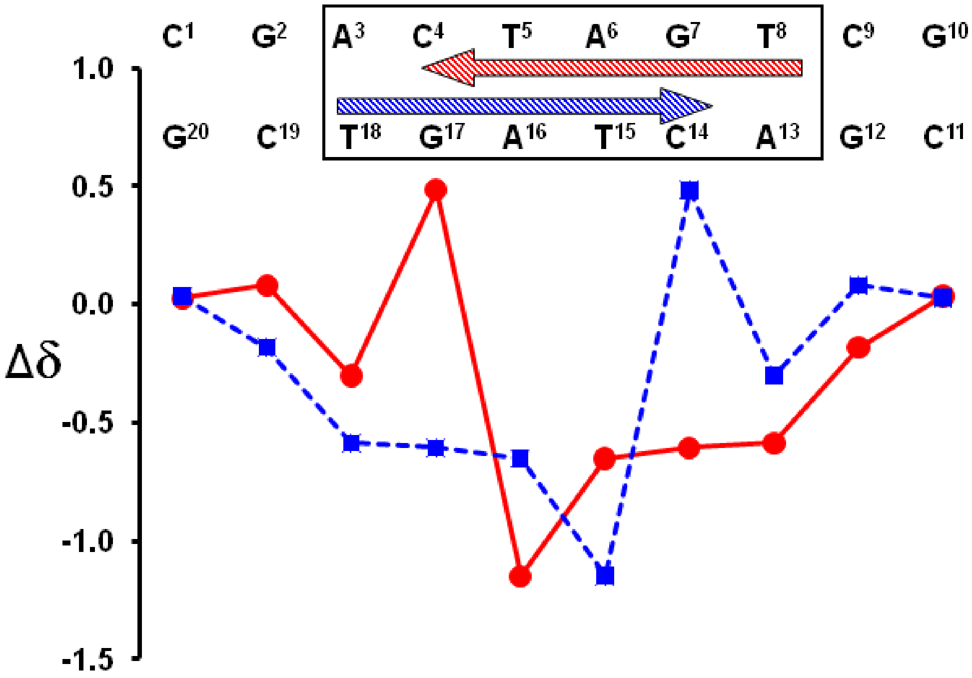


Figure 4.32 Chemical shift differences for H1′ resonances of ligand-bound and ligand-free DNA duplex [bound-free] for the complex between AIK18-51 and d(CGACTAGTCG)$_2$. Shaded arrows represent the location of the ligand relative to the DNA sequence. The dashed line shows the chemical shift changes for the opposing DNA strand.

Other changes in the chemical shifts (around $|\Delta\delta|$ = 0.4-1 ppm) are observed with H5″ (C4, A6 ,G7, & T8) and H2″ (A6 and T8). These changes were believed to be caused not only by the influence of nearby aromatic rings of the ligand (*via* van der Waals contacts) or the formation of hydrogen bonds with DNA bases, but also by changes in the DNA backbone conformation especially at the A^3pC^4 step where the highest positive change in the chemical shift was observed. The change in the chemical shift is attributed to an alteration in the geometry of the DNA backbone at these positions in the sequence. This was suspected from the changes in the ^{31}P NMR data of the DNA upon ligand binding (Figure 4.33). This perturbation in the chemical shift of ^{31}P upon AIK18-51 binding is in agreement with the previously determined complex between thiazotropsin A and d(CGACTAGTCG)$_2$. However, unlike the thiazotropsin A complex, the biggest change was seen at the A^3pC^4 step in the AIK18-51 complex, rather than the T^5pA^6 step. This may be explained by the fact that AIK18-51 is longer than thiazotropsin A, and the pyridine headgroup lying parallel to C4, would induce its effect at the A^3pC^4 step. The smaller formyl head group in thiazotropsin A, which also lies parallel to C4, would not induce such a change in the DNA. The change in the DNA backbone can be attributed to minor groove widening induced by the ligand binding. Resonance assignments for 2D [^{1}H, ^{31}P] HSQC NMR data (Figure 4.34) are listed for both free and ligand bound DNA in Table 4.11.

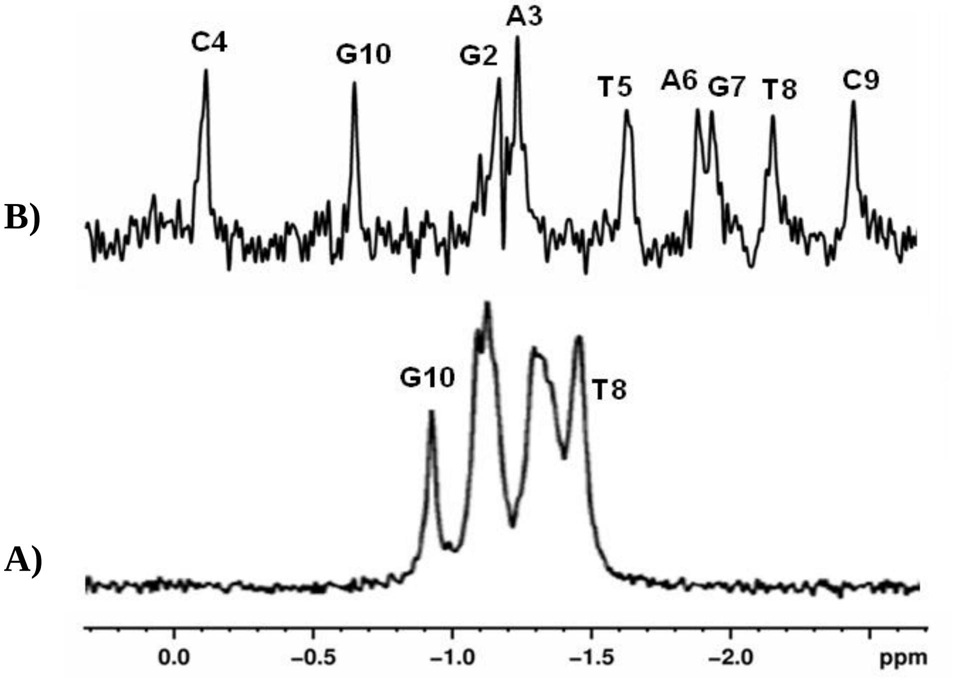


Figure 4.33 ^{31}P-{^{1}H} NMR spectra of d(CGACTAGTCG)$_2$ (A) and of the complex between d(CGACTAGTCG)$_2$ and AIK18-51 (B) acquired at 9.4 T. The effect of AIK18-51 binding to the DNA duplex was clear from the dispersion of signals that occurred for the complex, indicative of DNA backbone alteration.

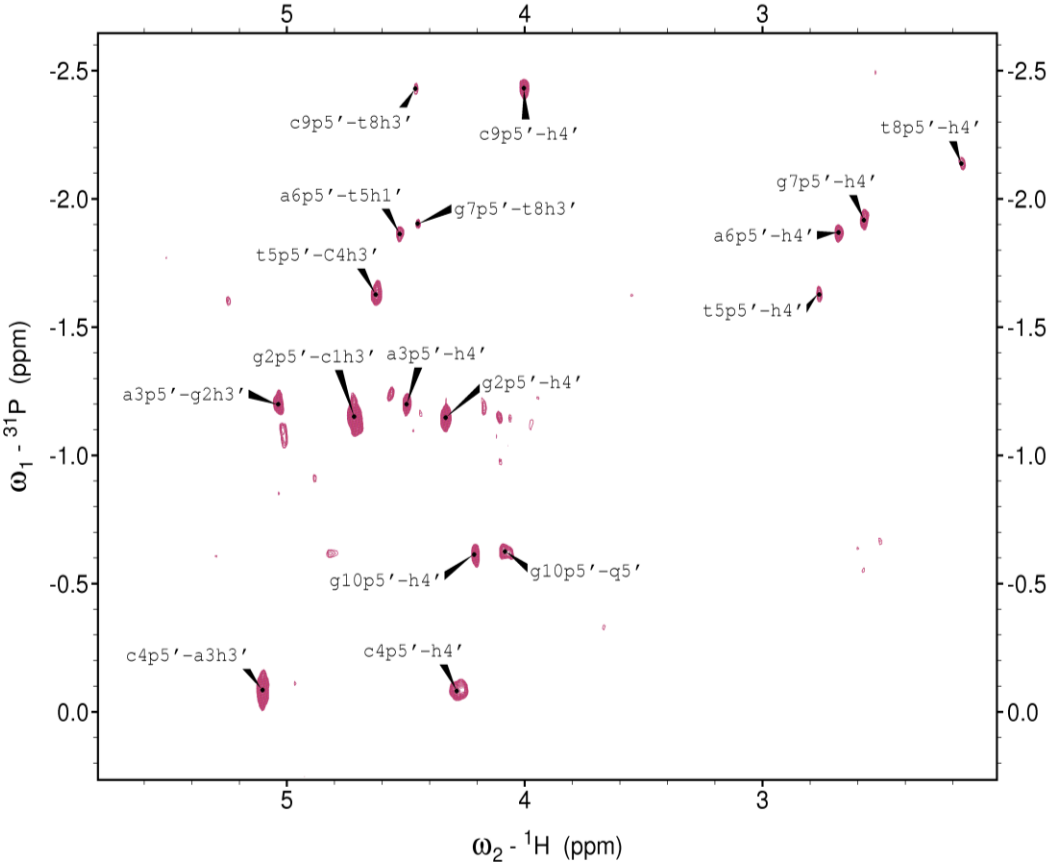


Figure 4.34 Part of the 2D [^{31}P,^{1}H] HSQC NMR data for the 2:1 complex between AIK18-51 and the d(CGACTAGTCG)$_2$ acquired at a magnetic field strength of 9.4 T.

Table 4.11 Comparison of ^{31}P chemical shift assignments for d(CGACTAGTCG)$_2$ in the absence (free) and presence (bound) of 2 equiv of ligand per duplex.

	Chemical Shift Assignment[a]: δ^{31}P (ppm)		**$\Delta\delta$ ^{31}P(δ_{bound} - δ_{free}) (ppm)**
Bases	**Bound**	**Free**	
C1	-	-	-
G2	-1.148	-1.04	-0.108
A3	-1.199	-1.09	-0.109
C4	-0.085	-1.29	**1.205**
G5	-1.627	-1.04	-0.587
C6	-1.869	-1.37	-0.499
G7	-1.905	-1.13	-0.775
T8	-2.133	-1.49	-0.643
C9	-2.43	-1.33	-1.100
G10	-0.624	-0.84	0.216

[a] Assignments are for the phosphates that are 5′ with respect to the base.

4.5 *Molecular modeling*

Proton-proton NOEs assigned from 2D [^{1}H, ^{1}H] NOESY NMR data acquired for d(CGACTAGTCG)$_2$ and for the 2:1 complex between AIK18-51 and d(CGACTAGTCG)$_2$ were used for model building. Intermolecular ligand-DNA and ligand-ligand NOE contacts unambiguously confirm the formation of the 2:1 AIK18-51-d(CGACTAGTCG)$_2$ complex at a six base pair binding site (5′-ACTAGT-3′) in the minor groove of DNA. The NOEs of the complex were consistent with the previously determined structure of thiazotropsin A-CGACTAGTCG and thiazotropsin B-CGACGCGTCG complexes, in which, the ligand is arranged in a 2:1, head-to-tail, side-by-side manner within the minor groove of DNA.

4.5.1 Solution structure of the free d(CGACTAGTCG)$_2$

The solution structure of d(CGACTAGTCG)$_2$ was determined in order to assess the effect of ligand binding on the structure of the DNA duplex. The Watson-Crick base pairing of the free duplex was intact as indicated by the presence of imino proton ^{1}H NMR resonances in the NMR data. The starting model of the free d(CGACTAGTCG)$_2$ was constructed in the canonical B-form as the NMR data were more consistent with the right-handed B-form of a DNA duplex than other DNA structures. A total of 391 NOEs were incorporated into the starting structure as distance restraints with a force constant of 1 kcal mol^{-1} Å^{-2}. Representations of the resulting calculated average structure are shown in Figure 4.35. The calculated structure of the free d(CGACGCGTCG)$_2$ was analysed using CURVES.[134] The overall nature of d(CGACTAGTCG)$_2$ did not deviate excessively from that of canonical B-form DNA. The average minor groove width increased from 5.9 Å in canonical B-DNA to 6.6 Å in the calculated structure. This may be caused by the salts present in solution used to stabilise the duplex, which can affect the conformation of nucleic acids.

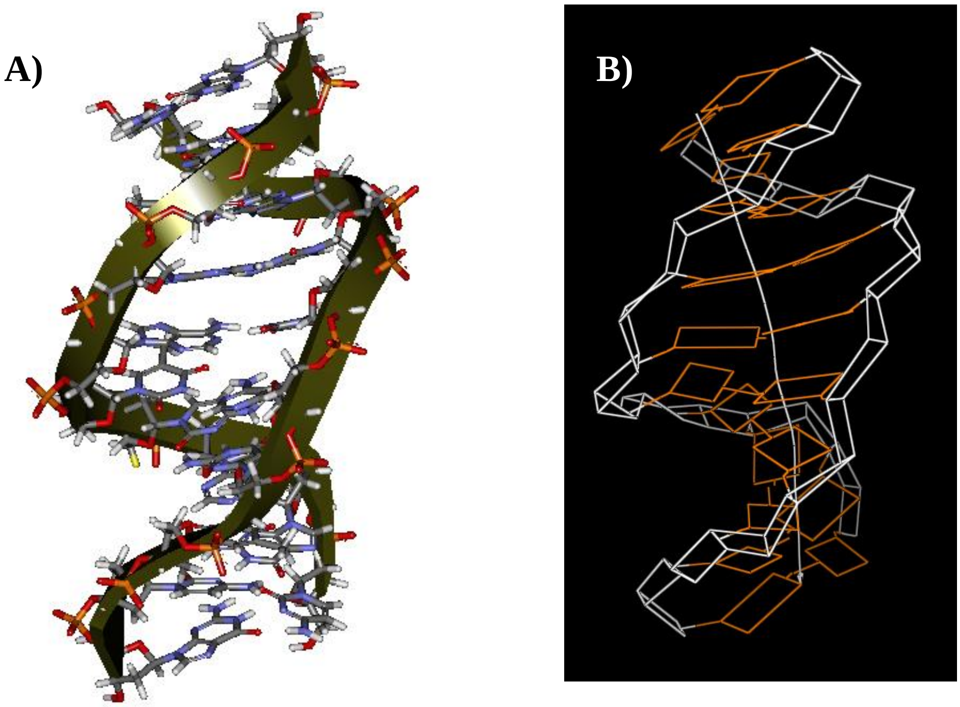


Figure 4.35 Representation of the solution structure of d(CGACTAGTCG)$_2$ alone based on restrained molecular dynamics simulations. **A)** Stick vs. arrows representation of the average structure of d(CGACTAGTCG)$_2$ taken from 100 ps of restrained molecular dynamics simulations. **B)** CURVES carton representation of the average structure of d(CGACTAGTCG)$_2$.

The structure displayed an overall axis curvature of 18.1 (see Figure 4.17 for the pictorial definitions of the parameters that relate to the DNA bases). Tip angle (θ), x- and y-displacement of base pairs was negligible (average θ = 1.48°, average dx = -1.57 Å, average dy= -0.19 Å) whilst inclination of the base pairs was significant (average η = 5.3°). Average values of base pair shear (S_x), stretch (S_y) and stagger (S_z) were similar to those of canonical B-DNA. Some variations in the buckle (κ), propeller twist (ω) and opening (σ) of base pairs were noted (average κ = -6.17, average ω = -4.45, average σ = -2.30). Average global inter-base pair parameters shift (Dx), slide (Dy), rise (Dz), tilt(τ), roll (ρ) and twist (Ω) were similar to those of canonical B-DNA. Sugar puckers fell into either C2′-endo or C3′-exo configurations.

4.5.2 Solution structure of the 2:1 AIK18-51/d(CGACTAGTCG)$_2$ complex.

A total of 757 intermolecular ligand-DNA and ligand-ligand together with intramolecular DNA-DNA and ligand-ligand distance restraints derived from NOE data were used to obtain the energy minimized model of the d(CGACTAGTCG)$_2$-AIK18-51 complex (See Appendix 2 for full details of NOE assignments). The two ligands fit snugly into the DNA minor groove, stacked next to each other with the positively charged DMAP tails pointing in opposite directions. As noted previously, 1D ^{1}H NMR data revealed that complex formation between AIK18-51 and d(CGACTAGTCG)$_2$ was symmetrical. This occurs only if the ligand is arranged in an antiparallel side-by side mode within the minor groove of a DNA duplex. The NOE data also confirmed this mode of binding through the observed NOEs between the DMAP tail of one ligand and the pyridine head of the partnering ligand and *vice versa*. The change in chemical shift of the DNA proton resonances firmly placed AIK18-51 in the minor groove of the duplex as expected and ligand/DNA NOEs supported this finding. A representation of the calculated solution structure of the complex is shown in Figure 4.36A. The structure shown represents the average structure of the complex taken from the final 100 ps of 300 ps of restrained molecular dynamics simulation in explicit solvent.

The hydrogen- bonding scheme for the complex was deduced indirectly on the basis of the calculated solution structure (Figure 4.37 and Table 4.12). The amide protons and the nitrogen of the thiazole ring of AIK18-51 face the floor of the DNA minor groove and form hydrogen bonds with the nearby DNA bases. Hydrogen bonds were assigned between the AIK18-51 amide protons NH-7, NH-14, NH-21, and NH-29 with T^5 O2, A^6 N3, G^7 N3, and T^8 O2, respectively. The thiazole nitrogen of AIK18-51 forms hydrogen bonds with the exocyclic NH$_2$ of G^7. Remarkably, a hydrogen bond was assigned between H33 of the DMAP tail and A^{13} N3 of the second strand, which has not previously been seen before in any of our complexes. Several NOEs were observed between DMAP of thiazotropsin B and the A^{13} H2 proton of d(CGACGCGTCG)$_2$, indicating a close contact between the two moieties. However, the distance between H26 in the DMAP group of thiazotropsin B and A^{13} N3 was not

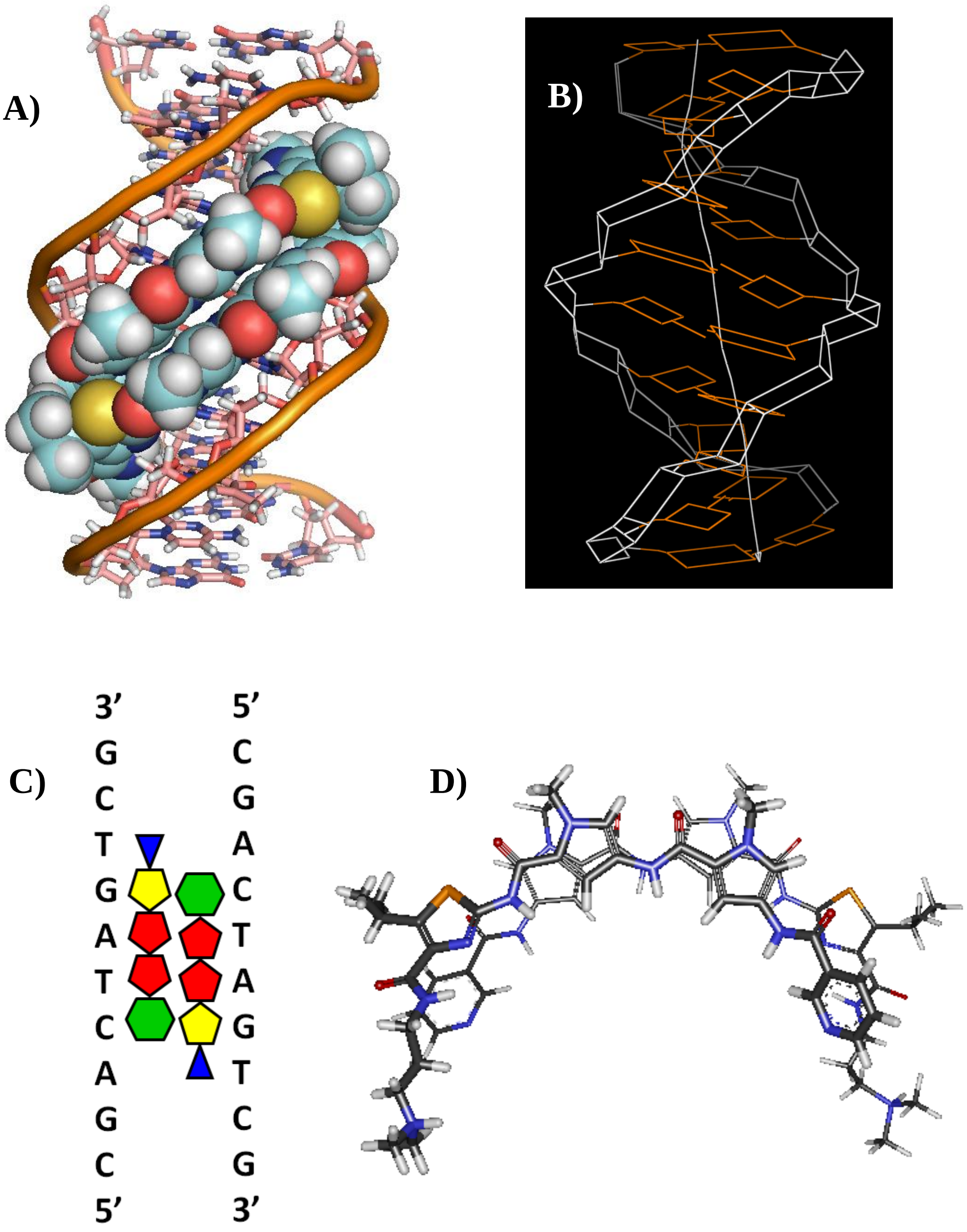


Figure 4.36 Cartoon and schematic representation of the complex between AIK18-51 and d(CGACTAGTCG)$_2$ showing the location of ligand with respect to the DNA sequence. **(A)** Representation of the refined solution structure of the complex between AIK18-51 (CPK drawing) and d(CGACTAGTCG)$_2$ (stick and tubes), **B)** CURVES cartoon representation of the average structure of the ligand-bound D(CGACTAGTCG)2. **C)** Schematic indicating the ligand alignment relative to the DNA sequence. Color coding: green hexagon = pyridine "head"; red pentagon =*N*-methylpyrrole; yellow pentagon = isopropylthiazole; blue triangle = DMAP "tail". **D)** Relationship between associated ligands in the complex.Thicker lines are shown for one ligand compared with its partner.

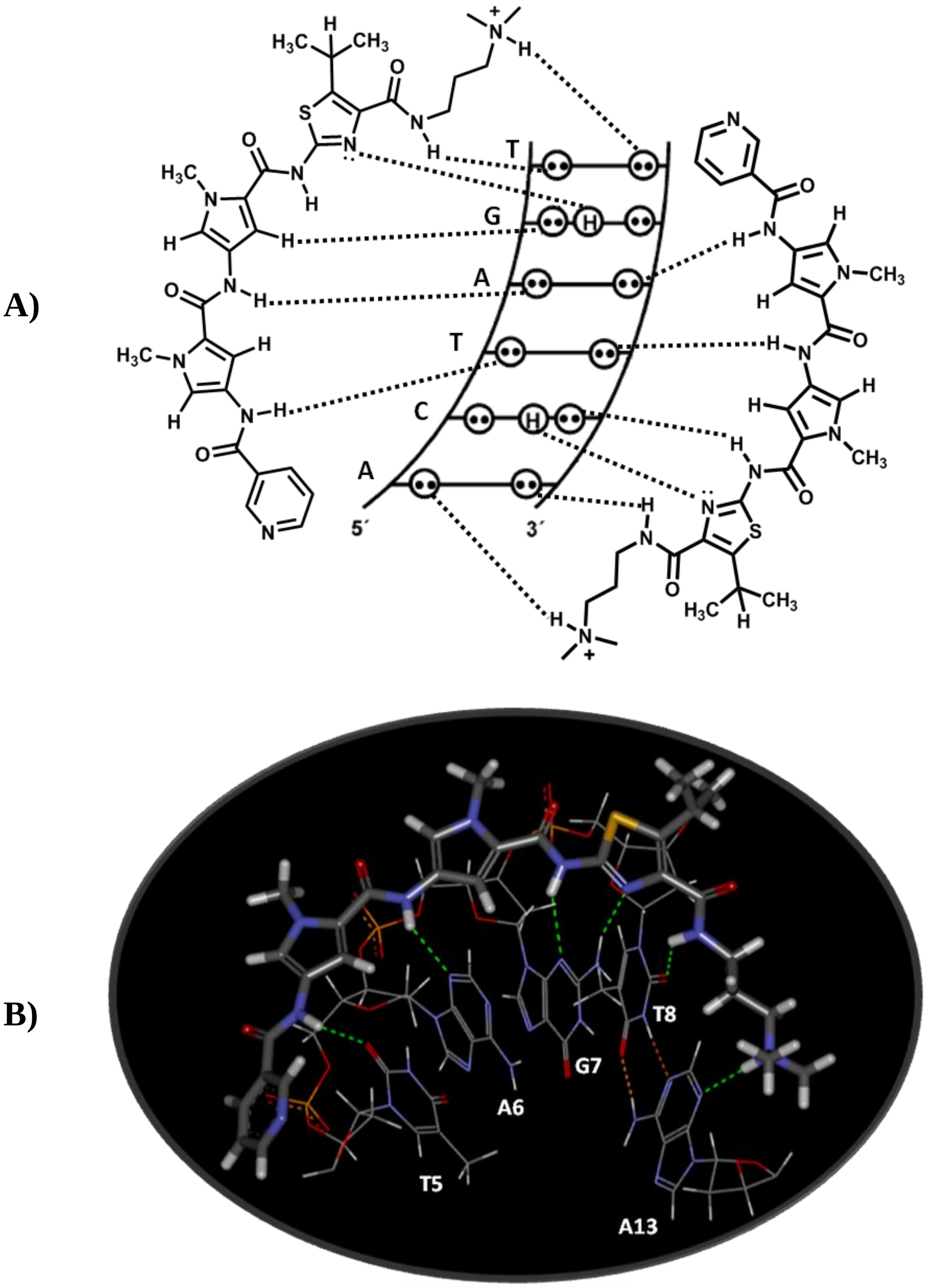


Figure 4.37 A) Deduced arrangement of hydrogen bonding between AIK18-51 and the DNA duplex d(CGACTAGTCG)$_2$; **B)** part of the calculated average structure showing the hydrogen bonds (green dashed lines) formed between one molecule of the AIK18-51 dimer (thick lines) and the DNA duplex. Hydrogen bonds were assigned using the Discovery Studio program.

Table 4.12 Summary of the hydrogen bonds in the 2:1 complex formed between AIK18-51 and d(CGACTAGTCG)$_2$ based on the labile proton exchange characteristics and solution structure information.

Ligand atom	**DNA atom**	**Distance (Å)**
NH-7	T^5 O2	1.9
NH-14	A^6 O2	2.2
NH-21	G^7 N3	2.5
NH-29	T^8 O2	2.2
NH-33	A^{13} H2	1.9
Thiazole N	G^7 H22	1.9

close enough for a hydrogen bond to be assigned as the the distance between the two atoms was > 2.5 Å. The labile protons of the ligand (amide protons) which form hydrogen bonds with DNA bases appeared to be inaccessible to solvent (see Figure 4.23). 2D [^{1}H, ^{1}H] NOESY NMR data acquired for the complex did not show cross-peaks between the ligands′ amide NH protons and the solvent resonance. The absence of an exchange correlation between the solvent resonance and that of the labile protons indicates that these protons are solvent inaccessible. This occurs as a consequence of the complex formation which makes the ligand-DNA binding interface less accessible to solvent when the ligand occupies the minor groove of DNA.

CURVES analysis of the ligand bound DNA duplex resulted in average DNA minor groove parameters of 7.5 Å wide and 4.6 Å deep in the-ACTAGT- region. Comparison is made in Table 4.13 of the dimensions of the minor groove for both canonical B-DNA and the NMR solution structures calculated for the ligand-free and ligand-bound DNA structures. These figures reflect the computational methods used rather than the data, which poorly defines features such as curvature and groove width. Nevertheless, it is of some interest to examine these values and particularly local perturbations, which may reflect true changes in DNA structure. In general, the global shape of the DNA structure upon ligand binding agreed well with that of the free DNA. However, some changes in the DNA structure conformation of the complex were observed as a result of side-by-side binding of AIK18-51. In addition to the increase in the width of minor groove of ligand-bound DNA duplex, the overall axis curvature increased from 18.1° in the calculated ligand-free B-DNA structure to 24° in the calculated ligand-bound B-DNA structure. Average global base pair axis parameters generally showed minor deviations from the values calculated for canonical B-DNA- (dx = -1.3 Å, dy = -0.28 Å, tip (θ) = -0.33°) although inclination (η) showed a higher average value (η = 8.8°). Remarkably, the central T^5A^6 base pairs showed significant inclination (22°) compared with the other base pairs.

Average values of base pair shear (S_x), stretch (S_y), stagger (S_z), buckle (κ), and opening (σ) were similar to those of canonical B-DNA but some variations in the propeller twist (ω) were noted (average ω = -16.1) especially at the central T^5A^6 base pairs which showed significant propeller twist (30°) compared with the other base pairs. Average global inter-base pair parameters namely shift (Dx), slide (Dy), rise (Dz), tilt(τ), roll (ρ) and twist (Ω) were in close agreement with the values for the calculated ligand-free DNA structure. Sugar puckers fell into either C2′-endo or C1′-

Table 4.13 Comparison of key global parameters for three DNA structures: canonical B-DNA, The calculated ligand-free B-DNA structure, and the calculated DNA structure in the presence of 2 molar equiv. of ligand (AIK18-51) bound in the minor groove of the DNA.

	Minor groove[1]		
	width (Å)	depth (Å)	overall axis bend (deg)
canonical B-DNA[2]	5.9	4.6	0
ligand-Free DNA[3]	6.6	3.9	18.1
DNA with 2 equiv of AIK18-51[3]	7.5	4.6	24

[1] Measured for the region -ACTAGT-. [2] Software-generated structure. [3] NMR structure

exo configurations. These changes in the DNA conformation are caused by the presence of side-by-side binding of AIK18-51 within the minor groove of the DNA duplex. The pyridine head group was orientated toward the 5′-end of the 5′-ACTAGT-3′ segment of DNA with the positively charged DMAP tail located toward the 3′ end.

4.6 Discussion of NMR –generated complex structures

4.6.1 Dimeric complex formation

Inter-molecular ligand-DNA and ligand-ligand NOE contacts obviously confirm the formation of 2:1 complex (ligand to DNA) for both thiazotropsin B and AIK18-51 with their DNA targets at a six base pair binding site in the minor groove of DNA. The thiazotropsin B dimer spans the 5′-ACGCGT-3′ sequence and the AIK18-51 dimer lies along the 5′-ACTAGT-3′ sequence. The positively charged DMAP group of each ligand points to the 3′-end of its DNA strand. The ligand arrangement is determined by a specific hydrogen bond between the imidazole/thiazole nitrogen of the carboxiamide ligands and the guanine amino group in the binding site. This hydrogen bond was deduced indirectly from the calculated solution structure and the observed NOE contacts between the amino protons of guanine and ligand protons in the 2D NOESY NMR spectrum. The appearance of these cross peaks indicates that the rotation of the amino group about the *N-C* bond is slowed significantly due to interactions with the imidazole/thiazole nitrogen.[150] Molecular modelling of these complexes suggests that the imidazole and thiazole nitrogens of the ligand are positioned almost ideally with respect to geometry and distance for the formation of a hydrogen bond to the guanine amino group.

Titrations of $d(CGACGCGTCG)_2$ and $d(CGACTAGTCG)_2$ with thiazotropsin B and AIK18-51, respectively yield only complexes with two ligands bound in the anti-parallel side-by-side fashion. No 1:l 1igand to DNA complexes were detected during the titrations. This is in agreement with the previously characterised dimeric complex of the Im-Pyr-Pyr analogue of distamycin with 5′-TGACT-3′[150] and the dimeric complex of the Pyr-Im-Pyr analogue of distamycin with various GC-containing sequences.[151-152] The relative positions and geometries of the dimeric ligands bound in the minor groove are very similar in all of these complexes. The ligand is arranged in a manner by which the amides of one ligand overlap the aromatic rings of the partner ligand. Moreover, only a single orientation of the ligand has been observed, with the DMAP group of the polyamide lying towards the 3′-end of the DNA strand with which it interacts.

The 2: l binding fashion (1igand to DNA) seems to be a general feature of the complexes formed by distamycin and its lexitropsin analogues with a minor groove of mixed GC/AT DNA sequences. This, perhaps, arises from the inherently wider grooves seen in GC-containing sequences.[153] Several studies have revealed that the minor groove width is very sequence dependent. DNA sequences containing four or more consecutive A bases show a characteristic narrowing of the minor groove.[154-155] The binding mode of distamycin has been shown to be very sensitive to this width.[156] Distamycin can form either a 1:1 complex or a 2:1 side-by-side binding complex depending on the DNA sequence with which it interacts. For sequences with five or more consecutive AT bases that have narrow minor grooves (e.g., 5′-AAAAA-3′) only 1:1 binding stoichiometry complexes are formed at low ligand ratios, while sequences with mixed AT or GC bases that have broad minor grooves (e.g. ATATA) only form 2:1 complexes regardless of ligand concentration.[157]

Although AIK18-51 is longer than thiazotropsin A and B (four aromatic rings compared with three aromatic rings), both of them span a six base pair binding site. This arises from the different configurations that can be adopted by the ligand dimer through their ability to adopt either the overlapped or the slipped dimer binding motifs. The AIK18-51 overlapped configuration (Figure 4.36 C, D), where all four aromatic rings are overlapped, allows the ligand to recognise a six base pair binding site (5′-ACTAGT-3′). The slipped configuration of thiazotropsin B (Figure 4.18 C,

D) , where two of the three aromatic rings stack on top of each other, enables the ligand dimer to also span a six base pair binding site (5′-ACGCGT-3′). Our ITC studies showed that thiazotropsin A, which is composed of three aromatic rings, can slide between five (5′-ACAGT-3′) and six (5′-ACTAGT-3′) base pair binding sites, a situation similar to that of distamycin bound to 5′-AAATT-3′ and 5′-AAATTT-3′.[156,158]

4.6.2 Chemical shift changes

With the NMR assignment of the free and bound duplexes for thiazotropsin B and AIK18-51 complexes, it was possible to make a comparison with the previously determined complex of thiazotropsin A. The difference in the chemical shift between the bound and free DNA for the three complexes were used to plot graphs for H1′, H2′, H2″, and H4′ protons (Figure 4.38 to Figure 4.41) in order to provide a visual comparison between the three complexes. A direct comparison was possible between AIK18-51 and thiazotropsin A as they both bind to the same DNA sequence. Although thiazotropsin B binds to a different DNA sequence, it was useful to compare the magnetic shielding effects induced by thiazotropsin B binding to the minor groove with the other ligands.

It was possible to assess from the different graphs the trend that $\Delta\delta$ is more negative for AIK18-51 than for thiazotropsin A at bases T^5 and A^6 . There is evidently a more pronounced shielding effect when DNA is bound to AIK18-51, which must be due to the presence of the pyridine group and the extended conjugation through four contiguous aromatic/amide ring systems.

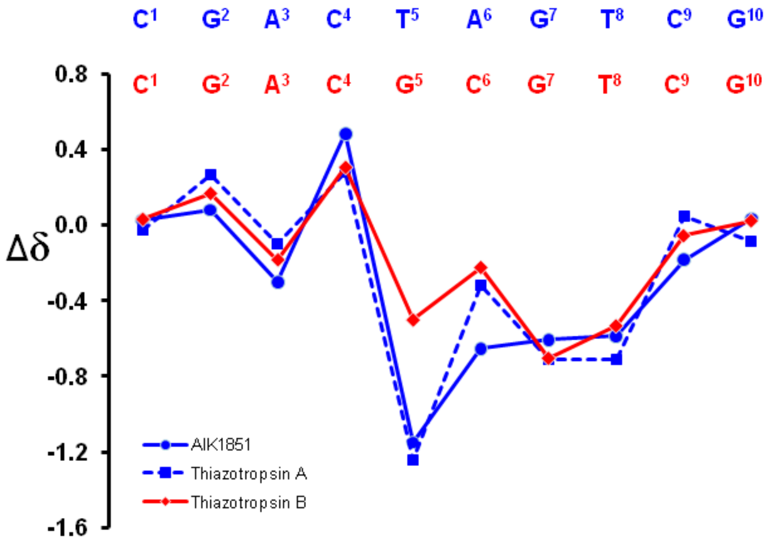


Figure 4.38 A graph showing chemical shift differences for H1′ resonances between bound and free DNA for AIK18-51 (represented by the blue curve), thiazotropsin A (represented by the dashed blue curve), and for thiazotropsin B (represented by the red curve). The DNA sequence to which AIK18-51 and thiazotropsin A were bound is coloured blue, whilst the DNA sequence to which thiazotropsin B was bound is coloured red.

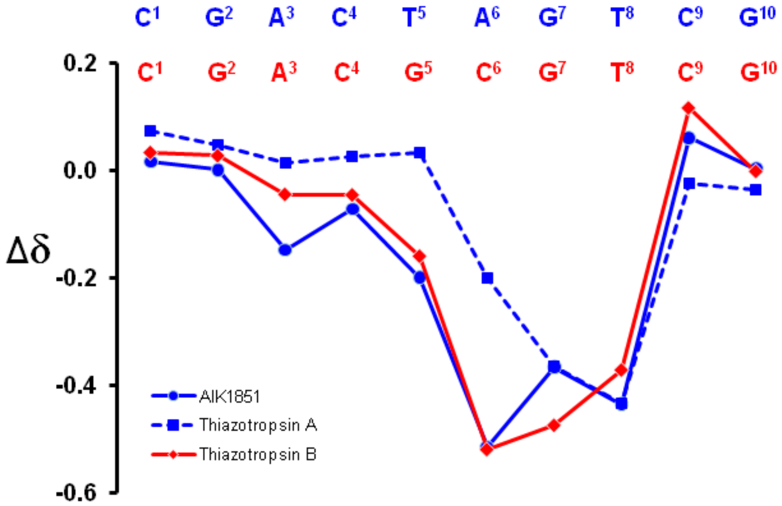


Figure 4.39 A graph showing chemical shift differences for H2′ resonances between bound and free DNA for AIK18-51 (represented by the blue curve), thiazotropsin A (represented by the dashed blue curve), and for thiazotropsin B (represented by the red curve). The DNA sequence to which AIK18-51 and thiazotropsin A were bound is coloured blue, whilst the DNA sequence to which thiazotropsin B was bound is coloured red.

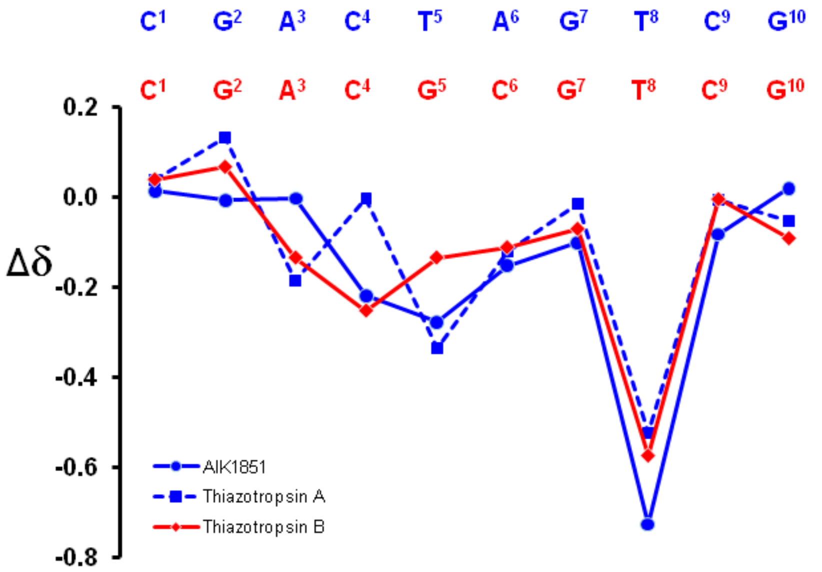


Figure 4.40 A graph showing chemical shift differences for H2″ resonances between bound and free DNA for AIK18-51 (represented by the blue curve), thiazotropsin A (represented by the dashed blue curve), and for thiazotropsin B (represented by the red curve). The DNA sequence to which AIK18-51 and thiazotropsin A were bound is coloured blue, whilst the DNA sequence to which thiazotropsin B was bound is coloured red.

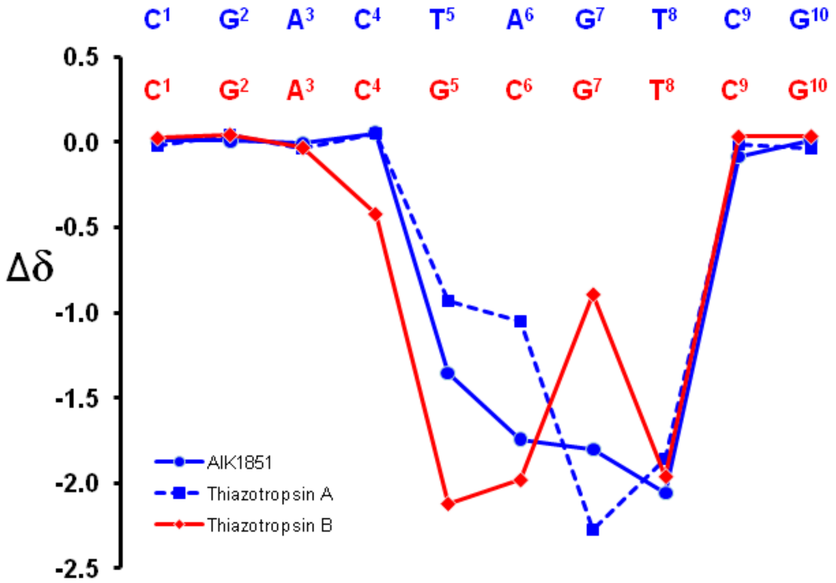


Figure 4.41 A graph showing chemical shift differences for H4′ resonances between bound and free DNA for AIK18-51 (represented by the blue curve), thiazotropsin A (represented by the dashed blue curve), and for thiazotropsin B (represented by the red curve). The DNA sequence to which AIK18-51 and thiazotropsin A were bound is coloured blue, whilst the DNA sequence to which thiazotropsin B was bound is coloured red.

The biggest perturbations in the chemical shift of DNA protons were observed for the sugar H1′ and H4′ protons, which are located in the minor groove wall, but significant (> 0.1 ppm) shifts were observed for other DNA protons as well. The chemical shift perturbations are quite similar for the three complexes, and the pattern indicates that the ligand dimers span a 6 base pairs DNA segment. The chemical shift changes for the H1′, H2′, H2″, and H4′ of the three complexes have the same trend in general with some variations at certain positions. For instance, the H4′ chemical shift changes were not uniformly high for thiazoptropsin B; G^7H4′ does not change as much as the G^5, C^6 or T^8 H4′ resonances. There are different factors that could affect these changes in chemical shift, such as the exact geometry of the DNA protons relative to the aromatic rings of the binding ligand, the shielding effect induced by van der Waals interaction, or the effects brought about by hydrogen bond formation. The close proximity of the aromatic rings of the binding ligand to the groove walls of DNA induced upfield shift of H4′ resonances indicating that the ligand penetrates deeply into the DNA minor groove and has close contacts with the walls. However, the close contact between the DNA atoms and the ligand is not the only factor that dictates the chemical shift perturbations. The geometry of the sugar rings plays a key role in dictating the occurrence of the shielding or deshielding effects. The sugar protons situated at the walls of the minor groove (e.g. H4′ and H1′) usually experience large upfield shifts if they are positioned perpendicular to the ring plane of the ligand, while the H2 resonances (located at the floor of the minor groove) experience large downfield shifts because they are positioned parallel with the ring plane[159] (Figure 4.40). A combined influence of all these factors may explain the variations observed in the chemical shift changes of the studied complexes.

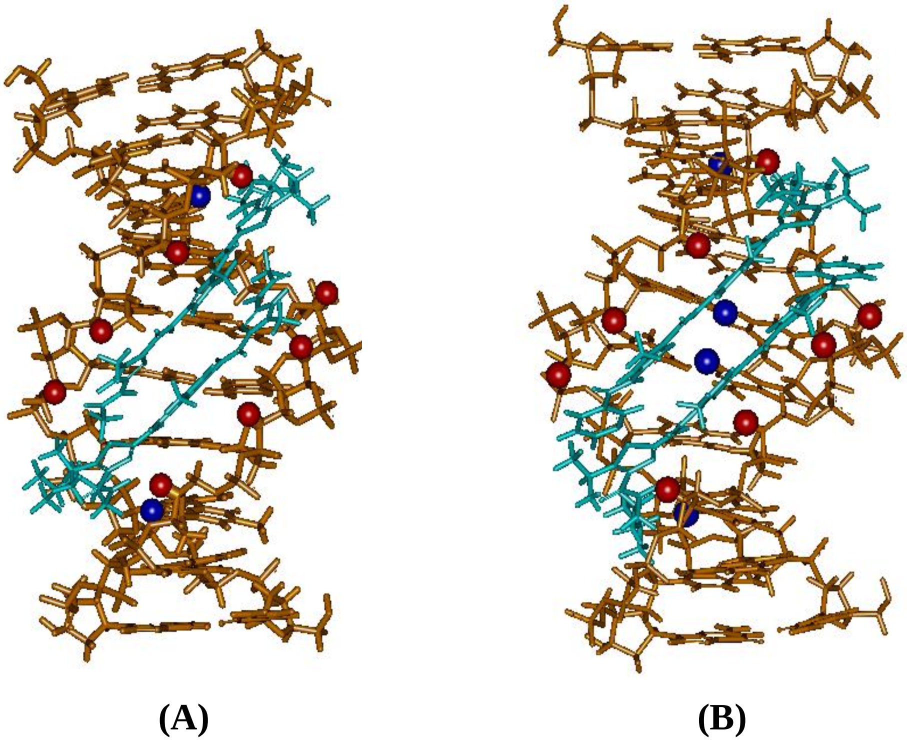

(A) **(B)**

Figure 4.42 The NOE-based models of **(A)** thiazotropsin B-CGACGCGTGC and **(B)** AIK18-51-CGACTAGTGC complexes showing the aromatic rings of ligands inserted edge-on into the DNA minor groove and highlighting the H4' protons experiencing an upfield shift greater than -0.8 ppm (in red). The H2 protons of adenine (highlighted in blue), which are located on the floor of the minor groove experience downfield shifts because they are positioned parallel to the ring plane (A^6 H2 and A^{16} H2 in the AIK18-51 complex), while the H2 proton resonances of A^3 and A^{13} were not affected as their position is distant from the aromatic rings of the ligand.

4.6.3 Lexitropsin-induced DNA structural perturbations

Comparison of the ^{31}P NMR chemical shifts for free and ligand-bound DNA allowed general deductions to be made concerning the changes in DNA backbone conformation. As shown in Figure 4.15 and Figure 4.33, all of the ^{31}P NMR resonances were observed within a narrow range of chemical shifts for the free oligonucleotide duplexes d(CGACGCGTCG)$_2$ and d(CGACTAGTCG)$_2$, but differed for the ligand-bound DNA duplexes. Changes in ^{31}P chemical shifts that occurred upon ligand binding (Table 4.5 and 4.11) are characteristic of changes in backbone conformation. These results are in agreement with the changes in the ^{31}P chemical shifts observed in the previously determined complex between thiazotropsin A and d(CGACTAGTCG)$_2$.[38] Chemical shift changes to higher ppms are indicative of changes in the phosphate-ester torsion angle.[160] In our study, the C^4-5′- phosphate in the AIK18-51-d(CGACTAGTCG)$_2$ complex and the G^5-5′ and C^6-5′ phosphates in thiazotropsin B-d(CGACGCGTCG)$_2$ complex experienced a substantial positive chemical shift change (downfield) when the ligand was bound. The resonances of other phosphates located in the binding site showed a negative chemical shift change (upfield) when the ligand was bound. This is highly likely to be caused by the magnetic shielding effects promoted by the nearby ligand through van der Waals interaction. The change in the chemical shift of the neighbouring H5′′ protons support this conclusion since their resonance experienced a dramatic upfield chemical shift change (negative) when the ligand was bound, a situation similar to that observed with the H4′ protons described earlier. We attributed the magnetic deshielding (downfield positive change) of the ^{31}P resonances to the influence of a large backbone conformational change, while the shielding (upfield negative change) observed in the ^{31}P chemical shifts to the influence of ligand binding. The highest positive change in ^{31}P chemical shift was observed at the A^{3}pC^4 step (1.2 ppm) in the AIK18-51-d(CGACTAGTCG)$_2$ complex and at the G^{5}pC^6 step (0.9 ppm) in the thiazotropsin B-d(CGACGCGTCG)$_2$ complex. These variations in the location and magnitude of chemical shift change may reflect the differences in the position and extent of the structural perturbation in the backbone of DNA between the two complexes, respectively, and is a consequence of the differing lengths of the ligands. The pyridine head of AIK18-51, which lies parallel to C^4, is more bulky than the formyl head of thiazotropsin A and is likely to be responsible for the greater changes observed in the ^{31}P chemical shifts of the AIK18-51-d(CGACTAGTCG)$_2$ complex .

The NOE based models of the studied lexitropsins-DNA complexes have shown that these ligands bind to the minor groove as dimers in an anti-parallel side-by-side fashion and the hydrogen bonding play a key role in the complexation process (see Figure 4.19 and 4.37). Although the NMR refined structure of the free

$d(CGACGCGTCG)_2$ and $d(CGACTAGTCG)_2$ revealed that the width of their minor grooves was intrinsically wider than that of canonical B-DNA (due to the salt concentration effects), ligand binding induced further widening in the minor grooves particularly with the AIK18-51-$d(CGACTAGTCG)_2$ complex and to less extent with the thiazotropsin B-$d(CGACGCGTCG)_2$ complex, presumably because GC rich sequences have wider minor grooves, (Table 4.7 and Table 4.13). The NOE based models of the complexes AIK18-51-$d(CGACTAGTCG)_2$ and thiazotropsin B-$d(CGACGCGTCG)_2$ revealed 1.0 and 0.5 Å widening in their minor grooves when compared to the ligand free DNA, respectively. The binding of thiazotropsin B to the minor groove was associated with a minimal disruption to the DNA structure contrary to what was observed with AIK18-51 and thiazotropsin A binding to the minor grooves. In fact, visual inspection for the dispersion of ^{31}P signals upon ligand binding, which can provide an indication of DNA backbone alteration, supports this observation. All the ^{31}P signals were completely dispersed upon thiazotropsin A (Figure 4.43) and AIK18-51 (Figure 4.33) binding to $d(CGACTAGTCG)_{2,}$ but in the case of thiazotropsin B binding with $d(CGACGCGTCG)_2$ (Figure 4.15), the ^{31}P signals of G^2, A^3, C^9 and G^{10}, and G^5 and C^6 were overlapped, which suggests that the binding is associated with slight structural altrations. These results are consistent with the ITC studies of thiazotropsin A and B, which have shown that the interaction thiazotropsin A with $d(CGACTAGTCG)_2$ is mainly enthalpically driven through hydrogen bonding and van der Waals interactions and associated with unfavourable entropy contribution, which is indicative of induced fit binding and regidification of the complex upon binding. Binding of thiazotropsin B to $d(CGACGCGTCG)_2$ is also mainly driven by enthalpy through hydrogen bonding and van der Walls forces. Contrary to thiazotropsin A, the binding is associated with a favourable entropy contribution, which is likely to be caused by both the release of more water molecules from the minor grooves and the absence of large perturbation to the DNA structure as a result of the wide minor groove that GC rich sequences possess, which allows the ligand to fit into the minor groove without inducing large DNA structural alterations.

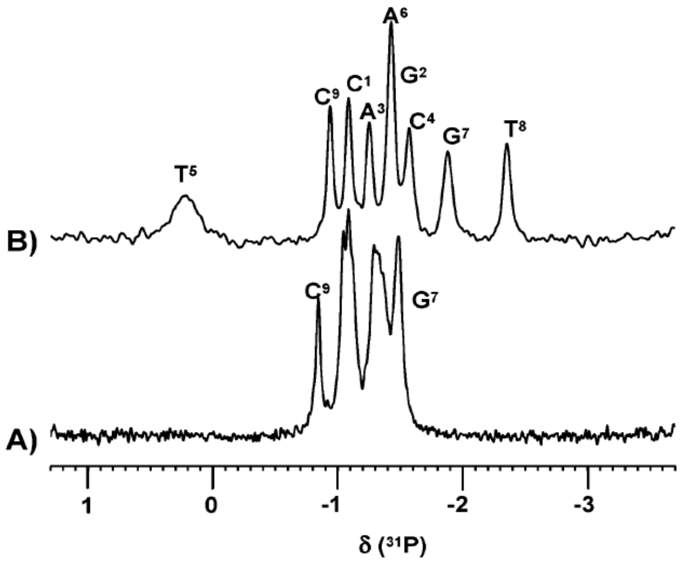


Figure 4.43 ^{31}P-{^{1}H} NMR spectra of d(CGACTAGTCG)$_2$ (A) and of the complex between d(CGACTAGTCG)$_2$ and thiazotropsin A (B) acquired at 9.4 T. The effect of AIK18-51 binding to the DNA duplex was clear from the dispersion of signals that occurred for the complex, indicative of DNA backbone alteration.[38]

The ligand induced widening of the minor groove and the bending of the helix axis towards the major groove generated a compressed major groove with a narrow and deep cleft (Figure 4.44). Average base pair axis parameters showed significant inclination (η) throughout the ligand binding site and particularly at the central base pairs in both complexes, which contributes to the significant bend in the DNA helix. Additionally, major perturbations in the DNA base-pair propeller twist were also observed upon ligand binding. These DNA structural alterations are consistent with those observed in the recently reported crystal structure of a cyclic polyamide minor groove binder ligand complexed to the DNA sequence 5′-CCAGTACTGG-3′[44] (Figure 4.45). Comparison of the DNA structural alterations induced by the DNA-binding transcription factors androgen receptor (AR) and glucocorticoid receptor (GR) in the major groove (Figure 4.46) with those induced by the cyclic polyamide binding in the minor groove revealed that the latter disrupts the DNA structure in such a way that the transcription factors are unable to recognise the major grooves. Such structural perturbations in the DNA helix, especially in relation to the groove

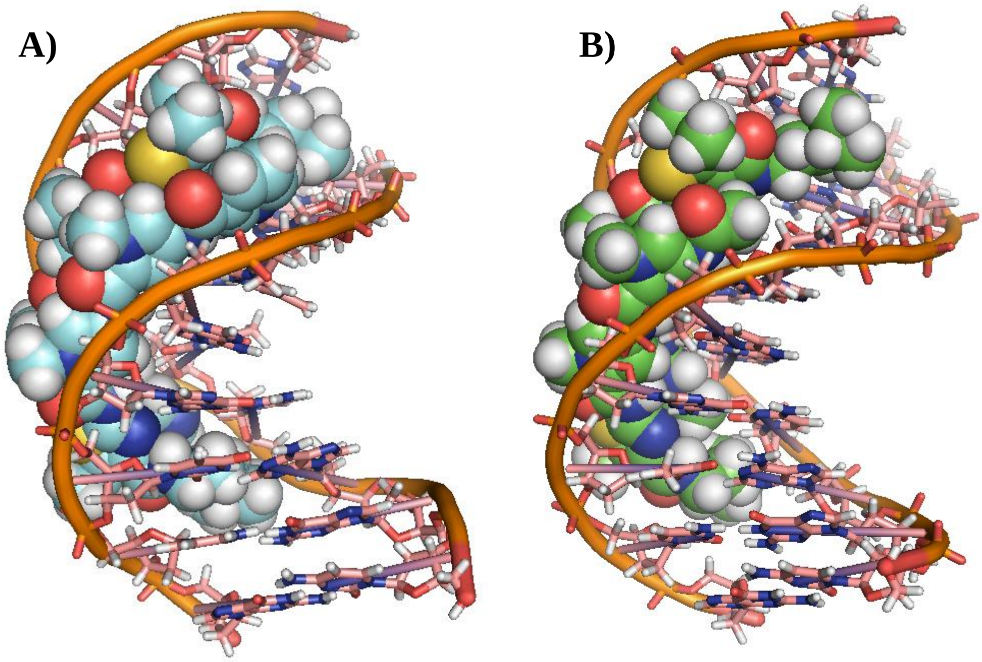


Figure 4.44 Representation of the NMR refined solution structure of the complexes **A)** between AIK18-51 (CPK) and d(CGACTAGTCG)$_2$ (stick and tubes); and **B)** between thiazotropsin B and d(CGACGCGTCG)$_2$ showing the ligand-induced bending of the DNA helix toward the major groove resulting in major groove compression.

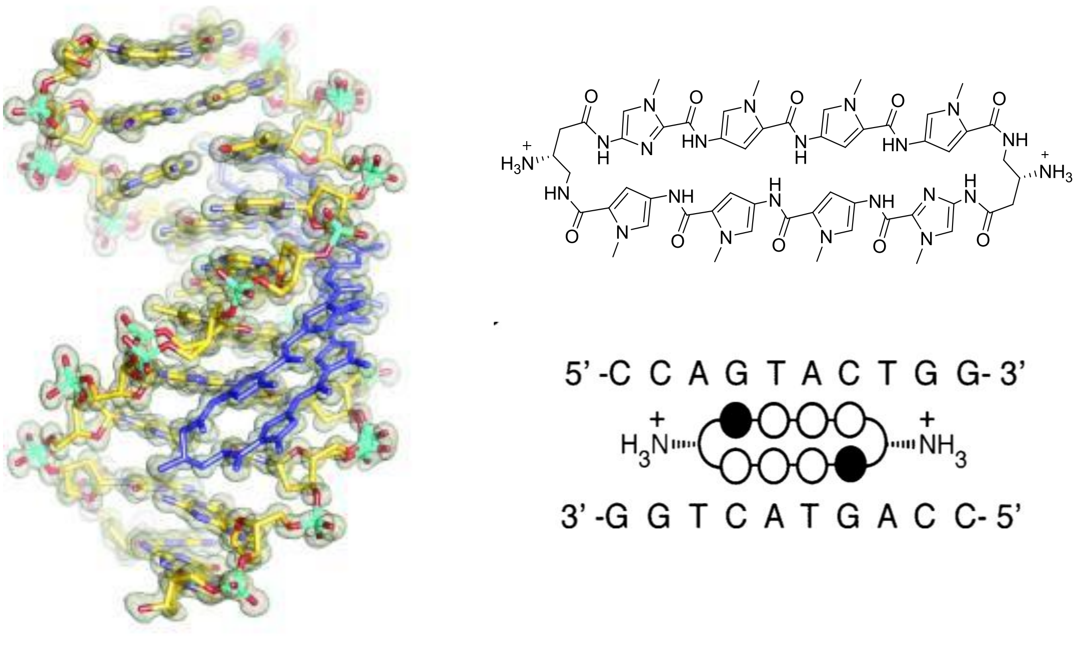


Figure 4.45 Crystal structure of the 8 ring cyclic polyamide that targets the 5′-AGTACT-3′ sequence.[44]

dimensions, are believed to be responsible for the disruption of transcription factor−DNA interfaces via allosteric modulation. The wide, shallow surface of the major groove in B-form DNA allows the sequence-specific DNA-binding proteins to interact with the major groove, because it exposes more base pair functional groups and its width can accommodate protein α-helical domains or β-sheets. MGB-induced structural perturbation of the DNA major grooves can prevent transcription factors from recognising their DNA targets, and such ligands can therefore be used to regulate gene expression.

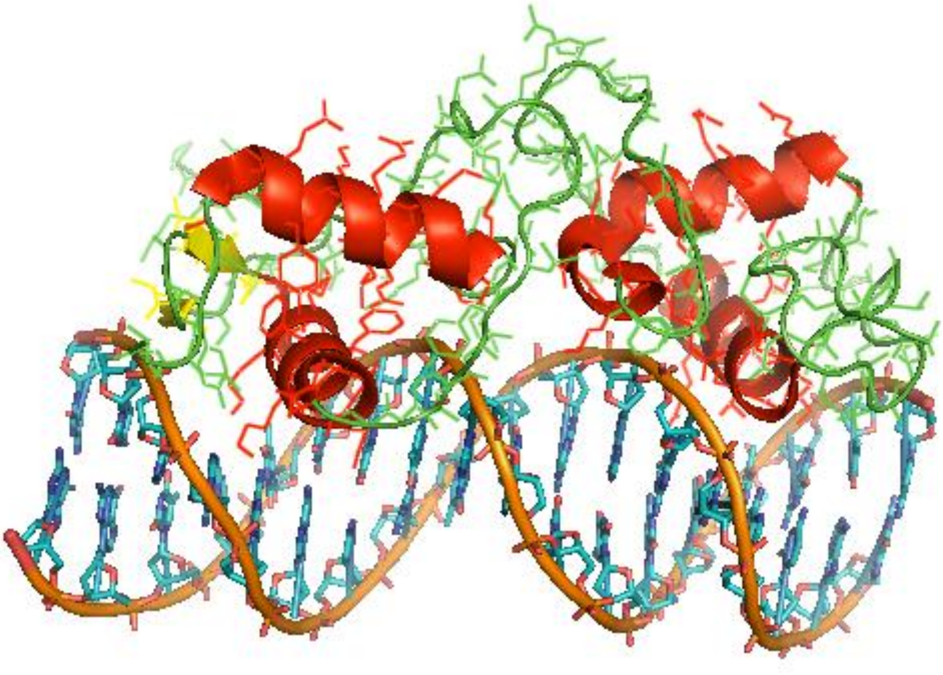


Figure 4.46 Molecular modeling representation of the crystal structure for the androgen receptor DNA-binding domain bound to a direct repeat response element [d(5′-CC$\underline{\text{AGAACA}}TCA\underline{\text{AGAACA}}$G-3′)$_2$] showing the importance of major groove width in protein–DNA interactions.[161]

4.7 Conclusions and future work

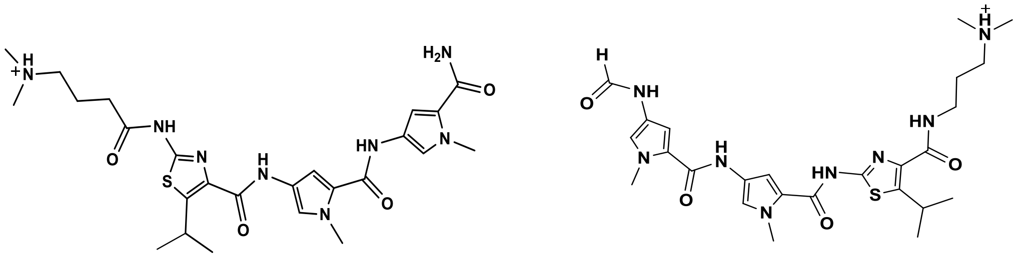
Our NMR studies have provided clear evidence that the lexitropsins investigated in this study bind in a sequence specific manner to the DNA minor groove in a 2:1 ratio through side-by-side binding in a head-to-tail fashion. The shape of the molecule and its ability to undergo induced-fit recognition and form a series of direct hydrogen bonds to the floor of the DNA minor groove allows these ligands to target specific DNA sequences. The amide NH′s form hydrogen bonds with the purine N3 and pyrimidine O2 lone pairs and the thiazole and imidazole rings enhance specificity for the exocyclic amine of guanine through hydrogen bond formation. A comparison of the ligand-free and ligand-bound forms of the DNA in solution indicates that the overall shape of the DNA is retained but ligand binding widened the minor groove and induced a significant bend in the axis of the DNA helix towards the major groove, thereby generating a compressed major groove with a narrow deep cleft. These alterations in the DNA structure form the molecular basis for the allosteric inhibition of protein-DNA interactions by small molecules.

Isothermal titration calorimetry (ITC) studies generated a complete thermodynamic profile for lexitropsin-DNA associations that includes binding affinity (K), stoichiometry (N), enthalpy (ΔH), entropy (ΔS) and free energy of binding (ΔG) for the interaction. ITC studies revealed that the lexitropsin-DNA interactions are mainly enthalpically driven through hydrogen bonding and van der Waals forces. The unfavourable entropies associated with these interactions are indicative of ″induced fit″ binding and conformational changes in either of the reactants. These findings are consistent with our NMR results. Comparing the thermodynamic binding characteristics of closely related ligand structures to specific binding sites helped to establish how modifications in the structure influence binding affinity. The results of these studies can be used to improve the binding selectivity and the physical properties of these ligands without sacrificing the binding affinity. For instance, lipophilic functional groups can be introduced at specific positions of the heterocyclic units to modify essential physicochemical properties without affecting binding capability, and in some cases, improve the binding affinity.

The ring slippage of the side-by-side minor groove binders investigated here has enabled these ligands to span six base pair DNA sequences, in contrast to the expectation of four base pairs based on the size of the ligand. The importance of this in terms of gene targeting is that such ligands with small molecular weight and enhanced lipophilicity can be used to disrupt the binding of transcription factors to the response element of the target gene that is composed of six base pair sequences. There are many examples of transcription factors that recognise hormone response elements composed of a six base pair DNA sequence, such as the androgen receptor (AR), glucocorticoid receptors (GR), mineralocorticoid receptor (MR) and progesterone receptor (PR). All these receptors recognise the same DNA response element,[162] which is organized as repeats of the consensus DNA sequence 5′-WGWWCW-3 (W is A or T), with a three nucleotide spacer. Malfunction of these transcription factors can lead to dysregulated gene expression, which is observed in many human cancers such as prostate cancer. The prostate specific antigen (PSA) gene which is commonly used as a marker for the diagnosis of prostate cancer is regulated through the interaction of AR with the androgen response element (ARE) of the PSA gene.[162] The AR-ARE interaction can be disrupted via small molecules (e.g. MGBs) targeting the ARE. Blocking the AR from binding to the ARE is expected to reduce the transcription of the PSA gene. Recent data from Dervan′s laboratory have shown that a cyclic polyamide minor groove binder, comprised of 8 imidazole/pyrrole rings (Figure 4.45), which binds to the ARE (5′-WGWWCW-3′) is able to regulate the expression of AR target genes in cell culture studies.[43] Although it was concluded from this study that the relatively large cyclic polyamide is cell permeable, absorption studies carried out on this compound showed low Caco-2 permeability, suggesting that it may not be orally available.[43]

The ability of our compounds to recognise a DNA sequence very similar to that of the ARE (for instance, thiazotropsin A can recognise the 5′-WCWWGW-3′ sequence, where W is A or T), is encouraging and development of these ligands is therefore of value in order to target the ARE sequence. The small molecular weight of the lexitropsins which are composed of only three aromatic rings and their enhanced lipophilicity via the isopropyl thiazole system could offer them additional beneficial physical properties compared with the large molecular weight cyclic polyamide ligands. Furthermore, additional lipophilic functional groups could be introduced without sacrificing binding affinity as these substituents point out from the minor grooves.

Different ligands can be designed in order to inhibit AR-ARE interactions based on the structure of our synthetic lexitropsins. One ligand structure is described here by way of example to target the 5′-WGWWCW-3′ sequence. The proposed structure is an analogue of thiazotropsin A (thiazotropsin C) (Figure 4.47).



Thiazotropsin C

Thiazotropsin A

Figure 4.47 Structure of thiazotropsin A and its proposed analogue, thiazotropsin C.

In the proposed ligand structure, the positions of both the DMAP tail and the isopropyl thiazole in thiazotropsin A are changed: the DMAP group is attached to the formyl head and the isopropyl thiazole replaces the pyrrole ring located towards the *N*-terminus. Studying the DNA binding behavior of this compound may reveal further factors that dictate the preferred orientation of 2:1 anti-parallel dimer. In all of the studies we have conducted, the ligands bind to the minor groove as dimers in an anti-parallel side-by-side fashion and only a single orientation of the ligand has been observed: the positively charged DMAP groups are orientated towards the 3′-ends of the DNA strands. In theory, if thiazotropsin C can bind to the minor groove as a dimer, there are two possible orientations that can be adopted by the ligand with respect to the DNA helix (Figure 4.48). In the first orientation (Figure 4.48.A), the ligand recognizes the 5′-WGWWCW-3′ sequence as a dimer, with its positively charged DMAP group lying toward the 5′-end and the carbonyl group of each amide (coloured red) linking the aromatic rings is closer to the the 5′-end than its partnering NH. If the biophysical experiments confirm the binding of thiazotropsin C to the 5′-WGWWCW-3′ sequence, such results would suggest that the factor that dictates the orientation of lexitropsin ligands in the minor groove is the orientation of the amide links with respects to the 5′-3′-ends. In this case, the amide link always prefers an orientation where its carbonyl group is closer to the the 5′-end than the partnering NH regardless of the DMAP group position in the ligand structure whether it is attached to the head or the tail position.

In the second orientation (Figure 4.48.B), the ligand recognises the 5′-WCWWGW-3′ sequence as a dimer, with its positively charged DMAP group lying towards the 3′-end and the carbonyl group of each amide (coloured blue) linking the aromatic rings is closer to the the 3′-end than the partnering NH. If the biophysical experiments confirm the binding of thiazotropsin C to the 5′-WCWWGW-3 sequence, such results would suggest that the orientation of lexitropsins in the minor groove is dictated by the positively charged DMAP tail of the ligand. In this case, the DMAP group always prefers to point towards the 3′-end regardless of its position in the ligand structure either at the head or the tail position, and regardless of the two possible orientations of the amide links (red/ blue coloured amides) with respect to the 5′-3′-ends.

It is expected that the ligand would not exhibit both orientations, because lexitropsins usually prefer to adopt a single orientation. If no binding or very weak binding is observed with this ligand, this presumably indicates a significant steric clash between the DMAP group and the backbone of DNA. The possibility that the ligand dimer slides between six and five base pairs sequence cannot be ruled out. However, footprinting studies performed on thiazotropsin A[147] and thiazotropsin B[129] have shown that these ligand are selective for the six base pair sequences 5′-ACTAGT-3′ and 5′-ACGCGT-3′ respectively. The outcome of such a study would be of great importance in terms of both the development of lexitropsins to target specific genes and to gain more insight into the molecular basis of minor groove recognition by lexitropsins. Finally, a proposed synthetic pathway for thiazotropsin C is summarized in Figure 4.49.

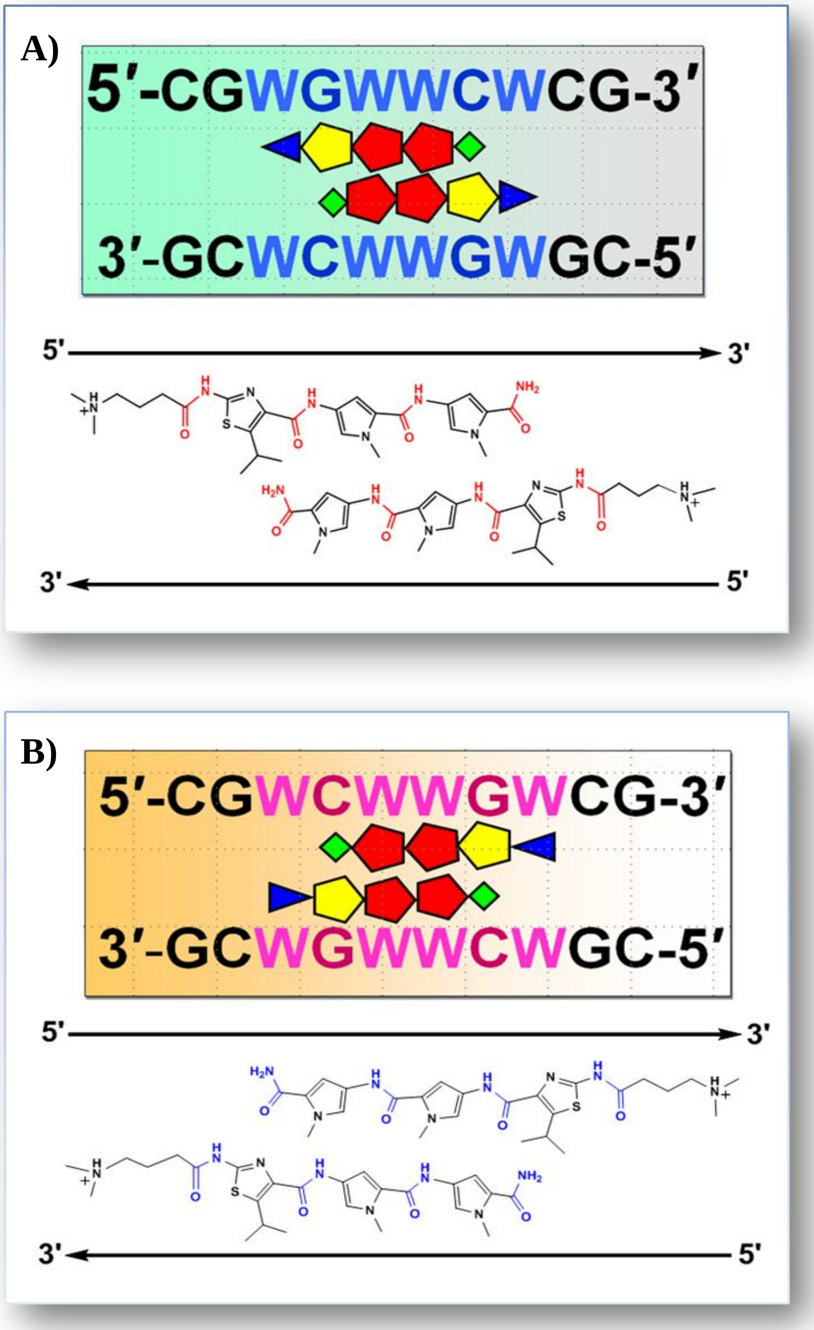


Figure 4.48 Representation of the possible binding orientations that may be adopted by thiazotropsin C with the DNA sequence 5′-WGWWCW-3′ **A)** The ligand aligned in a 3′-5′ direction. **B)** The ligand aligned in a 5′-3′ direction.

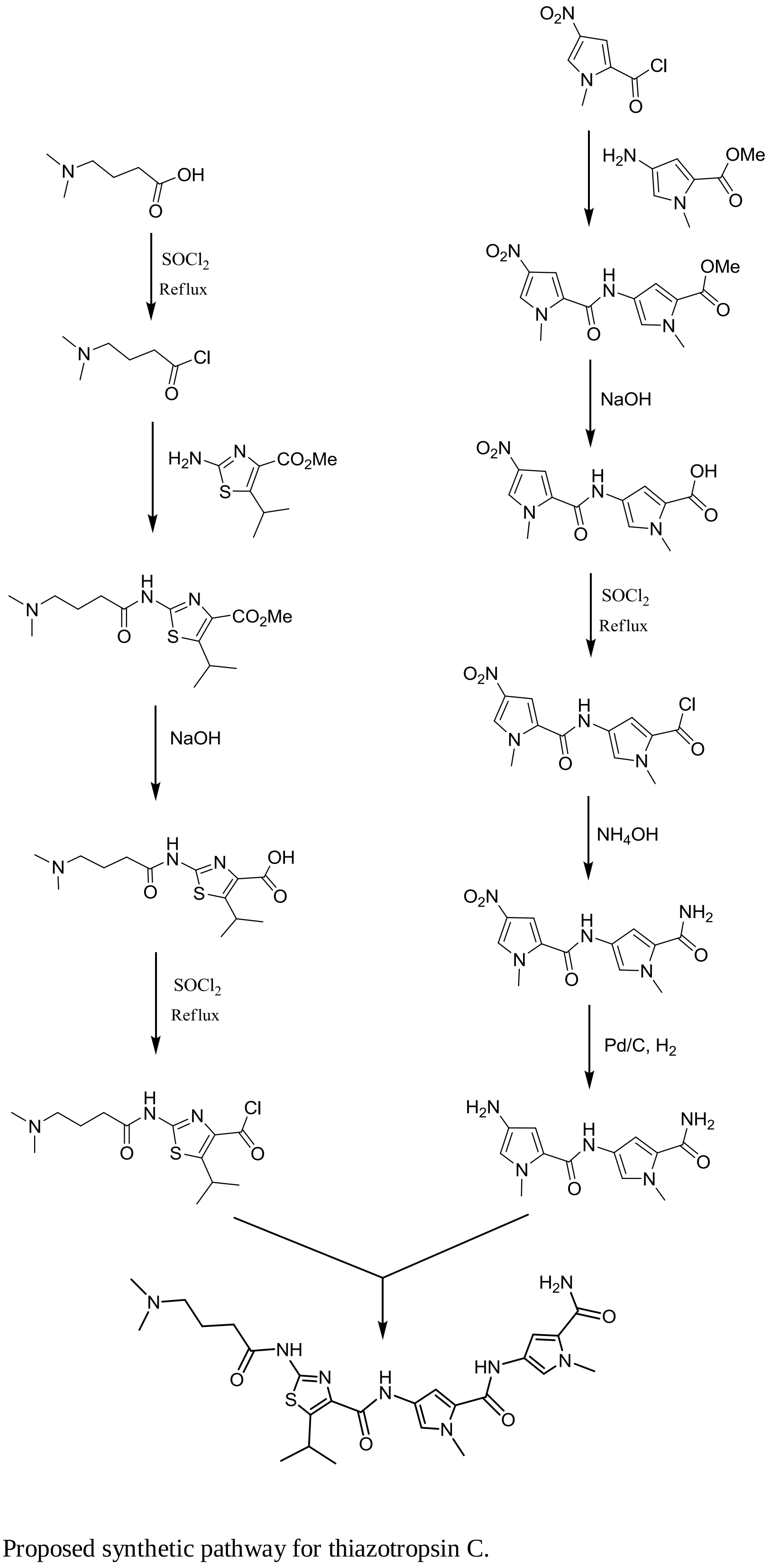


Figure 4.49 Proposed synthetic pathway for thiazotropsin C.

5 CHAPTER 5: Synthesis of alkene-linked MGBs

5.1 Introduction

ITC studies have revealed that hydrogen bonding and/or van der Waals interactions are the main forces that drive the binding of the lexitropsins synthesised at Strathclyde. In order to investigate the role of the traditional amide links of these ligands in the binding to the DNA minor grooves, and to evaluate their importance in forming hydrogen bonds with DNA bases, analogous compounds where an alkene group replaced the amides were proposed , such as examples **1** and **2** (Figure 5.1).

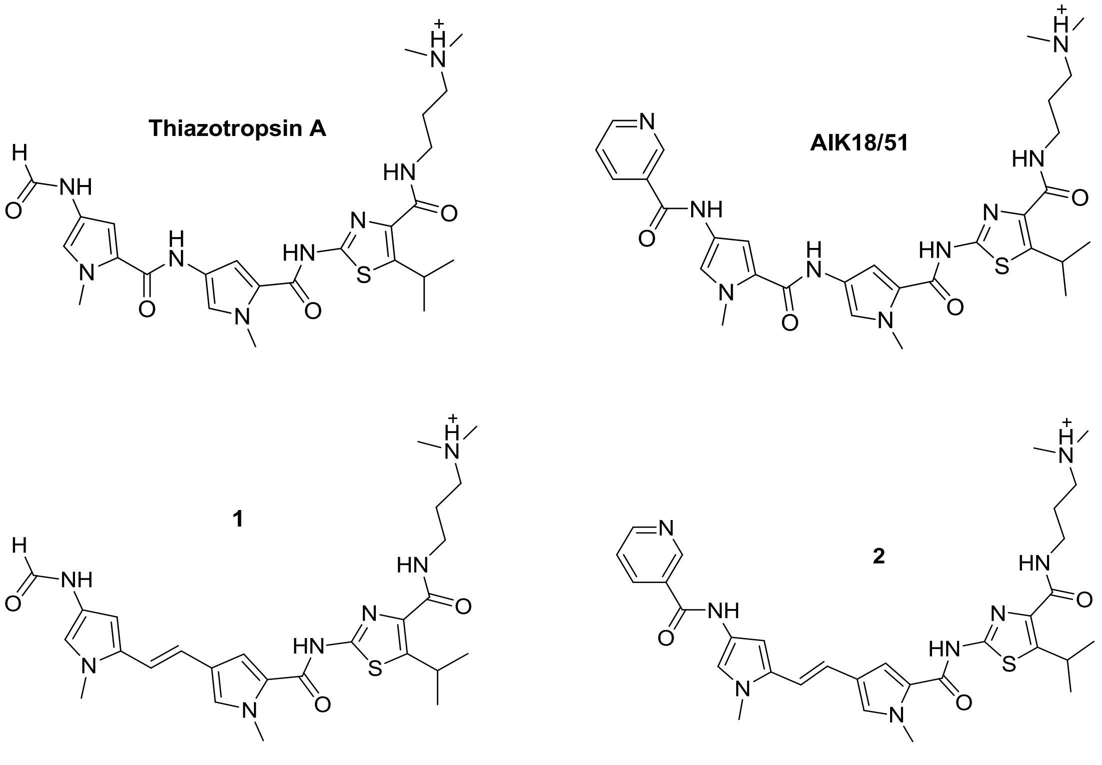


Figure 5.1 Examples of alkene-linked MGBs: 1 is an analogue of thiazotropsin A, and 2 is an analogue of AIK18-51.

Assessing the binding of the alkene-linked MGBs and studying the thermodynamic characteristics of their interaction with DNA would enable a direct evaluation of an amide functional group compared with an alkene, and reveal the thermodynamic role that the amide plays in the binding process.

If the alkene linked MGBs bind to DNA, they may have additional beneficial physical properties such as enhanced lipophilicity. However, this variation may also lead to differences in the binding pattern or affinity by replacing a hydrogen bonding group with one which does not.

5.2 Proposed pathway for synthesis of alkene-linked MGBs

From the retrosynthetic analysis shown below (Figure 5.2), these compounds can be prepared using a convergent synthesis through coupling the standard thiazole tail **3** prepared from **7** and **8**, with the alkene-linked pyrrole dimer **4** described in the literature,[163] prepared from **5** and **6** using a Wadsworth-Emmons reaction. To generate the final product **1** would require the selective reduction of the nitro group followed by formylation to cap the amino head group.

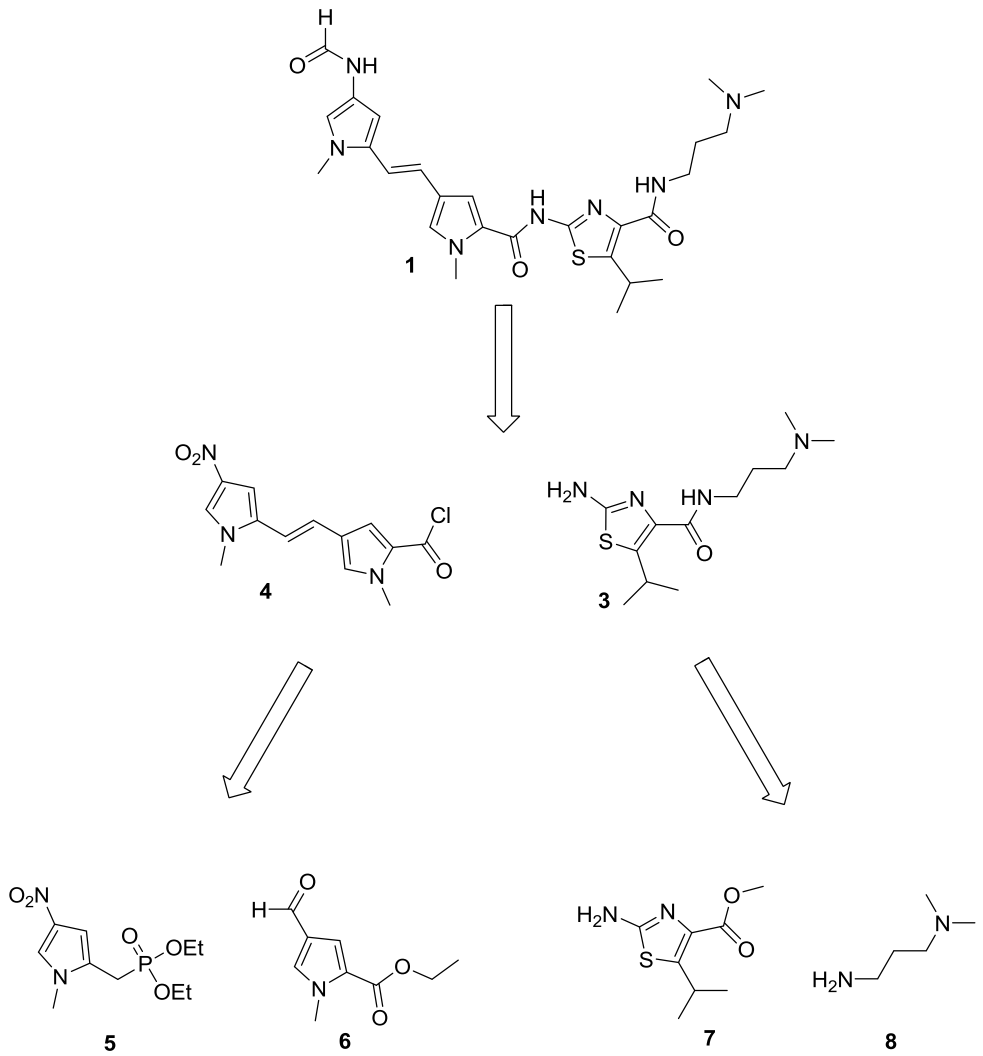
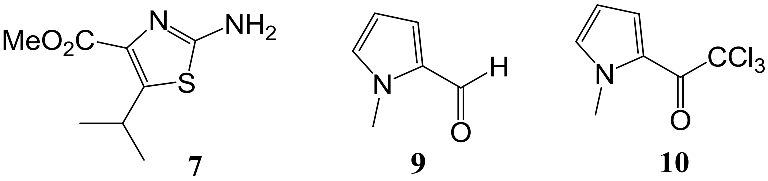


Figure 5.2 Retrosynthetic analysis of the alkene-linked analogue of thiazotropsin A.

5.3 *Results and discussion*

5.3.1 Synthesis of the alkene-linked pyrrole dimer

4-Nitro-*N*-methylpyrrolealdehyde **11** (Figure 5.4) was prepared by the nitration of *N*-methylpyrrole-2-aldehyde **9** with nitric acid in acetic anhydride, in a relatively low yield (23 %). This low yield of nitration at position 4 may be explained by the additional nitrations at positions 3 or 5 on the pyrrole ring. Fortunately, it was possible to separate the desired product by recrystallisation at -40 °C. 2-Trichloro acetyl-*N*-methylpyrrole **10**, was also prepared in a good yield (70 %) by the reaction of trichloroacetylchloride and *N*-methylpyrrole.[164]



7 **9** **10**

Figure 5.3 Structure of some monomers used in the synthesis of MGBs.

The alkene linked pyrole dimer **16** (ethyl-*N*-methyl-4- [2-(*N*-methyl-4-nitropyrrol-2-yl)- vinyl]pyrrole-2- carboxylate) was prepared in eight steps from the precursors **9** and **14** (Figure 5.4). The final step in this synthesis is the Horner-Wadsworth-Emmons reaction between **5** and **6**. This route which has been described in the literature[163] was used successfully to generate the alkene-linked pyrrole **16** (Figure 5.4).

In our synthesis, 4-nitro-*N*-methylpyrrolealdehyde **11** was reduced by sodium borohydride to give **12** in an excellent yield (92%). This was followed by refluxing in thionyl chloride to give **13** quantitatively. The phosphonate product **5** was obtained by refluxing **13** in neat triethyl phosphate, which gave an excellent yield (88 %). The aldehyde **6** was obtained in three steps; *N*-methyl pyrrole **14** was reacted with trichloroacetic chloride to give **10**, which was then reacted with sodium methoxide to give the ester **15** in excellent yield. The desired aldehyde was then obtained using dichloromethylmethyl ether in a Freidel-Crafts acylation which proceeded smoothly and in a good yield to give **6.**

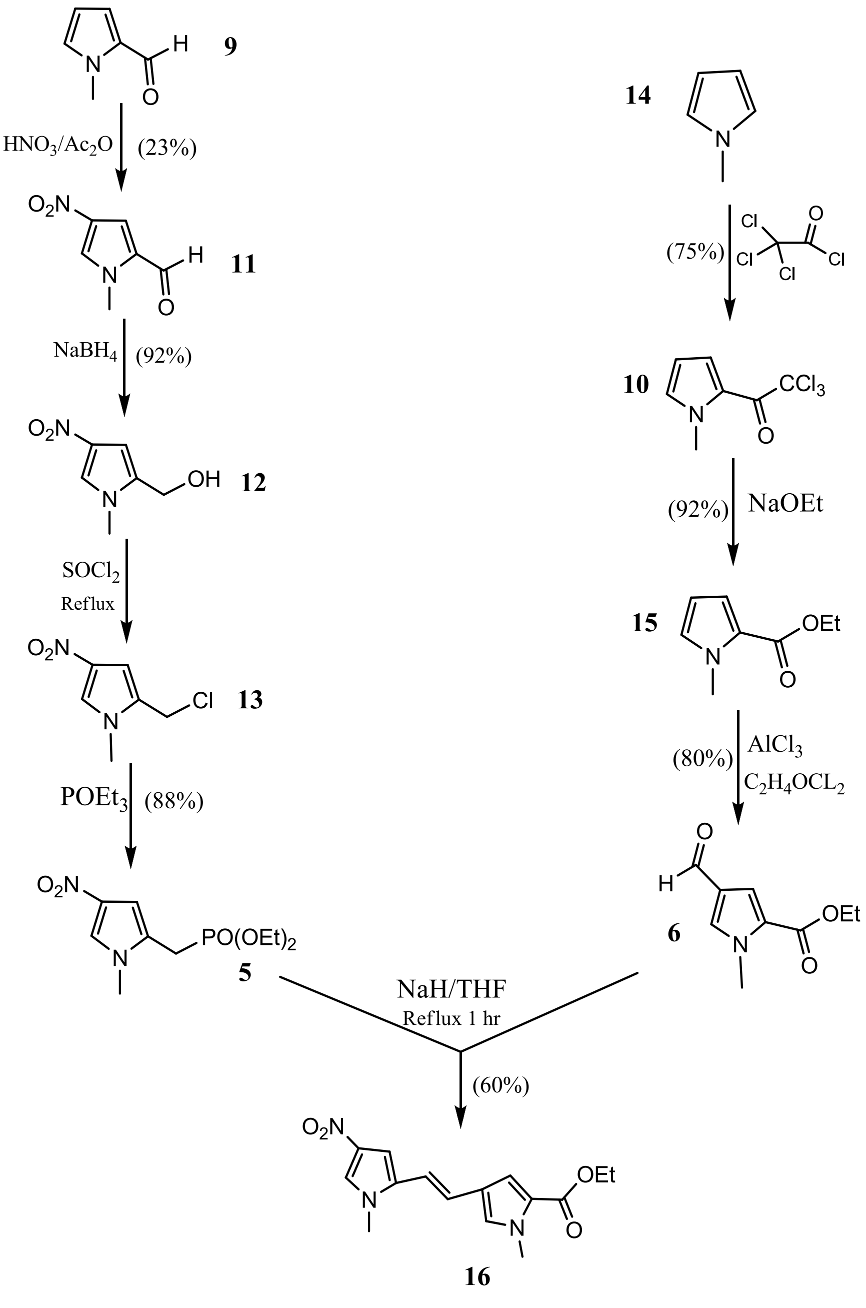


Figure 5.4 Synthesis of the alkene-linked minor groove binder precursor.

The Wadsworth -Horner -Emmons reaction gave **16** in a moderate yield (60%) by reacting **5** with **6** under reflux using sodium hydride (Figure 5.5). The mechanism of this reaction is outlined in Figure 5.6. The reaction proceeds via deprotonation of the phosphonate **5** using sodium hydride to generate the phosphonate carbanion **18** which is followed by nucleophilic attack of the carbanion on the aldehyde to give the α-hydroxy phosphonate intermediates **19, 20,** which is the rate determining step. The presence of a hydrogen atom (highlighted in red) α to the phosphonate allows the reaction to proceed via the intermediates **19-22**. . This reaction also requires the presence of an electron withdrawing group α to the phosphonate for the final elimination to occur. In the absence of such an electron-withdrawing group, the final product would be the α-hydroxyphosphonate **19** and **20**.[165] In our case, the reaction proceeded smoothly and gave a 60 % yield. This suggests that the electron withdrawing characteristics of the nitro substituted *N*-methyl pyrrole (highlighted in blue) was enough for the elimination to occur. The ratio of E/Z isomers (**16a**/**16b**) is dependent on the stereochemical outcome of the initial carbanion addition and the ability of the intermediates to equilibrate. Furthermore, the formation of the *Z* isomer is hindered by a steric clash between the protons on C-3 and/or between the nitro and ethyl carbxoylate substituent on C-4 and C-2, respectively of each pyrrole heterocycle, which drive the reaction toward sterically favoured *E*-isomer. A *J* coupling of 16 Hz for the alkene protons (δ_H = 6.52 and 6.79 ppm, *J* = 16.0 Hz) in the NMR spectrum confirmed the formation of the *E*-isomer.

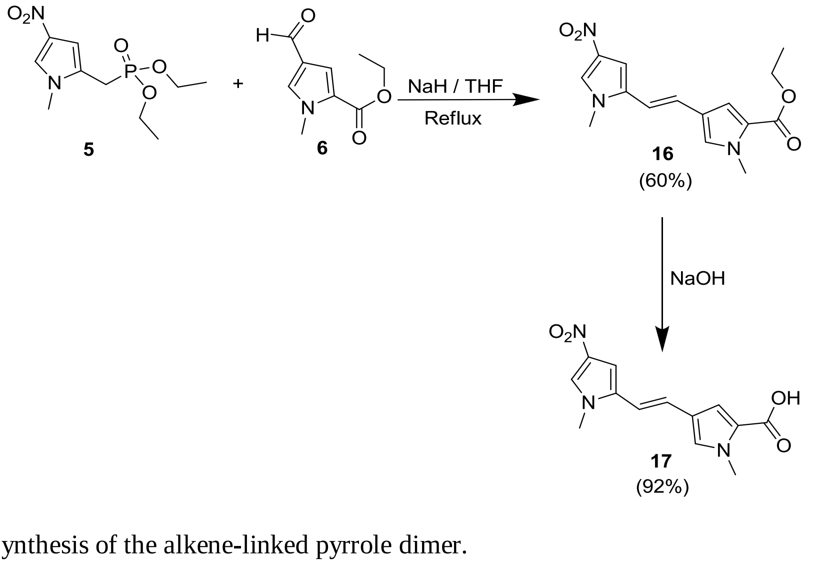


Figure 5.5 Synthesis of the alkene-linked pyrrole dimer.

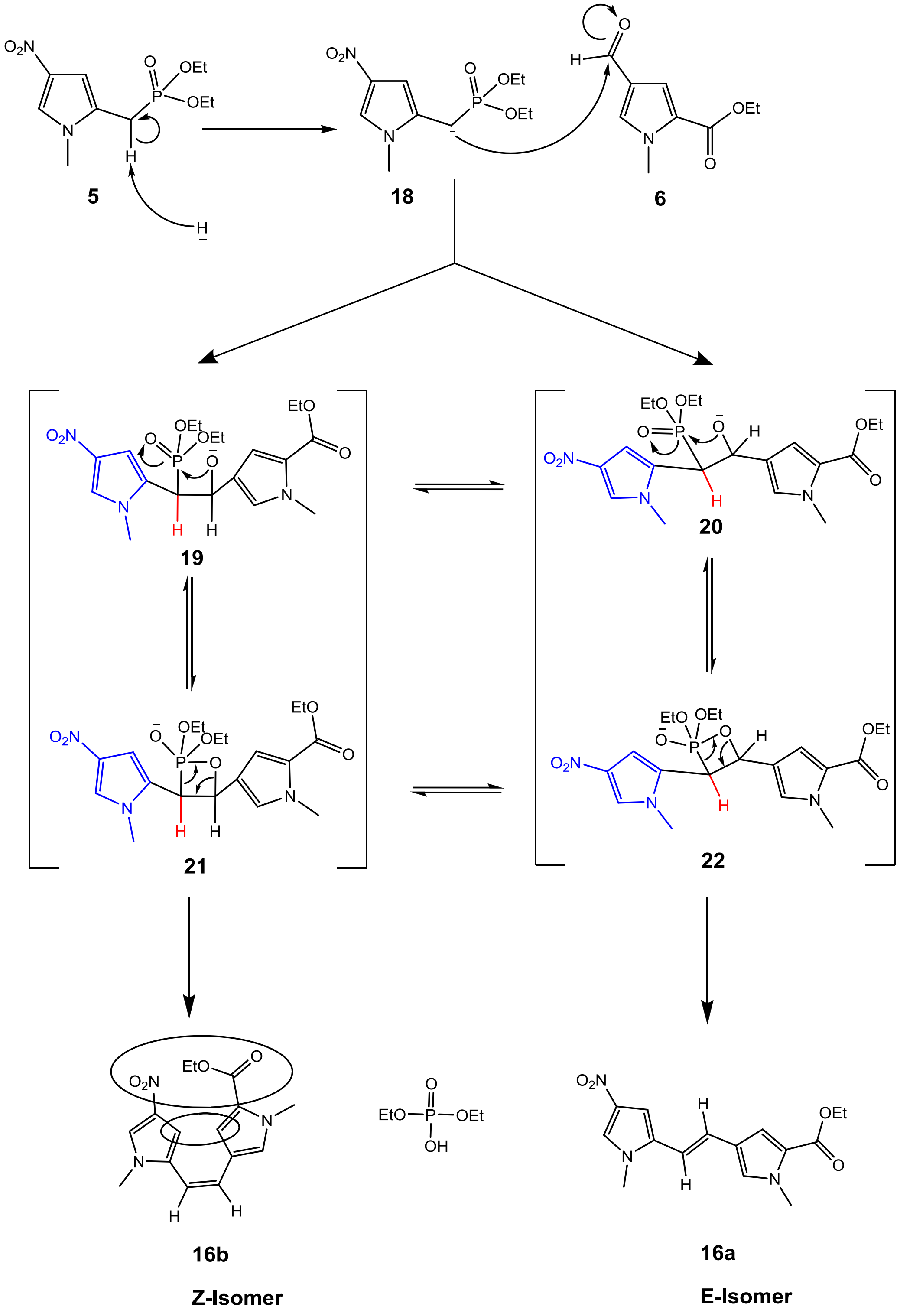
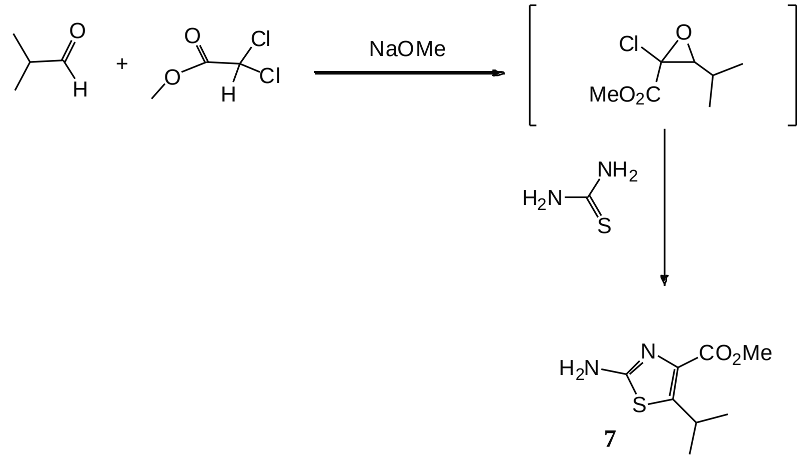


Figure 5.6 The mechanism of the Wadsworth-Horner-Emmons reaction.

The Wadsworth-Horner-Emmons reaction is a modified version of the Wittig reaction using phosphonate-stabilized carbanions.[166] This reaction has several practical advantages over the Wittig reaction. The phosphonate-stabilized carbanions are more nucleophilic and more basic in contrast to phosphonium ylides used in the Wittig reaction. The dialkylphosphate salt by-products of the reaction are water soluble and easily removed during the work-up by aqueous extraction. Moreover, the time of reaction is short (1 hour) and is, generally, high yielding.

Methyl-2-amino-5-isopropyl-1,3-thiazole-4-carboxylate **7** was obtained using a previously described synthesis[167] as the free amine and used without further purification (Figure 5.7).



7

Figure 5.7 Synthesis of methyl-2-amino-5-isopropyl-1,3-thiazole-4-carboxylate **7**.[166]

5.3.2 Amide bond formation under anhydrous conditions

Basic hydrolysis of the ester to the corresponding carboxylic acid **17** proceeded in excellent yield (92%) (Figure 5.5). The next stage was to synthesise the trimer **24** by coupling **17** with **7**. The most established way to couple the aromatic amino thiazole **7** with a carboxylic acid is via the acid chloride. The presence of the alkene in **17** limited our options of using the common chlorinating agents such as thionyl chloride because the production of HCl under reflux conditions would chlorinate the alkene. Therefore, coupling agents such as HBTU and T_3P were tried initially to react **17** with **7** (Figure 5.8), but were not successful.

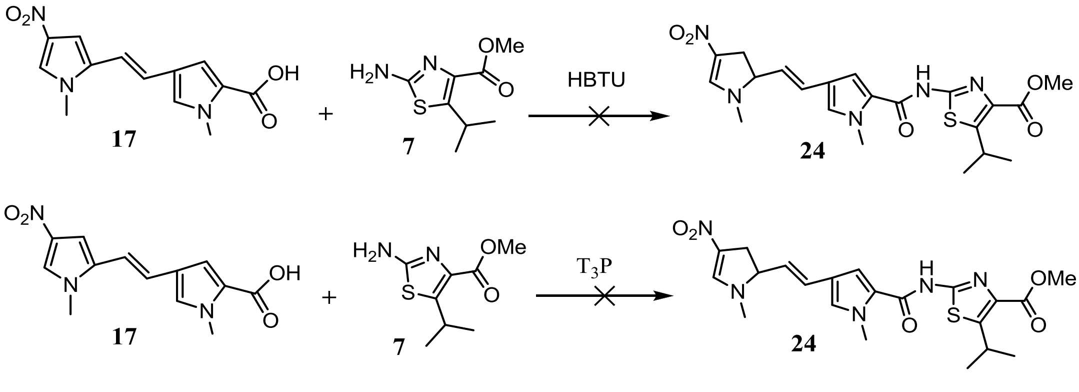


Figure 5.8 Attempted coupling reactions between the dimer **17** and the nonomer **7**.

Despite these coupling agents being successfully deployed previously in coupling aromatic amines (e.g. the amino substituted pyrroles) and aliphatic amines (dimethylaminopropylamine, DMAP) with carboxylic acids,[168] their failure in this case is likely to be due to the electron withdrawing characteristics of the thiazole ring, which reduces the nucleophilicity of the exocyclic amine.

Ghosez *et al.* have used the mild method of employing tetramethyl-alpha-chloroenamine[169] to avoid formation of HCl, which is extremely useful in the presence of acid sensitive groups (Figure 5.9).

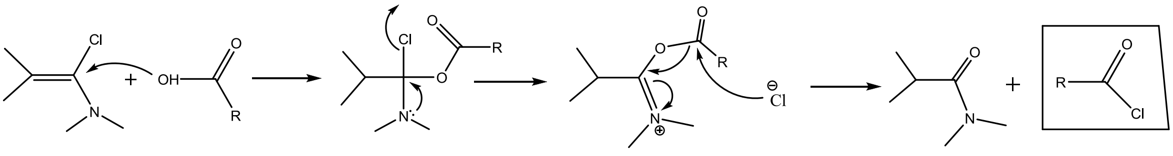


Figure 5.9 The mechanism of acid chloride formation using Ghosez reagent.

This reagent was used successfully in transforming the carboxylic acid **17** via the corresponding acid chloride **23** (not isolated) and the amine **7** to the desired trimer **24** in good yield (70 %) (Figure 5.10).

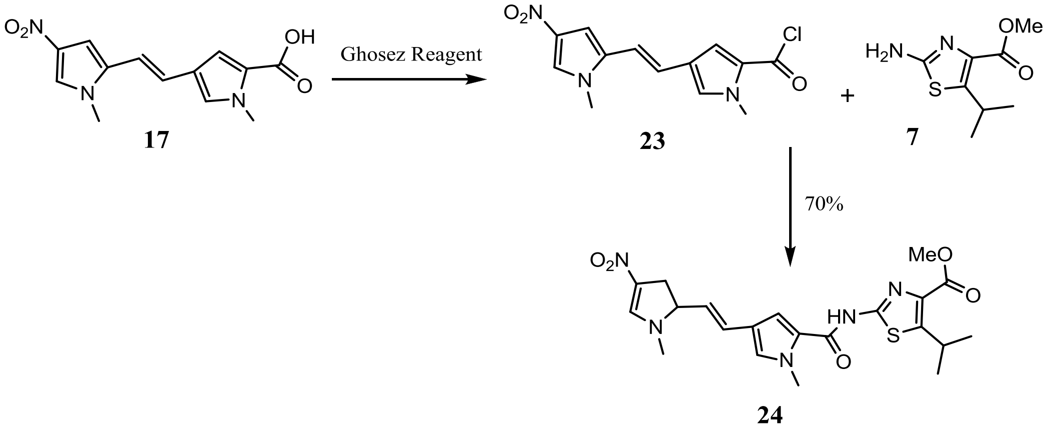


Figure 5.10 Coupling between the dimer **17** and the nonomer **7** using Ghosez reagent.

Coupling of the trimer **24** to produce **25** by heating at moderate temperature in dimethylaminopropylamine (used as both reagent and solvent) proceeded in excellent yield (Figure 2.3.6).

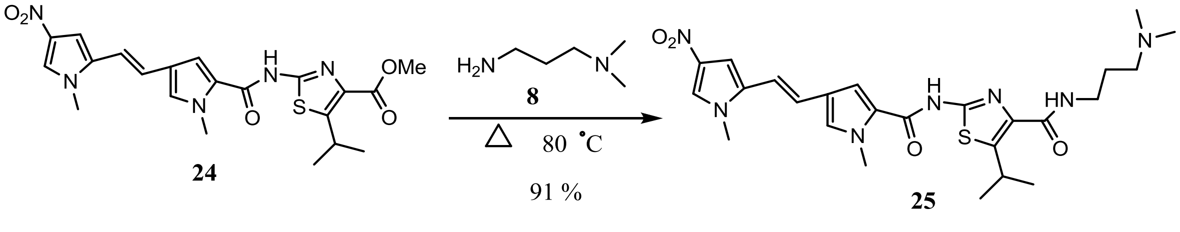


Figure 5.11 Coupling between the trimer **24** and dimethylaminopropyl amine.

5.3.3 Selective reduction of the aromatic nitro group in the presence of an alkene

To obtain target molecules **1** and **2** required the initial reduction of the nitro group of **25** followed by reaction with the formyl ester or nicotinoyl chloride. An alkene in **25** meant that selective reduction of the nitro group was a critical step. Classical catalytic hydrogenation methods using Pd-C/$H_{2,}$ which had previously been used for the reduction of the nitro group in amide linked MGB synthesis[168] could not be applied in this case because it lacked selectivity. The reduction of nitro groups in the presence of alkenes was reported by Anthony *et al,*[163] which involved sodium borohydride in water as a source of hydrogen with palladium on carbon as the catalyst over a short time (15 min). The selectivity of this reaction is based on relative rates of reduction; the nitro group is reduced around 1000 times faster than the alkene, which allows the desired product to be obtained with only small amounts of the alkene being reduced. Using sodium borohydride dissolved in water allows the reduction to be carried out in a short time to avoid reducing the double bond (Figure 5.12).

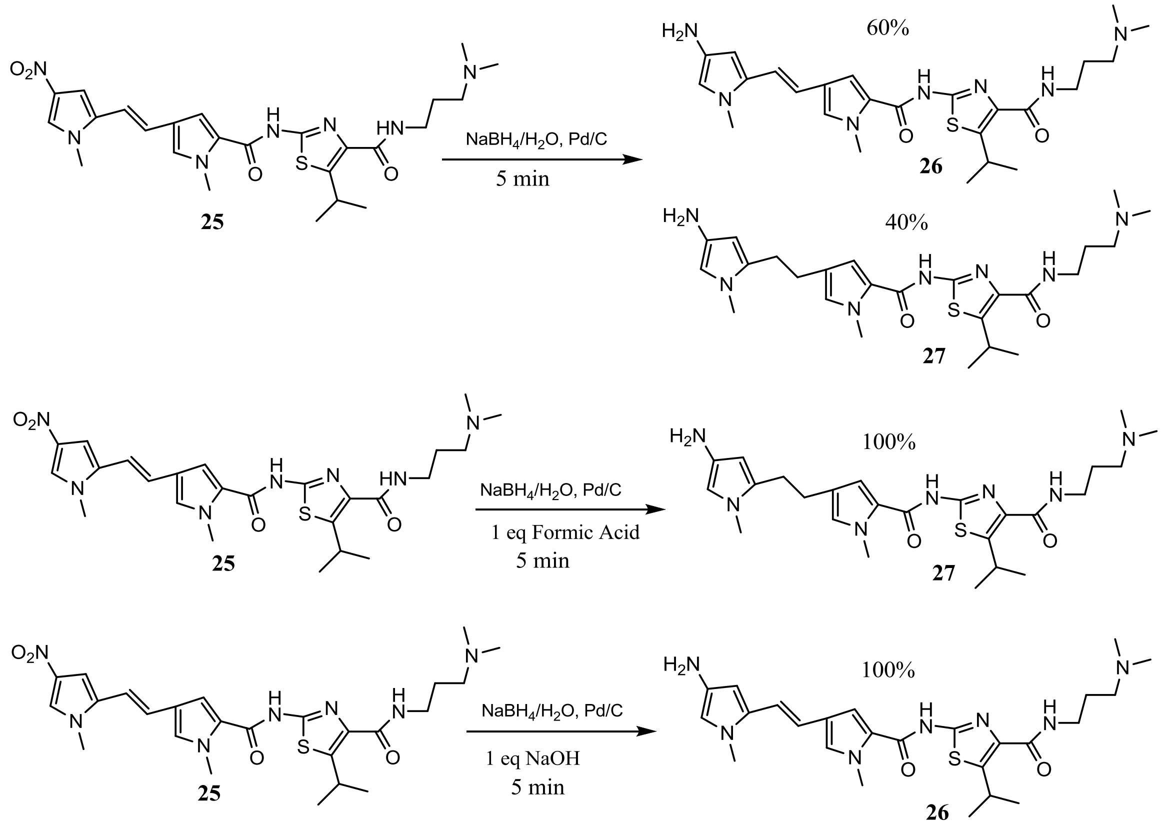
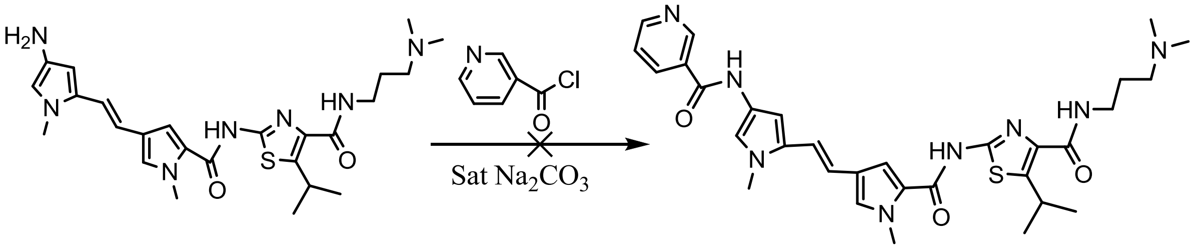


Figure 5.12 Attempted selective reduction of the nitro group in the presence of an alkene.

Using this method, 60 % of the desired product **26** was formed, but with 40 % of the fully reduced alkane **27** (Figure 5.12). To neutralise the sodium hydroxide generated from the reaction of sodium borohydride with water (which was thought may cause decomposition of the product later on), one equivalent of formic acid was added to the reaction, but this led to complete reduction of both the nitro group and alkene, which suggested acidic conditions may catalyse the reduction of the alkene and inferred that the pH of the reaction may play a role in the selectivity of these reactions. This was confirmed when the reaction was performed at pH 11.5 by the addition of 1-2 equivalents of sodium hydroxide; the nitro group was completely reduced with the alkene unaffected. Unfortunately, it was not possible to isolate **26** as a free base due to an unforseen stability problem, and the formation of **26** was only confirmed by mass spectrometry.

5.3.4 A water based method for amide bond formation between the acid chloride/ester and the amine

With the reduction method optimised, we attempted to couple nicotinoyl chloride with **26** (Figure 5.13). Because it was not possible to isolate **26** in its free form, the reaction with nicotinoyl chloride was carried out immediately after the reduction step. Reaction of **26** with nicotinoyl chloride in a saturated solution of sodium carbonate was unsuccessful, possibly because the reactants were not stable under the aqueous reaction conditions and decomposition was occurring.

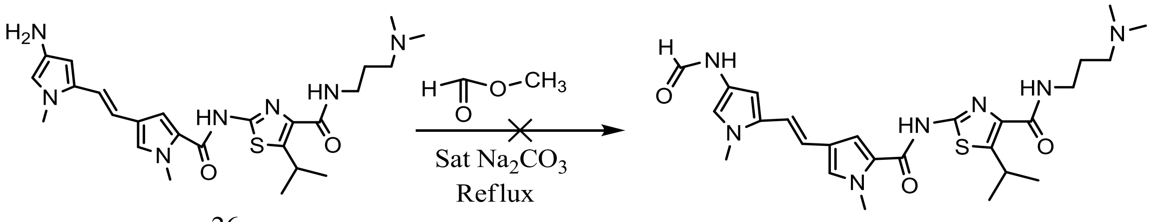


26

2

Figure 5.13

The same procedure was also attempted with methyl formate, but again, the reaction was unsuccessful (Figure 5.14).



26

1

Figure 5.14

During the procedure, the pH of reaction dropped from a pH of 11 to 5, which suggests that the methyl formate was not stable under the aqueous reaction conditions and hydrolyzed to formic acid.

In an attempt to work under anhydrous conditions, **26** was stabilized by converting it to the hydrochloride salt immediately after the reduction of the nitro group and collected by freeze-drying. Unfortunately, it was difficult to dissolve the salt in an organic solvent such as THF or DCM even in the presence of an excess non-nucleophilic base such as *N*-methylmorpholine or Hunig′s base. Using DMF improved the solubility of the salt a little, but the desired product was never obtained despite several attempts. Alternative methods that could be used for amide bond formation under aqueous conditions and different synthetic approaches that avoided decomposition were therefore explored (Figure 5.15).

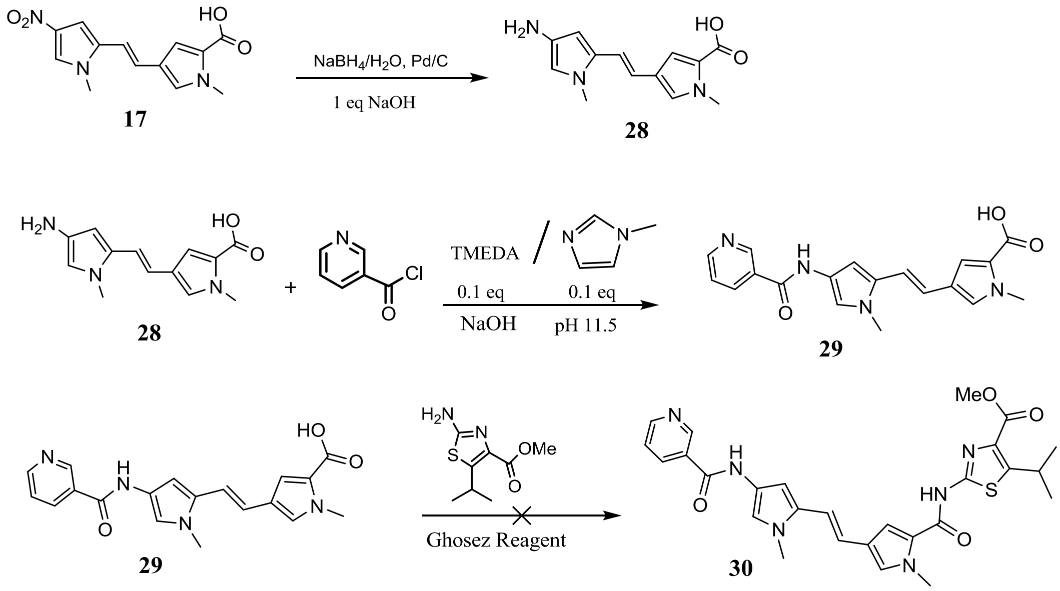


Figure 5.15

The first step was the reduction of **17** using aqueous sodium borohydride (as before) to give **28,** which was then directly coupled with nicotinoyl chloride to give **29** in a good yield (60 %) using the Nakatsuji method.[170] *N*-methylimidazole and TMEDA have a synergistic effect and play two different roles; *N*-methylimidazole forms highly reactive ammonium intermediates with the acid chloride, whilst TMEDA traps the formed HCl and avoids the conversion of the amine into its unreactive HCl salt compound. **29** was obtained as the TFA salt, but was difficult to dissolve in an organic solvent, which made the coupling of **29** with the thiazole **7** unsuccessful due to this solubility problem.

Another water-based method was attempted for the formylation step[171] (Figure 5.16) using a two-phase system of benzene and water (C_6H_6/H_2O). Formyl imidazole was obtained by reacting carbonyl diimidazole (CDI) with formic acid in benzene, which was added dropwise to a solution containing the amine in 1 M Na_2CO_3 under vigorous stirring. Despite several attempts, the reaction did not work and the desired product was never obtained.

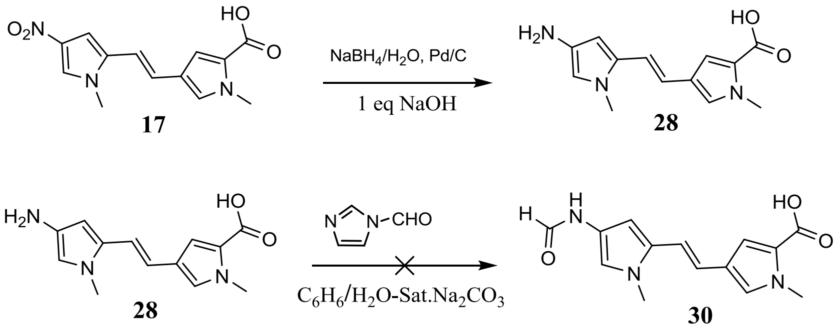


Figure 5.16

5.3.5 Conclusions

The final step of the synthetic pathway was unsuccessful primarily because the reactants were unstable under the aqueous conditions and decomposed during the reaction. The use of anhydrous conditions to carry out these reactions was also unsuccessful because the acidic salt of **26** was insoluble in organic solvents even in the presence of excess non-nucleophilic base such as Hunig′s base. Furthermore, the main problem we faced was the inability to isolate **26** as a free base due to unknown stability problem, which made the next coupling step more difficult, because it limited our choices to perform the reaction under anhydrous conditions. Unfortunately, time limitations did not allow us to address these problems. The precursor to the target compound **1** was therefore used in the assessment of binding (HA10 – see section 3.2.5)

5.4 *Experimental Section*

General

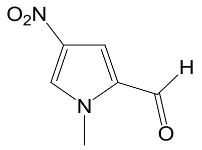
Abbreviations: br, broad; s, singlet; d, doublet; t, triplet; q, quartet; Q, quintet; hept, heptet; exch, echangable; DMAP, dimethylaminopropylamine; HBTU, O-benzotriazol-1-yl-*N*,*N*,*N*,*N*-tetramethyluronium hexaflourophosphate; TMEDA, *N*,*N*,*N*′,*N*′-tetramethylethylenediamine ; DMF, *N*,*N*-dimethylformamide; HPLC, high performance liquid chromatography; HRESI-MS, high-resolution electrospray ionisation mass spectrometry; LRESI-MS, low-resolution electrospray ionisation mass spectrometry Pd/C, palladium on carbon; TFA, triflouroacetic acid. NMR spectra were obtained on either Bruker AMX 400 or Bruker AMX 500 spectrometers. IR spectra were run as KBr disks and liquids as films, using a Perkin Elmer, 1 FT-IR spectrometer. HRESI-MS spectra were obtained on an LTQ orbitrap. Column chromatography was performed using 200-400 mesh silica gels. Melting points were recorded on a Reichert hot stage microscope. HPLC purifications were carried out on a Waters system using a C18 Luna column with a Waters 1525 Binary HPLC pump, and a Waters 2487 dual λ absorbance detector at 254 nm using the gradient eluting system given in Table 5.1.

Table 5.1 HPLC gradient conditions.

Time (min)	A	B	Flow rate (ml/min)
0	90	10	4
28	30	70	4
33	10	90	4
38	90	10	4
40	90	10	0

Mobile phase
A = Water + 0.1% TFA
B = Methanol + 0.1% TFA

Preparation of 1-methyl-4-nitro-1*H*-pyrrole-2-carbaldehyde 11[127]



Nitric acid (70% v/v, 10 ml) was added slowly (by pastuer pipette , 1 ml) to acetic anhydride (40ml) at -40 ºC while stirring, and allowed to stir for a further 20 min. This solution was added slowly (by pastuer pipette, 1 ml) to a solution of 1-methylpyrrole-2-carboxaldehyde **9** (10 g, 91.63 mmol) in acetic anhydride (60 ml) at -40 ºC and allowed to return to 0 ˚C over 2 hours. The solution was cooled to -40 ºC, at which point a precipitate formed. This was filtered and washed with hexane, before being dried under reduced pressure to give the desired product as a yellow solid.

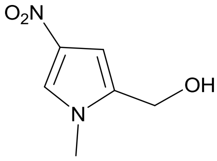
Yield; 3.24 g, 23 %, m.p. = 158-160 ºC, (Lit = 158-160 ºC).[127]

IR (KBr): 3139, 3126, 2958, 1672, 1504, 1404, 1310, 1164, 1099 cm.$^{-1}$

δ_H ($CDCl_3$): 4.04 (3H, s, CH_3), 7.43 (1H, d, Ar-H, *J* = 1.84 Hz), 7.68 (d, Ar-H, *J* = 1.84 Hz) , 9.63 (1H, s, CO(H)).

HRESI-MS (M+H): Found 155.04519 calculated for $C_6H_7O_3$ N_2 155.04512 (0.43771 ppm).

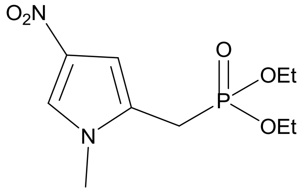
Preparation of (1-methyl-4-nitro-1H-pyrrol-2-yl)methanol 12



1-Methyl-4-nitro-1*H*-pyrrole-2-carbaldehyde **11** (0.90 g, 5.84 mmol) was placed in anhydrous ethanol (50 ml) under nitrogen. Sodium borohydride (0.11 g, 2.9 mmol) was added in small portions over 5 min and the solution allowed to stir for 30 min or until the disappearance of the starting material monitored by TLC. Water (10 ml) was then added slowly to quench the reaction. The solution was then extracted with ethyl acetate (2 x 50 ml). The organic layer was dried over magnesium sulphate, filtered and the solvent removed under reduced pressure to yield (1-methyl-4-nitro-1*H*-pyrrol-2-yl) methanol **12** as a light brown solid.

Yield; 0.838 g, 92 %, m.p. = 89-90 ˚C.
IR (KBr): 3522, 3131, 2933, 2888, 1624, 1520, 1490, 1412, 1337, 1311,1103,1017 cm.$^{-1}$
δ_H ($CDCl_3$): 3.76 (3H, s, N-Me), 4.59 (2H, s, CH_2), 6.67 (1H, d, Ar-H, *J* = 1.4 Hz)), 7.51 (1H, d, Ar-H, *J* = 1.4Hz).
HRESI-MS (M+H): Found 157.06079 calculated for $C_6H_9O_3$ N_2 157.06079 (0.14227 ppm).

Preparation of diethyl (1-methyl-4-nitro-1*H*-pyrrol-2-yl) methylphosphonate 5[163]



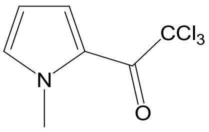
(1-Methyl-4-nitro-1*H*-pyrrol-2-yl) methanol **12** (0.83 g, 5.32 mmol) was dissolved in thionyl chloride (10 ml). The solution was then refluxed for 1 hour, and the excess thionyl chloride removed under reduced pressure. The residue was then dissolved in 20 ml DCM and filtered to get rid of the inorganic impurities which might be formed during the reaction. The DCM was removed under reduced pressure and the residue was refluxed in triethylphosphite (5 ml) for 2 hours at 160 ˚C. The excess phosphite was removed under high vacuum (1.5 mmHg @ 70 ˚C) to give the desired product initially as a brown oil, which was purified through a silica column (mobile phase: ethyl acetate ; Rf 0.18) to yield the product as a yellow-orange oil.

Yield; 1.29 g, 88 %
IR (NaCl): 3457, 3137, 2984, 1555,1519, 1497, 1425, 1346, 1308, 1249,, 1161,1025 cm.$^{-1}$

δ_H ($CDCl_3$): 1.30 (6H, t, CH_3, *J* = 7.0 Hz), 3.09 (2H, d, (CH_2)P, *J* = 20.45 Hz), 3.72 (3H, s, N-Me), 4.01 (4H, m, (CH_2)CH_3, *J* = 7.0 Hz), 6.63 (1H, d, Ar-H, *J* = 1.6 Hz), 7.46 (1H, d, Ar-H, *J* = 1.6 Hz), δ_P ($CDCl_3$), 23.44.

HRESI-MS (M+H): Found 277.09442 calculated for $C_{10}H_{18}O_5\ N_2P$ 277.09478 (-1.31094 ppm).

Preparation of 2-trichloroacetyl-*N*-methylpyrrole 10 [127]



Trichloroacetylchloride (8.97 g, 49.43 mmol) in DCM (50 ml) was placed in a round-bottomed flask at room temperature under nitrogen. A solution of *N*-methylpyrrole **14** (4.01 g, 49.43 mmol) in DCM (20 ml) was then added dropwise over 2.5 h. The solution was then allowed to stir overnight and the solvent removed under reduced pressure to yield the crude product, which was purified through a silica column to yield the product as a light brown solid.

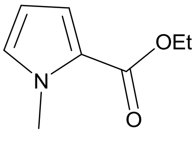
Yield = 8.396g, 75 %, m.p. = 63-64 ºC, (Lit = 64-65 ºC).[127]

IR (KBr): 3119, 2952, 1664, 1508, 1456, 1404, 1361 1243, 1100, 1067, 741 cm.$^{-1}$

δ_H ($CDCl_3$): 3.98 (3H, s, N-CH_3), 6.23 (1H, dd, Ar-H, *J* = 2.44 Hz & *J* = 4.44 Hz), 6.97 (1H, t, Ar-H, *J* = 1.76 Hz), 7.51 (1H, dd, Ar-H, *J* = 1.56 Hz & *J* = 4.44 Hz).

δ_C ($CDCl_3$): 38.68 (CH_3), 96.52 (CCl_3), 109.07 (C), 122.02 (C-H), 124.18 (C-H), 133.80 (C-H), 173.04 (C=O).

Preparation of ethyl 1-methyl-1*H*-pyrrole-2-carboxylate 15 [172]



2-Trichloroacetyl-*N*-methylpyrrole **10** (4.51 g, 19.86 mmol) and ethanol (40 ml) were placed in a round-bottomed flask under nitrogen. Sodium ethoxide (1.35g, 19.87 mmol) was then added and the solution was refluxed for 2 hrs. The reaction was then quenched with water and extracted with DCM (2 x 50 ml). The DCM fractions were dried over Mg_2SO_4, and the solvent was removed under reduced pressure to yield a brown-yellow oil.

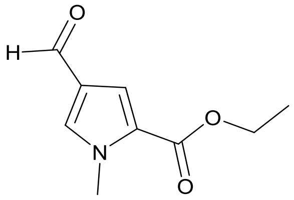
Yield; 2.79 g, 92 %.

IR (KBr): 3112, 2981, 1698, 1531, 1445, 1481, 1415, 1385, 1320, 1247, 1113, 742 cm^1.

δ_H ($CDCl_3$): 1.35 (3H, t, $CH_2(CH_3)$, J = 7.12 Hz), 3.93 (3H, s, N-CH_3), 4.27 (2H, q, CH_2, J = 7.12 Hz), 6.11 (1H, dd, Ar-H, J = 2.52 Hz & J = 3.96 Hz), 6.78 (1H, t, Ar-H, J = 2.12 Hz), 6.94 (1H, dd, Ar-H, J = 1.84 Hz & J = *3.96* Hz).

HRESI-MS (M+H): Found 154.08620 calculated for $C_8H_{12}O_2N$ 154.08626 (-0.3781 ppm).

Preparation of ethyl 4-formyl-1-methyl-1*H*-pyrrole-2-carboxylate 6 [173]



Ethyl 1-methyl-1*H*-pyrrole-2-carboxylate **15** (3.5 g, 28.24 mmol) and aluminium trichloride (8.0 g, 59.99 mmol) were added to a solution of nitromethane (40 ml) and 1,2-dichloroethane (40 ml) at -30 ˚C. Dichloromethylmethyl ether (2.5 ml, 28 mmol) in 1,2-dichloroethane (10 ml) was then added rapidly and the solution allowed to stir at -30 ˚C for 16 h. The solution was then poured onto ice (50 g) and the layers separated. The aqueous layer was extracted with ether (50 ml), the organic layers combined and dried ($MgSO_4$), filtered and the solvent removed under reduced pressure, to yield the crude product as a crystalline brown/black solid which was purified using a silica column (mobile phase: ethyl acetate: hexane, 1:4 ; Rf 0.16) to yield the desired product as a crystalline yellow solid.

.

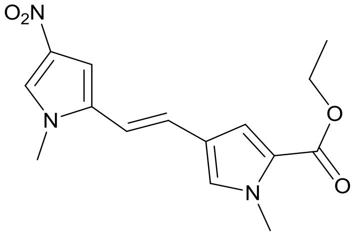
Yield; 3.313 g, 80%, m.p. = 67-69 ˚C, (Lit = 66-68 ºC).[173]

IR (NaCl): 3130, 2980, 2767, 2719, 2690, 1703, 1676, 1547, 1501, 1471, 1437, 1264, 1210, 1131, 1098, 1067, 1019cm.$^{-1}$

δ_H ($CDCl_3$): 1.37 (3H, t, $CH_2(CH_3)$, *J* = 7.12 Hz), 3.99 (3H, s, N-Me), 4.32 (2H, q, $(CH_2)CH_3$, *J* = 7.12 Hz), 7.38 (2H, m, 2(Ar-H)), 9.77 (1H, s, CO(H)).

HRESI-MS (M-H): Found 180.06597 calculated for $C_9H_{10}O_3N$ 180.06662 (-3.58617 ppm).

Preparation of ethyl 1-methyl-4-[(*E*)-2-(1-methyl-4-nitro-1*H*-pyrrol-2-yl) ethenyl]-1*H*-pyrrole-2-carboxylate 16 [163]



1-Methyl-4-nitro-1*H*-pyrrol-2-yl)methylphosphonate **5** (0.50 g, 1.81 mmol) and 4-formyl-1-methyl-1*H*-pyrrole-2-carboxylate **6** (0.328 g, 1.81 mmol) were dissolved in THF (4 ml) under nitrogen. Sodium hydride (0.26 g, 10.83 mmol) was then added in small portions over 5 min and the solution refluxed for 1h. Ice water (20ml) was then added and the solution was extracted with DCM (2x30 ml). The organic layers were combined, dried ($MgSO_4$) and the solvent removed under reduced pressure to yield the crude product which was filtered through a silica column (mobile phase: ethyl acetate: hexane, 1:3 ; Rf 0.21) to yield the desired product as an orange solid.

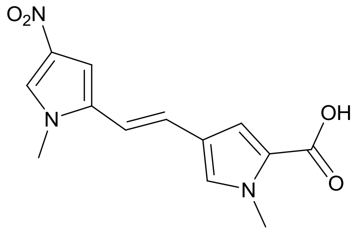
Yield; 0.329 g, 60 %, m.p. = 158-160˚C , (Lit = 158-161 ºC).[163]

IR (NaCl): 3139,3118, 2956, 2926, 1688, 1638, 1559, 1534, 1498, 1509, 1439, 1405, 1297, 1247, 1100 cm.$^{-1}$

δ_H ($CDCl_3$): 1.38 (3H, t, $CH_2(CH_3)$, *J* = 7.15 Hz), 3.69 (3H, s, N-Me), 3.93 (3H, s, N-Me), 4.32 (2H, q, $(CH_2)CH_3$, *J* = 7.15 Hz), 6.52 (1H, d, *CH*=CH, *J* = 16.0 Hz)),

6.79(1H, d, CH=*CH*, *J* = 16 Hz), 6.84 (1H, d, Ar-H, *J* = 1.8 Hz)), 6.88 (1H, d, Ar-H, *J* = 1.8 Hz), 7.11 (1H, d, Ar-H, (*J* = 1.8 Hz), 7.45 (1H, d, Ar-H, *J* = 1.8 Hz).
HRESI-MS (M-H): Found 302.11493 calculated for C_{15} $H_{16}O_4N_3$ 302.11463 (0.99273 ppm).

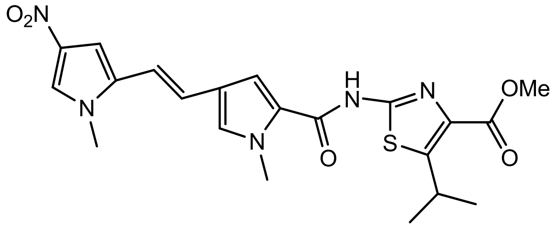
Preparation of 1-methyl-4-[(*E*)-2-(1-methyl-4-nitro-1*H*-pyrrol-2-yl)ethenyl]-1*H*-pyrrole-2-carboxylic acid 17[147]



Ethyl 1-methyl-4-[(*E*)-2-(1-methyl-4-nitro-1*H*-pyrrol-2-yl)ethenyl]-1*H*-pyrrole-2-carboxylate **16** (0.329 g, 1.084 mmol) was dissolved in acetonitile (5ml). To this solution, sodium hydroxide (0.26 g, 6.5 mmol) in water (5 ml) was added and the solution was refluxed for 1 h. The solvent was then evaporated and the residue was dissolved in water (10 ml) .The solution was cooled to 0˚C and diluted hydrochloric acid added until a pH of 2 was reached, at which point the product precipitated as a yellow solid.

Yield; 0.274 g, 92 %, m.p. = no distinct m.p. decomposes at >230 ˚C.
IR (KBr): 3300-2700, 3134, 292, 1672, 1559, 1493, 1453, 1413, 1302, 1259, 1137, 1111 cm.$^{-1}$
δ_H (DMSO): 3.71 (3H, s, N-Me), 3.82 (3H, s, N-Me), 6.74 (1H, d, *CH*=CH, *J* = 16.0 Hz), 6.89 (1H, d, Ar-H, *J* = 1.8 Hz), 6.90 (1H, d, CH=*CH*, *J* = 16.0 Hz), 7.13 (1H, d, Ar-H, *J* = 1.8 Hz), 7.23 (1H, d, Ar-H, *J* = 1.8 Hz), 7.92 (1H, d, Ar-H, *J* = 1.8 Hz), 12.30 (1H, s, COO*H*).
HRESI-MS (M+H): Found 276.09778 calculated for C_{13} $H_{14}O_4N_3$ 276.09788 (-0.36142 ppm).

Preparation of (E)-methyl 5-isopropyl-2-(1-methyl-4-(2-(1-methyl-4-nitro-2,3-dihydro-1 H –pyrrol-2-yl)vinyl)-1 H-pyrrole-2-carboamido)thiazole-4-carboxylate 24

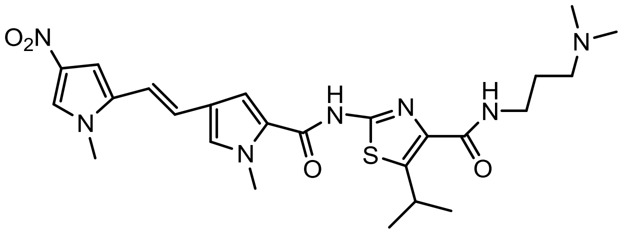


1-Methyl-4-[(*E*)-2-(1-methyl-4-nitro-1*H*-pyrrol-2-yl)ethenyl]-1*H*-pyrrole-2-carboxylic acid **17** (0.20 g ,0.727 mmol) was dissolved in anhydrous THF (10 ml) at room temperature. To this solution, Ghozes reagent (0.485 g, 3.635 mmol) was added and the solution stirred overnight under nitrogen or until the disappearance of the starting material monitored by TLC. Methyl 2-amino-5-isopropyl-1,3-thiazole-4-carboxylate **7** (0.145 g,0.727 mmol) and anhydrous *N*,*N*-diisopropylethylamine (0.281g , 2.181mmol) were dissolved in anhydrous THF (5 ml) . This solution was added to the reaction mixture after being cooled to 0 C for 20 min. The reaction mixture was then allowed to stir overnight at room temperature under nitrogen. The solvent was removed under reduced pressure and dil. hydrochloric acid (30 ml) was added to the crude mixture. The product was extracted with ethyl acetate (2x30 ml) and the solvent was dried ($MgSO_4$), filtered, and removed under reduced pressure to yield the crude product which was purified through a silica column (mobile phase: ethyl acetate: hexane, 1:1; Rf 0.27) to yield the product as an orange solid.

Yield; 0.233 g, 70 %, m.p. = no distinct m.p. decomposes at >230 ˚C.
IR (KBr): 3600-3300, 3429, 3343, 2922, 2852, 2349, 1713, 1651, 1538, 1487, 1463, 1432, 1413, 1290 ,1171 cm.$^{-1}$
δ_H (DMSO): 1.29 (6H, d, CH_3, *J* = 6.84 Hz) 3.72 (3H, s, N-Me), 3.80 (3H, s, N-Me), 3.89(3H, s, OCH_3) , 4.01 (1H, Q, CH, *J* = 6.8 Hz), 6.64 (1H, d, *CH*=CH, *J* = 16.0 Hz), 6.97 (1H, d, Ar-H, *J* = 1.8 Hz), 6.98 (1H, d, CH=*CH*, *J* = 16.0 Hz), 7.36 (1H, d, Ar-H, *J* = 1.8 Hz), 7.68 (1H, d, Ar-H, *J* = 1.8 Hz), 7.96 (1H, d, Ar-H, *J* = 1.8 Hz), 12.39 (1H, s, N-H).

HRESI-MS (M+H): Found 458.14896 calculated for $C_{21}H_{24}O_5N_5S$ 458.14927 (-0.67603 ppm).

Preparation of (E)-N-(3-(dimethylamino)propyl)-5-isopropyl-2-(1-methyl-4-(2-(1-methyl-4-nitro-1H-pyrrol-2-yl)vinyl)-1H-pyrrole-2-carboxamido)thiazole-4-carboamide 25



(E)-Methyl 5-isopropyl -2-(1-methyl-4-(2-(1-methyl-4-nitro-2,3-dihydro-1*H* –pyrrol-2-yl)vinyl)-1*H*-pyrrole-2- carboamido)thiazole-4-carboxylate **24** (0.233, 0.509 mmol) was dissolved in 10 ml of DMAP. The solvent was heated to 100 °C and allowed to stir for 12 hrs or until the disappearance of the starting material monitored by TLC. The solvent was then removed under reduced pressure to yield the crude product which was purified through a base activated aluminium oxide silica column (mobile phase: ethyl acetate: methanol, 1:0.05; Rf, 0.42) to yield the product as an orange-red solid.

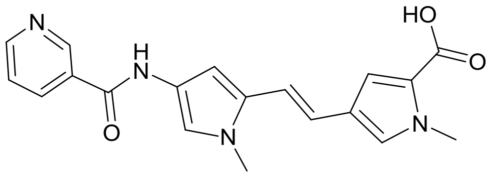
Yield; 0.244 g, 91%, m.p. = no distinct m.p. decomposes at >230 ˚C.

IR (KBr): 3600-3300, 3432, 3116, 2922, 2853, 2360, 2341, 1647, 1545, 1507, 1476, 1423, 1395, 1305, 1136 cm.$^{-1}$

δ_H (MeOD): 1.35 (6H, d, CH_3, *J* = 6.88 Hz), 2.03 (2H, Q, $CH_2(CH_2)CH_2$, *J* = 7.2 Hz), 2.91 (6H,s, CH_3), 3.18 (2H, t, CH_2 , J= 7.3 Hz), 3.48(2H , t , CH_2, J= 6.5 Hz), 3.75 (3H, s, N-Me), 3.98 (3H, s, N-Me), 4.31(1H, Q, CH, J= 6.8 Hz) , 6.75 (1H, d, *CH*=CH, *J* = 16.0 Hz), 6.86 (1H, d, Ar-H, *J* = 1.6 Hz), 6.92 (1H, d, CH=*CH*, *J* = 16.0 Hz), 7.25 (1H, d, Ar-H, *J* = 1.6 Hz), 7.28 (1H, d, Ar-H, *J* = 1.6 Hz), 7.69 (1H, d, Ar-H, *J* = 1.6 Hz). [CONH protons are exch. with the solvent]

HRESI-MS (M+H): Found 528.23879 calculated for $C_{25}H_{34}O_4N_7S$ 528.23875 (0.07158 ppm).

Preparation of (E)-1-methyl-4-(2-(1-methyl-4-(nicotinamido)-1H-pyrrol-2-yl)vinyl)-1H-pyrrole-2-carboxylic acid 29



1-Methyl-4-[(*E*)-2-(1-methyl-4-nitro-1*H*-pyrrol-2-yl)ethenyl]-1*H*-pyrrole-2-carboxylic acid **17** (0.2 g, 0.27 mmol) was dissolved in dioxane (3 ml). The solution was basified by adding sodium hydroxide (0.058 g/ 290 µl of 5 M NaOH, 1.45 mmol) dissolved in 2 ml of water to a pH of 11.5. Pd/C (10 %, 0.200 g) was then added in small portions, followed by a solution of sodium borohydride (0.1 g, 2.64 mmol) in water (2 ml). The suspension was then allowed to stir for 5 min, filtered over Kieselguhr directly into a flask containing *N*-methylimidazole (6 mg, 0.073 mmol), TMEDA (8.5 mg, 0.073 mmol) and Na_2CO_3 (77 mg, 0.727 mmol) in water (5 mL). The reaction mixture was allowed to stir at room temperature and a pH meter probe was used to monitor the pH of the reaction. Nicotinoyl chloride hydrochloride (129 mg, 0.727 mmol) was added in small portions over 1 hr simultaneously with the addition of 1 M aqueous sodium hydroxide using a micro-syringe to strictly maintain the pH at 11.5. The solution was then allowed to stir overnight. The solvent was removed under reduced pressure and the crude product was purified through a reverse phase C18 silica column (mobile phase: acetonitrile: water: TFA, 1:9:0.1) to yield the desired product as a brown solid.

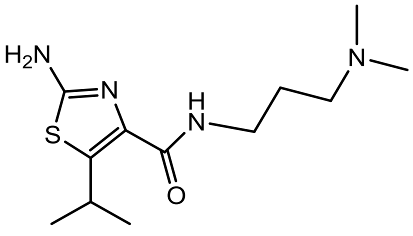
Yield; 0.132 g, 51%, m.p. = no distinct m.p. decomposes at >230 ˚C.

IR (KBr): 3500-2700, 3413, 2922, 2853, 1693, 1626, 1538, 1445, 1389, 1278, 1186 cm.$^{-1}$

δ_H (DMSO): 3.62 (3H, s, N-Me), 3.82 (3H, s, N-Me), 6.42(1H, d, Ar-H, J = 1.6 Hz), 6.61 (1H, d, *CH*=CH, *J* = 16.0 Hz), 6.78 (1H, d, CH=CH, *J* = 16.0 Hz), 7.08 (1H, d, Ar-H, *J* = 1.96 Hz), 7.19 (1H, d, Ar-H, *J* = 1.8 Hz), 7.22 (1H, d, Ar-H, *J* = 1.92 Hz),

7.74 (1H, dd, Ar-H, *J* = 5.0 and 8.0 Hz), 8.50 (1H, d, Ar-H, *J* = 9.0 Hz), 8.81 (1H, dd , Ar-H, *J* = *1.6* and 5.0 Hz), 9.18 (1H, Ar-H, *J* = 1.6 Hz), 10.54 (1H, s, CONH).
LRESI-MS (M+H): Found 351.09 calculated for $C_{19}H_{19}O_4N_3$ 351.14517.
HRESI-MS (M+H): Found 351.14514 calculated for $C_{19}H_{19}O_4N_3$ 351.14517.

Preparation of methyl 2-amino-5-isopropylthiazole-4-carboxylate



Methyl 2-amino-5-isopropyl-1,3-thiazole-4-carboxylate **7** (0.31 g, 1.548 mmol)) was dissolved in 10 ml of DMAP. The solvent was heated to 100 °C and allowed to stir for 12 hrs or until the disappearance of the starting material monitored by TLC. The solvent was removed under reduced pressure to yield the crude product which was purified through a silica column (mobile phase: ethyl acetate: methanol: triethylamine, 4:1:0.05; Rf, 0.1) to yield the product as a yellow oil.

Yield; 0.359 g, 86%
δ_H ($CDCl_3$): 1.23 (6H, d, CH_3, *J* = 6.84 Hz), 1.73(2H, Q, $CH_2(CH_2)CH_2$, *J* = 6.92Hz), 2.22 (6H, s, CH_3N), 2.33(2H , t, CH_2, *J*= 6.84Hz), 3.39 (2H, q, CH_2 , *J*= 6.84Hz), 4.31 (1H, hept, CH, *J*= 6.84 Hz), 4.86 (2H, br, s, NH_2 ,exch), 7.52(1H, br, s, CONH).
HRESI-MS (M+H): Found 271.15852 calculated for $C_{12}H_{23}ON_4S$ 271.15871 (-0.71368 ppm).

6 References

1 Belmont, P.; Constant, J. F. & Demeunck, M. Nucleic acid conformation diversity: From structure to function and regulation *Chemical society Reviews* **30**, 70-81, (2001).

2 Dickerson, R. E. Definitions and nomenclature of nucleic acid structure parameters. *J Biomol Struct Dyn* **6**, 627-634, (1989).

3 Ladbury, J. E. & Chowdhry, B. Z. Sensing the heat: the application of isothermal titration calorimetry to thermodynamic studies of biomolecular interactions. *Chem Biol* **3**, 791-801, (1996).

4 Watson, J. D. & Crick, F. H. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**, 737-738, (1953).

5 Saenger, W. Principles of Nucleic Acid Structure. *Springer-Verlag: New York*, (1984).

6 http://www.cem.msu.edu/~reusch/VirtualText/nucacids.htm. *Michigan State University,* Accessed on 27 May 2010.

7 Dervan, P. B. Molecular recognition of DNA by small molecules. *Bioorg Med Chem* **9**, 2215-2235, (2001).

8 McCammon, J. A. & Harvey, S. C. Dynamics of Proteins and Nucleic Acids. *Cambridge University Press,* Cambridge, (1987).

9 Ellis, R. J. & Pinheiro, T. J. Medicine: danger--misfolding proteins. *Nature* **416**, 483-484, (2002).

10 Radford, S. E. & Dobson, C. M. From computer simulations to human disease: emerging themes in protein folding. *Cell* **97**, 291-298, (1999).

11 Fleming, A. Observations on the bacteriostatic action of MI & B 693 and on the influence thereon of bacteria and peptone. *J. Path. Bact.* **50**, 69-81, (1940).

12 Schreiber, S. L. Target-oriented and diversity-oriented organic synthesis in drug discovery. *Science* **287**, 1964-1969, (2000).

13 Yang, X. L. & Wang, A. H. Structural studies of atom-specific anticancer drugs acting on DNA. *Pharmacol Ther* **83**, 181-215, (1999).

14 Geierstanger, B. H. & Wemmer, D. E. Complexes of the minor groove of DNA. *Annu Rev Biophys Biomol Struct* **24**, 463-493, (1995).

15 van Drie, J. A.; Roher, D. C.; Blinn, J. R. & Gao, H. Modern Methods of Drug Discovery (Hillisch, A. and Hilgenfeld, R., Eds.), Birkhauser Verlag-Basel, 203-222, (2003).

16 McPherson, J. D.; Marra, M.; Hillier, L.; Waterston, R. H.; Chinwalla, A.;Wallis, J.;Sekhon, M.;Wylie, K.;Mardis, E. R.;Wilson, R. K.;Fulton, R.; Kucaba, T. A.; Wagner-McPherson, C.; Barbazuk, W. B.; Gregory, S. G.;Humphray, S. J.; French, L.; Evans, R. S.; Bethel, G.; Whittaker, A.; Holden, J. L.; McCann, O. T.; Dunham, A.; Soderlund, C.; Scott, C. E.; Bentley, D. R.; Schuler, G.; Chen, H. C.;Jang, W.; Green, E. D.; Idol, J. R.;Maduro, V. V.;Montgomery, K. T.;Lee, E.;Miller, A.;Emerling, S.; Kucherlapati;Gibbs, R.; Scherer, S.; Gorrell, J. H.; Sodergren, E.; Clerc-Blankenburg, K.; Tabor, P.; Naylor, S.; Garcia, D.; de Jong, P. J.; Catanese, J. J.; Nowak, N.; Osoegawa, K.; Qin, S.; Rowen, L.; Madan, A.; Dors, M.;

Hood, L.; Trask, B.; Friedman, C.;Massa, H.; Cheung, V. G.; Kirsch, I. R.;Reid, T.; Yonescu, R.; Weissenbach, J.; Bruls, T.;Heilig, R.; Branscomb, E.;Olsen, A.; Doggett, N.; Cheng, J. F.; Hawkins, T.; Myers, R. M.; Shang, J.; Ramirez, L.; Schmutz, J.; Velasquez, O.; Dixon, K.; Stone, N. E.; Cox, D. R.; Haussler, D.; Kent, W. J.; Furey, T.;Rogic, S.; Kennedy, S.; Jones, S.; Rosenthal, A.; Wen, G.; Schilhabel, M.; Gloeckner, G.; Nyakatura, G.; Siebert, R.; Schlegelberger, B.; Korenberg, J.; Chen, X. N.; Fujiyama, A.; Hattori, M.; Toyoda, A.; Yada, T.; Park, H. S.; Sakaki, Y.; Shimizu, N.; Asakawa, S.; Kawasaki, K.; Sasaki, T.; Shintani, A.; Shimizu, A.; Shibuya, K.; Kudoh, J.; Minoshima, S.; Ramser, J.; Seranski, P.; Hoff, C.; Poustka, A.; Reinhardt, R. & Lehrach, H. A physical map of the human genome. *Nature* **409**, 934-941, (2001).

17 Stephen Oliver reviews, Proteins and Proteomics: A Laboratory Manual by Richard J. Simpson. Practical proteomics. *Nature* **422**, 473, (2003).

18 Gavin, A. C.; Bosche, M.; Krause, R.; Grandi, P.; Marzioch, M.; Bauer, A.; Schultz, J.; Rick, J. M.; Michon, A. M.; Cruciat, C. M.; Remor, M.; Hofert, C.; Schelder, M.; Brajenovic, M.; Ruffner, H.; Merino, A.; Klein, K.; Hudak, M.; Dickson, D.; Rudi, T.; Gnau, V.; Bauch, A.; Bastuck, S.; Huhse, B.; Leutwein, C.; Heurtier, M. A.; Copley, R. R.; Edelmann, A.; Querfurth, E.; Rybin, V.; Drewes, G.; Raida, M.; Bouwmeester, T.; Bork, P.; Seraphin, B.; Kuster, B.; Neubauer, G. & Superti-Furga, G. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**, 141-147, (2002).

19 Drews, J. Biotechnology's metamorphosis into a drug discovery industry. *Nat Biotechnol* **16 Suppl**, 22-24, (1998).

20 Dabrowiak, J. C.; Stankus, A. A. & Goodisman, J. Sequence Specificity of Drug-DNA Interactions. In Nucleic Acid Targeted Drug Design. New York. (1992).

21 Chaires, J. B.; Leng, F.; Przewloka, T.; Fokt, I.; Ling, Y. H.; Perez-Soler, R. & Priebe, W. Structure-based design of a new bisintercalating anthracycline antibiotic. *J Med Chem* **40**, 261-266, (1997).

22 Crick, F. Central dogma of molecular biology. *Nature* **227**, 561-563, (1970).

23 Ptashne, M. A Genetic Switch. *Blackwell Scientific Publications*, Palo Alto, California, (1986).

24 Bailly, C. & Henichart, J. P. DNA recognition by intercalator-minor-groove binder hybrid molecules. *Bioconjug Chem* **2**, 379-393, (1991).

25 Dervan, P. B. Design of sequence-specific DNA-binding molecules. *Science* **232**, 464-471, (1986).

26 White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E. & Dervan, P. B. Recognition of the four Watson-Crick base pairs in the DNA minor groove by synthetic ligands. *Nature* **391**, 468-471, (1998).

27 http://users.ugent.be/~avierstr/principles/centraldogma.html.
Ghent University. Accessed on 12/1/2008. .

28 Oleksy, A.; Blanco, A. G.; Boer, R.; Uson, I.; Aymami, J.; Rodger, A.; Hannon, M. J. & Coll, M. Molecular recognition of a three-way DNA junction by a metallosupramolecular helicate. *Angew Chem Int Ed Engl* **45**, 1227-1231, (2006).

29 Clark, A. S.; Deans, B.; Stevens, M. F.; Tisdale, M. J.; Wheelhouse, R. T.; Denny, B. J. & Hartley, J. A. Antitumor imidazotetrazines. 32. Synthesis of novel imidazotetrazinones and related bicyclic heterocycles to probe the mode of action of the antitumor drug temozolomide. *J Med Chem* **38**, 1493-1504, (1995).

30 Goldacre, R. J.; Loveless, A. & Ross, W. C. Mode of production of chromosome abnormalities by the nitrogen mustards; the possible role of cross-linking. *Nature* **163**, 667-669, (1949).

31 Sarma, R. H. (Editor). Nucleic acid geometry and dynamics. *Pergamon Press. New York* (1980).

32 Pommier, Y. & Cherfils, J. Interfacial inhibition of macromolecular interactions: nature's paradigm for drug discovery. *Trends Pharmacol Sci* **26**, 138-145, (2005).

33 Hurley, L. H. DNA and its associated processes as targets for cancer therapy. *Nat Rev Cancer* **2**, 188-200, (2002).

34 Dickinson, L. A.; Gulizia, R. J.; Trauger, J. W.; Baird, E. E.; Mosier, D. E.; Gottesfeld, J. M. & Dervan, P. B. Inhibition of RNA polymerase II transcription in human cells by synthetic DNA-binding ligands. *Proc Natl Acad Sci U S A* **95**, 12890-12895, (1998).

35 Ehley, J. A.; Melander, C.; Herman, D.; Baird, E. E.; Ferguson, H. A.; Goodrich, J. A.; Dervan, P. B. & Gottesfeld, J. M. Promoter scanning for transcription inhibition with DNA-binding polyamides. *Mol Cell Biol* **22**, 1723-1733, (2002).

36 McBryant, S. J.; Baird, E. E.; Trauger, J. W.; Dervan, P. B. & Gottesfeld, J. M. Minor groove DNA-protein contacts upstream of a tRNA gene detected with a synthetic DNA binding ligand. *J Mol Biol* **286**, 973-981, (1999).

37 Dervan, P. B. & Burli, R. W. Sequence-specific DNA recognition by polyamides. *Curr Opin Chem Biol* **3**, 688-693, (1999).

38 Anthony, N. G.; Johnston, B. F.; Khalaf, A. I.; MacKay, S. P.; Parkinson, J. A.; Suckling, C. J. & Waigh, R. D. Short lexitropsin that recognizes the DNA minor groove at 5'-ACTAGT-3': understanding the role of isopropyl-thiazole. *J Am Chem Soc* **126**, 11338-11349, (2004).

39 Neidle, S. DNA minor-groove recognition by small molecules. *Nat Prod Rep* **18**, 291-309, (2001).

40 Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P. & Dickerson, R. E. Binding of an antitumor drug to DNA, Netropsin and C-G-C-G-A-A-T-T-BrC-G-C-G. *J Mol Biol* **183**, 553-563, (1985).

41 Abrescia, N. G.; Malinina, L. & Subirana, J. A. Stacking interaction of guanine with netropsin in the minor groove of d(CGTATATACG)2. *J Mol Biol* **294**, 657-666, (1999).

42 Chalikian, T. V.; Plum, G. E.; Sarvazyan, A. P. & Breslauer, K. J. Influence of drug binding on DNA hydration: acoustic and densimetric characterizations of netropsin binding to the poly(dAdT).poly(dAdT) and poly(dA).poly(dT) duplexes and the poly(dT).poly(dA).poly(dT) triplex at 25 degrees C. *Biochemistry* **33**, 8629-8640, (1994).

43 Chenoweth, D. M.; Harki, D. A.; Phillips, J. W.; Dose, C. & Dervan, P. B. Cyclic pyrrole-imidazole polyamides targeted to the androgen response element. *J Am Chem Soc* **131**, 7182-7188, (2009).

44 Chenoweth, D. M. & Dervan, P. B. Structural basis for cyclic py-im polyamide allosteric inhibition of nuclear receptor binding. *J Am Chem Soc* **132**, 14521-14529, (2010).

45 Perrin, C. L. & Nielson, J. B. "Strong" hydrogen bonds in chemistry and biology. *Annu Rev Phys Chem* **48**, 511-544, (1997).

46 Baker, E. N. & Hubbard, R. E. Hydrogen bonding in globular proteins. *Prog Biophys Mol Biol* **44**, 97-179, (1984).

47 *http://en.wikipedia.org/wiki/File:3D_model_hydrogen_bonds_in_water.jpg*, Maňas, M. (2007).

48 Tanford, C. The hydrophobic effect and the organization of living matter. *Science* **200**, 1012-1018, (1978).

49 *http://www.bio.brandeis.edu/classes/biochem104/hydrophobic_effect.pdf.*, Cohen, C. The hydrophobic effect. (2005).

50 *http://wwwarpe.snv.jussieu.fr/td_2_eng/hydhyd.html*. Bouvier, D. Hydrophobic effect and van der Waals interactions. (1999).

51 Misra, V. K. & Honig, B. On the magnitude of the electrostatic contribution to ligand-DNA interactions. *Proc Natl Acad Sci U S A* **92**, 4691-4695, (1995).

52 Baginski, M.; Fogolari, F. & Briggs, J. M. Electrostatic and non-electrostatic contributions to the binding free energies of anthracycline antibiotics to DNA. *J Mol Biol* **274**, 253-267, (1997).

53 Kostjukov, V. V.; Khomytova, N. M.; Davies, D. B. & Evstigneev, M. P. Electrostatic contribution to the energy of binding of aromatic ligands with DNA. *Biopolymers* **89**, 680-690, (2008).

54 Davis, M. E. & Mccammon, J. A. Electrostatics in biomolecular structure and dynamics. *Chem Rev* **90**, 509–521, (1990).

55 Noskov, S. Y. & Lim, C. Free energy decomposition of protein-protein interactions. *Biophys J* **81**, 737-750, (2001).

56 *http://en.wikipedia.org/wiki/Lennard-Jones_potential.* Lennard-Jones potential (2009).

57 Knapp, M.; Bellamacina, C.; Murray, J. M. & Bussiere, D. E. Targeting cancer: the challenges and successes of structure-based drug design against the human purinome. *Curr Top Med Chem* **6**, 1129-1159, (2006).

58 Ababou, A. & Ladbury, J. E. Survey of the year 2004: literature on applications of isothermal titration calorimetry. *J Mol Recognit* **19**, 79-89, (2006).

59 Cliff, M. J.; Gutierrez, A. & Ladbury, J. E. A survey of the year 2003 literature on applications of isothermal titration calorimetry. *J Mol Recognit* **17**, 513-523, (2004).

60 Cliff, M. J. & Ladbury, J. E. A survey of the year 2002 literature on applications of isothermal titration calorimetry. *J Mol Recognit* **16**, 383-391, (2003).

61 Wiseman, T.; Williston, S.; Brandts, J. F. & Lin, L. N. Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal Biochem* **179**, 131-137, (1989).

62 Haq, I.; Chowdhry, B. Z. & Jenkins, T. C. Calorimetric techniques in the study of high-order DNA-drug interactions. *Methods Enzymol* **340**, 109-149, (2001).

63 Haq, I.; Jenkins, T. C.; Chowdhry, B. Z.; Ren, J. & Chaires, J. B. Parsing free energies of drug-DNA interactions. *Methods Enzymol* **323**, 373-405, (2000).

64 Microcal Inc Northampton, M. *VP-ITC Instruction Manual*, <http://www.microcalorimetry.com> (2001).

65 Holdgate, G. A. Making cool drugs hot: isothermal titration calorimetry as a tool to study binding energetics. *Biotechniques* **31**, 164-166, 168, 170 passim, (2001).

66 O'Brien, R.; Ladbury, J. E. & Chowdhry, B. Z. In Protein-Ligand Interactions: Hydrodynamics and Calorimetry. . *Oxford University Press: Oxford, UK*, 263-286, (2001).

67 Horn, J. R.; Brandts, J. F. & Murphy, K. P. van't Hoff and calorimetric enthalpies II: effects of linked equilibria. *Biochemistry* **41**, 7501-7507, (2002).

68 Chaires, J. B. Possible origin of differences between van't Hoff and calorimetric enthalpy estimates. *Biophys Chem* **64**, 15-23, (1997).

69 Naghibi, H.; Tamura, A. & Sturtevant, J. M. Significant discrepancies between van't Hoff and calorimetric enthalpies. *Proc Natl Acad Sci U S A* **92**, 5597-5599, (1995).

70 Todd, M. J. & Gomez, J. Enzyme kinetics determined using calorimetry: a general assay for enzyme activity? *Anal Biochem* **296**, 179-187, (2001).

71 Ladbury, J. E.; Klebe, G. & Freire, E. Adding calorimetric data to decision making in lead discovery: a hot tip. *Nat Rev Drug Discov* **9**, 23-27, (2010).

72 Haq, I.; Ladbury, J. E.; Chowdhry, B. Z.; Jenkins, T. C. & Chaires, J. B. Specific binding of hoechst 33258 to the d(CGCAAATTTGCG)2 duplex: calorimetric and spectroscopic studies. *J Mol Biol* **271**, 244-257, (1997).

73 Haq, I. Thermodynamics of drug-DNA interactions. *Arch Biochem Biophys* **403**, 1-15, (2002).

74 Bergqvist, S.; Williams, M. A.; O'Brien, R. & Ladbury, J. E. Heat capacity effects of water molecules and ions at a protein-DNA interface. *J Mol Biol* **336**, 829-842, (2004).

75 Freyer, M. W.; Buscaglia, R.; Hollingsworth, A.; Ramos, J.; Blynn, M.; Pratt, R.; Wilson, W. D. & Lewis, E. A. Break in the heat capacity change at 303 K for complex binding of netropsin to AATT containing hairpin DNA constructs. *Biophys J* **92**, 2516-2522, (2007).

76 Freyer, M. W.; Buscaglia, R.; Nguyen, B.; Wilson, W. D. & Lewis, E. A. Binding of netropsin and 4,6-diamidino-2-phenylindole to an A2T2 DNA hairpin: a comparison of biophysical techniques. *Anal Biochem* **355**, 259-266, (2006).

77 Schwabe, J. W. The role of water in protein-DNA interactions. *Curr Opin Struct Biol* **7**, 126-134, (1997).

78 Lafont, V.; Armstrong, A. A.; Ohtaka, H.; Kiso, Y.; Mario Amzel, L. & Freire, E. Compensating enthalpic and entropic changes hinder binding affinity optimization. *Chem Biol Drug Des* **69**, 413-422, (2007).

79 Velazquez-Campoy, A.; Todd, M. J. & Freire, E. HIV-1 protease inhibitors: enthalpic versus entropic optimization of the binding affinity. *Biochemistry* **39**, 2201-2207, (2000).

80 Velazquez-Campoy, A.; Kiso, Y. & Freire, E. The binding energetics of first- and second-generation HIV-1 protease inhibitors: implications for drug design. *Arch Biochem Biophys* **390**, 169-175, (2001).

81 Ohtaka, H.; Velazquez-Campoy, A.; Xie, D. & Freire, E. Overcoming drug resistance in HIV-1 chemotherapy: the binding thermodynamics of Amprenavir and TMC-126 to wild-type and drug-resistant mutants of the HIV-1 protease. *Protein Sci* **11**, 1908-1916, (2002).

82 Bacha, U. M.; Barrila, J. A.; Velazquez-Campoy, A.; Leavitt, S. & Freire, E. Development of potent inhibitors of the SARS associated coronavirus protease 3CL^{pro} *Biophys J. 86, 97a. (Poster at 2004 Biophysical Society Meeting)*, (2004).

83 Nezami, A.; Luque, I.; Kimura, T.; Kiso, Y. & Freire, E. Identification and characterization of allophenylnorstatine-based inhibitors of plasmepsin II, an antimalarial target. *Biochemistry* **41**, 2273-2280, (2002).

84 Ward, W. H. & Holdgate, G. A. Isothermal titration calorimetry in drug discovery. *Prog Med Chem* **38**, 309-376, (2001).

85 Han, F.; Taulier, N. & Chalikian, T. V. Association of the minor groove binding drug Hoechst 33258 with d(CGCGAATTCGCG)2: volumetric, calorimetric, and spectroscopic characterizations. *Biochemistry* **44**, 9785-9794, (2005).

86 Mazur, S.; Tanious, F. A.; Ding, D.; Kumar, A.; Boykin, D. W.; Simpson, I. J.; Neidle, S. & Wilson, W. D. A thermodynamic and structural analysis of DNA minor-groove complex formation. *J Mol Biol* **300**, 321-337, (2000).

87 Lah, J. & Vesnaver, G. Energetic diversity of DNA minor-groove recognition by small molecules displayed through some model ligand-DNA systems. *J Mol Biol* **342**, 73-89, (2004).

88 Pilch, D. S.; Poklar, N.; Gelfand, C. A.; Law, S. M.; Breslauer, K. J.; Baird, E. E. & Dervan, P. B. Binding of a hairpin polyamide in the minor groove of DNA: sequence-specific enthalpic discrimination. *Proc Natl Acad Sci U S A* **93**, 8306-8311, (1996).

89 Bailly, C.; Chessari, G.; Carrasco, C.; Joubert, A.; Mann, J.; Wilson, W. D. & Neidle, S. Sequence-specific minor groove binding by bis-benzimidazoles: water molecules in ligand recognition. *Nucleic Acids Res* **31**, 1514-1524, (2003).

90 Ladbury, J. E. Just add water! The effect of water on the specificity of protein-ligand binding sites and its potential application to drug design. *Chem Biol* **3**, 973-980, (1996).

91 Chaires, J. B. A thermodynamic signature for drug-DNA binding mode. *Arch Biochem Biophys* **453**, 26-31, (2006).

92 Privalov, P. L. & Gill, S. J. Stability of protein structure and hydrophobic interaction. *Adv Protein Chem* **39**, 191-234, (1988).

93 Sturtevant, J. M. Heat capacity and entropy changes in processes involving proteins. *Proc Natl Acad Sci U S A* **74**, 2236-2240, (1977).

94 Kozlov, A. G. & Lohman, T. M. Adenine base unstacking dominates the observed enthalpy and heat capacity changes for the Escherichia coli SSB tetramer binding to single-stranded oligoadenylates. *Biochemistry* **38**, 7388-7397, (1999).

95 Morton, C. J. & Ladbury, J. E. Water-mediated protein-DNA interactions: the relationship of thermodynamics to structural detail. *Protein Sci* **5**, 2115-2118, (1996).

96 Sharp, K. A. In Thermodynamics in Biology *Oxford University Press: New York*, 113-130, (2000).

97 Oda, M.; Furukawa, K.; Ogata, K.; Sarai, A. & Nakamura, H. Thermodynamics of specific and non-specific DNA binding by the c-Myb DNA-binding domain. *J Mol Biol* **276**, 571-590, (1998).

98 Ha, J. H.; Capp, M. W.; Hohenwalter, M. D.; Baskerville, M. & Record, M. T., Jr. Thermodynamic stoichiometries of participation of water, cations and anions in specific and non-specific binding of lac repressor to DNA. Possible thermodynamic origins of the "glutamate effect" on protein-DNA interactions. *J Mol Biol* **228**, 252-264, (1992).

99 O'Brien, R.; DeDecker, B.; Fleming, K. G.; Sigler, P. B. & Ladbury, J. E. The effects of salt on the TATA binding protein-DNA interaction from a hyperthermophilic archaeon. *J Mol Biol* **279**, 117-125, (1998).

100 Kaul, M. & Pilch, D. S. Thermodynamics of aminoglycoside-rRNA recognition: the binding of neomycin-class aminoglycosides to the A site of 16S rRNA. *Biochemistry* **41**, 7695-7706, (2002).

101 Holbrook, J. A.; Capp, M. W.; Saecker, R. M. & Record, M. T., Jr. Enthalpy and heat capacity changes for formation of an oligomeric DNA duplex: interpretation in terms of coupled processes of formation and association of single-stranded helices. *Biochemistry* **38**, 8409-8422, (1999).

102 Baker, B. M. & Murphy, K. P. Evaluation of linked protonation effects in protein binding reactions using isothermal titration calorimetry. *Biophys J* **71**, 2049-2055, (1996).

103 Nguyen, B.; Stanek, J. & Wilson, W. D. Binding-linked protonation of a DNA minor-groove agent. *Biophys J* **90**, 1319-1328, (2006).

104 Berova, N.; Nakanishi, K. & Woody, R. W. Circular Dichroism: Principles and Applications, 2nd Edition. *Wiley-VCH*, (2000).

105 Eriksson, M. & Norden, B. Linear and circular dichroism of drug-nucleic acid complexes. *Methods Enzymol* **340**, 68-98, (2001).

106 Engel, G. Estimation of binding parameters of enzyme-ligand complex from fluorometric data by a curve fitting procedure: seryl-tRNA synthetase-tRNA Ser complex. *Anal Biochem* **61**, 184-191, (1974).

107 Lane, A. N. Nuclear magnetic resonance studies of drug-DNA complexes in solution. *Methods Enzymol* **340**, 252-281, (2001).

108 Wuthrich, K. NMR of Proteins and Nucleic Acids. *Wiley, New York*, (1986).

109 Wuthrich, K. Protein structure determination in solution by NMR spectroscopy. *J Biol Chem* **265**, 22059-22062, (1990).

110 Bloch, F.; Hansen, W. W. & Packard, M. E. The nuclear induction experiment *Phys.Rev.* **70**, 474-485., (1946).

111 Purcell, E. M.; Torrey, H. C. & Pound, R. V. Resonance Absorption by Nuclear Magnetic Moments in a Solid *Phys.Rev.* **69**, 37-38, (1946).

112 Keeler, J. Understanding NMR spectroscopy. *Wiley, Chichester, England*, (2005).

113 Hore, P. J. Nuclear magnetic resonance. *Oxford University Press*, (1995).

114 http://www.chem.ucalgary.ca/courses/351/Carey/Ch13/ch13-nmr-1.html#basics. Carey, F. A..
115 Propst, C. L. & Perun, T. J. Nucleic acid targeted drug design. *Marcel Dekker: New York, NY, ,* (1992).
116 Jeener, J. *Ampere International Summer School, Basko Polje, Yugoslavia,* (1971).
117 Martin, G. E. & Zekter, A. S. Two-Dimensional NMR Methods for Establishing Molecular Connectivity; VCH Publishers, Inc: New York. (1988).
118 Pastor, A. & Vivienteb, E. M. NMR spectroscopy in coordination supramolecular chemistry: A unique and powerful methodology *Coordination Chemistry Reviews* **252**, 2314-2345, (2008).
119 Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; DeBolt, I., S.; Ferguson, D.; Seibel, G. & Kollman, P. AMBER, a package of computer programs forapplying molecular mechanics, normal mode analysis, molecular dynamics and free energycalculations to simulate the structural and energetic properties of molecules. *Comp.Phys. Commun.* **91**, 1-41 (1995).
120 Case, D. A.; Darden, T. A.; Cheatham, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E. R. L. & Kollman, P. A. AMBER 10. *University of California, San Francisco.,* (2008).
121 Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Kenneth M. Merz; David M. Ferguson; David C. Spellmeyer; Thomas Fox; James W. Caldwell & Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules *J. Am. Chem. Soc.* **117**, 5179–5197, (1995).
122 Accessed on 20.08.2010. http://www.scripps.edu/rc/softwaredocs/msi/cerius45/FFBSim/2_Forcefields.html#568091 (Eq.19).
123 Grant, G. H. & Richards, G. W. Computational chemistry. *New York, Oxford University Press Inc.,* (1998).
124 Ryckaert, J.-P.; Ciccotti, G. & Berendsen, H. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes *Journal of Computational Physics* **23**, 327–341, (1977).
125 Wang J; Wolf RM; Caldwell JW; Kollman PA & DA., C. Development and testing of a general amber force field. *J Comput Chem* **25**, 1157-1174, (2005).
126 Goddard, T. D.; Kneller, D. G. SPARKY 3; Univeristy of California, San Francisco, CA.
127 Khalaf, A. I.; Pitt, A. R.; Scobie, M.; Suckling, C. J.; Urwin, J.; Waigh, R. D.; Fishleigh, R. V.; Young, S. C. & Wylie, W. A. The synthesis of some head to head linked DNA minor groove binders. *Tetrahedron* **56**, 5225-5239, (2000).
128 Buurma, N. J. & Haq, I. Calorimetric and spectroscopic studies of Hoechst 33258: self-association and binding to non-cognate DNA. *J Mol Biol* **381**, 607-621, (2008).
129 Hampshire, A. J.; Khairallah, H.; Khalaf, A. I.; Ebrahimabadi, A. H.; Waigh, R. D.; Suckling, C. J.; Brown, T. & Fox, K. R. DNA sequence recognition by an imidazole-containing isopropyl-substituted thiazole polyamide (thiazotropsin B). *Bioorg Med Chem Lett* **16**, 3469-3474, (2006).

130 Hwang, T. L. & Shaka, A. J. Water Suppression That Works. Excitation Sculpting Using Arbitrary Wave-Forms and Pulsed-Field Gradients *Journal of Magnetic Resonance, Series A* **112**, 275-279 (1995).

131 Maurer, T. & Kalbitzer, H. R. Indirect Referencing of 31P and 19F NMR Spectra. *J Magn Reson B* **113**, 177-178, (1996).

132 Clark, M.; Cramer, R. D. & van Opdenbosch, N. Molecular mechanics calculations of 1 and 10 were performed using Tripos Force Field without electrostatics. *J. Comput. Chem.*, 982-1012., (1989).

133 Darden, T.; York, D. & Pederson, L. Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089-10092, (1993).

134 Lavery, R. & Sklenar, H. Helical Analysis of Irregular Nucleic Acids. *CurVes 5.2:; Laboratorie de Biochimie Theorique, Institut de Biologie Physico-Chimique, CNRS, Paris, France*, (1997).

135 Harris, S. A.; Gavathiotis, E.; Searle, M. S.; Orozco, M. & Laughton, C. A. Cooperativity in drug-DNA recognition: a molecular dynamics study. *J Am Chem Soc* **123**, 12658-12663, (2001).

136 Kwong, P. D.; Doyle, M. L.; Casper, D. J.; Cicala, C.; Leavitt, S. A.; Majeed, S.; Steenbeke, T. D.; Venturi, M.; Chaiken, I.; Fung, M.; Katinger, H.; Parren, P. W.; Robinson, J.; Van Ryk, D.; Wang, L.; Burton, D. R.; Freire, E.; Wyatt, R.; Sodroski, J.; Hendrickson, W. A. & Arthos, J. HIV-1 evades antibody-mediated neutralization through conformational masking of receptor-binding sites. *Nature* **420**, 678-682, (2002).

137 O'Brien, R. & Haq, I. Applications of Biocalorimetry:Binding, Stability and EnzymeKinetics. *Biocalorimetry 2. Edited by John E. Ladbury and Michael Doyle. John Wiley & Sons, Ltd.* , (2004).

138 Spolar, R. S. & Record, M. T., Jr. Coupling of local folding to site-specific binding of proteins to DNA. *Science* **263**, 777-784, (1994).

139 Borea, P. A.; Varani, K.; Gessi, S.; Gilli, P. & Gilli, G. Binding thermodynamics at the human neuronal nicotine receptor. *Biochem Pharmacol* **55**, 1189-1197, (1998).

140 Chenoweth, D. M. & Dervan, P. B. Allosteric modulation of DNA by small molecules. *Proc Natl Acad Sci U S A* **106**, 13175-13179, (2009).

141 Eftink, M. R.; Anusiem, A. C. & Biltonen, R. L. Enthalpy-entropy compensation and heat capacity changes for protein-ligand interactions: general thermodynamic models and data for the binding of nucleotides to ribonuclease A. *Biochemistry* **22**, 3884-3896, (1983).

142 Clarke, E. C. W. & Glew, D. N. Evaluation of thermodynamic functions from equilibrium constants. *Trans. Faraday Soc.* **62**, 539–547, (1966).

143 Blandamer, M. J. Chemical Equilibria inSolution. In Physical Chemistry Series (Horwood, E. &Kemp, T. J., eds), 1st edit. Ellis Horwood, Chichester., (1992).

144 Buurma, N. J. & Haq, I. Advances in the analysis of isothermal titration calorimetry data for ligand-DNA interactions. *Methods* **42**, 162-172, (2007).

145 Fukada, H. & Takahashi, K. Enthalpy and heat capacity changes for the proton dissociation of various buffer components in 0.1 M potassium chloride. *Proteins* **33**, 159-166, (1998).

146 Trauger, J. W.; Baird, E. E.; Mrksich, M. & Dervan, P. B. Extension of Sequence-Specific Recognition in the MinorGroove of DNA by Pyrrole-Imidazole Polyamides to 9-13 Base Pairs. *J. Am. Chem. Soc.* **118**, 6160-6166, (1996).

147 Anthony, N. G.; Fox, K. R.; Johnston, B. F.; Khalaf, A. I.; Mackay, S. P.; McGroarty, I. S.; Parkinson, J. A.; Skellern, G. G.; Suckling, C. J. & Waigh, R. D. DNA binding of a short lexitropsin. *Bioorg Med Chem Lett* **14**, 1353-1356, (2004).

148 Treesuwan, W.; Wittayanarakul, K.; Anthony, N. G.; Huchet, G.; Alniss, H.; Hannongbua, S.; Khalaf, A. I.; Suckling, C. J.; Parkinson, J. A. & Mackay, S. P. A detailed binding free energy study of 2:1 ligand-DNA complex formation by experiment and simulation. *Phys Chem Chem Phys* **11**, 10682-10693, (2009).

149 IUPAC-IUB Joint Commission on Biochemical Nomenclature (JCBN). Abbreviations and symbols for the description of conformations of polynucleotide chains. Recommendations 1982. *Eur J Biochem* **131**, 9-15, (1983).

150 Mrksich, M.; Wade, W. S.; Dwyer, T. J.; Geierstanger, B. H.; Wemmer, D. E. & Dervan, P. B. Antiparallel side-by-side dimeric motif for sequence-specific recognition in the minor groove of DNA by the designed peptide 1-methylimidazole-2-carboxamide netropsin. *Proc Natl Acad Sci U S A* **89**, 7586-7590, (1992).

151 Dwyer, T. J.; Geierstanger, B. H.; Bathini, Y.; Lown, J. W. & Wemmer, D. E. Design and Binding of a Distamycin A Analog to d(C G C A A G T T GG C) • d(G C C A A C T T G C G): Synthesis, NMR Studies, and Implicationsfor the Design of Sequence-Specific Minor Groove Binding Oligopeptides. *J.Am. Chem. Soc.* **114**, 5911-5919., (1992).

152 Geierstanger, B. H.; Dwyer, T. J.; Bathini, Y., Lown, J. W., & Wemmer, D. E. NMR characterization of a heterocomplex formed by distamycin (Dst) and its analog 2-imidazole-distamycin (2 ImD) with d(CGCAAGTTGGC):d(GCCAACTTGCG): preference for the 1:1:1 2-ImD:Dst:DNA complex over the 2:1 2-ImD:DNA and the 2:1 Dst:DNA complexes. *J. Am. Chem. SOC.* **115**, 4474-4482, (1993).

153 Yoon, C.; Prive, G. G.; Goodsell, D. S. & Dickerson, R. E. Structure of an alternating-B DNA helix and its relationship to A-tract DNA. *Proc Natl Acad Sci U S A* **85**, 6332-6336, (1988).

154 Chuprina, V. P.; Lipanov, A. A.; Fedoroff, O.; Kim, S. G.; Kintanar, A. & Reid, B. R. Sequence effects on local DNA topology. *Proc Natl Acad Sci U S A* **88**, 9087-9091, (1991).

155 Katahira, M.; Sugeta, H. & Kyogoku, Y. A new model for the bending of DNAs containing the oligo(dA) tracts based on NMR observations. *Nucleic Acids Res* **18**, 613-618, (1990).

156 Pelton, J. G. & Wemmer, D. E. Structural characterization of a 2:1 distamycin A.d(CGCAAATTGGC) complex by two-dimensional NMR. *Proc Natl Acad Sci U S A* **86**, 5723-5727, (1989).

157 Fagan, P. & Wemmer, D. E. Cooperative binding of distamycin-A to DNA in the 2-1 mode. *J. Am. Chem. Soc.* **114**, 1080–1081., (1992).

158 Pelton, J. G. & Wemmer, D. E. Binding modes of distamycin A with d(CGCAAATTTGCG)2 determined by two-dimensional NMR. *J. Am. Chem. Soc.* **112**, 1393-1399., (1990).

159 Bunkenborg, J.; Behrens, C. & Jacobsen, J. P. NMR characterization of the DNA binding properties of a novel Hoechst 33258 analogue peptide building block. *Bioconjug Chem* **13**, 927-936, (2002).

160 Roongta, V. A.; Jones, C. R. & Gorenstein, D. G. Effect of distortions in the deoxyribose phosphate backbone conformation of duplex oligodeoxyribonucleotide dodecamers containing GT, GG, GA, AC, and GU base-pair mismatches on 31P NMR spectra. *Biochemistry* **29**, 5245-5258, (1990).

161 Shaffer, P. L.; Jivan, A.; Dollins, D. E.; Claessens, F. & Gewirth, D. T. Structural basis of androgen receptor binding to selective androgen response elements. *Proc Natl Acad Sci U S A* **101**, 4758-4763, (2004).

162 Culig, Z.; Klocker, H.; Bartsch, G. & Hobisch, A. Androgen receptors in prostate cancer. *Endocr Relat Cancer* **9**, 155-170, (2002).

163 Anthony, N. G.; Breen, D.; Donoghue, G.; Khalaf, A. I.; Mackay, S. P.; Parkinson, J. A. & Suckling, C. J. A new synthesis of alkene-containing minor-groove binders and essential hydrogen bonding in binding to DNA and in antibacterial activity. *Org Biomol Chem* **7**, 1843-1850, (2009).

164 Khalaf, A. I.; Pitt AR; Scobie, M.; Suckling CJ; Urwin J; Waigh RD; Young SC; RV, F. & WA, W. The Synthesis of Some Head-to-head Linked DNA Minor Groove Binders. *Tetrahedron* **56** 5225-5239. , (2000).

165 Corey, E. J. & Kwiatkowski, G. T. The Synthesis of Olefins from O,O'-Dialkyl β-Lithioalkylphosphonothioate Esters. *J. Am. Chem. Soc.* **88** 5654–5656, (1966).

166 Horner, L.; Hoffmann, H. & Wippel, H. G. Phosphine oxides as reagents for the olefin formation. . *Chem. Ber* **91**, 61-63., (1958).

167 Barton, A.; Breukelman, S. P.; Kaye, P. T.; Meakins, G. D. & Morgan, D. J. The preparation of thiazole-4- and -5-carboxylates, and an infrared study of their rotational isomers. *J. Chem. Soc. Perkin Trans* **1**, (1982).

168 Khalaf, A. I.; Waigh, R. D.; Drummond, A. J.; Pringle, B.; McGroarty, I.; Skellern, G. G. & Suckling, C. J. Distamycin analogues with enhanced lipophilicity: synthesis and antimicrobial activity. *J Med Chem* **47**, 2133-2156, (2004).

169 Devos, A.; Remion, J.; Frisquehesbain, A. M.; Colens, A. & Ghosez, L. SYNTHESIS OF ACYL HALIDES UNDER VERY MILD CONDITIONS. *J. Chem. Soc. Chem. Commun*, 1180-1181, (1979).

170 Nakatsuji, H.; Morita, J.; Misaki, T. & Tanabe, Y. Water Solvent Method for Esterification and Amide Formation between Acid Chlorides and Alcohols Promoted by Combined Catalytic Amines: Synergy between N-Methylimidazole and N,N,N′,N′-Tetramethylethylenediamine (TMEDA). *Advanced Synthesis & Catalysis* **348**, 2057– 2062, (2006).

171 Himati, F.; Arcamone, F.; Conte, M. R.; Patrizia Felicetti; Aldo Galeone; Paolo Lombardi; Luigi G. Paloma & Rossit., C. Synthesis of Two Distamycin Analogs and Their Binding Mode to d(CGCAAATTTGCG)2 in the 2:1 Solution Complexes as Determined by Two-Dimensional 1H-NMR. *J. Med. Chem.* **38**, 1140-1149, (1995).

172 Brittain, J. M. & Jones, R. A. Pyrrole Studies XXVII. Utilisation of 1-Methyl-2-pyrrolyl Lithium in the Synthesis of 1-Methyl-2-substituted Pyrroles *Synthetic Communications* **12**, 231 - 248 (1982).

173 Gupton, J. T.; Krolikowski, D. A. & Sikorski, J. A. Application of 2-substituted vinamidinium salts to the synthesis of 2,4-disubstituted pyrroles. *J. Org. Chem.* **55** 4735–4740, (1990).

7 Appendices

7.1 Appendix 1

List of NOE assignments together with restraints calculated using MARDIGRAS for the 2:1 complex between thiazotropsin B and D(CGACGCGTCG)2 . DNA bases are represented by letter codes A, T, C and G. The ligands are represented by L21 and L22. Q indicates a pseudo atom whilst M corresponds to a methyl group.

[H–H] NOEs	**Distance (Å)**	**[H–H] NOEs**	**Distance (Å)**	**[H–H] NOEs**	**Distance (Å)**
C1H1′–C1H4′	2.939	A3H2′2–A3H1′	2.645	C4H5–C4H6	2.759
C1H2′2–C1Q5′	4.198	A3H2′1–A3H4′	3.601	G5H4′–G5H5′2	2.390
C1H2′2–C1H4′	2.393	A3H2′1–A3H1′	2.55	G5H4′–G5H5′1	2.592
C1H2′2–C1H1′	2.484	A3H3′–A3Q5′	3.36	G5H2′2–G5H1′	2.604
C1H2′1–C1Q5′	3.705	A3H3′–A3H4′	2.682	G5H2′1–G5H1′	2.830
C1H2′1–C1H4′	3.036	A3H3′–A3H1′	3.371	G5H8–C4Q5′	3.828
C1H2′1–C1H1′	2.941	A3H3′–A3H2′2	3.084	G5H8–C4H1′	3.999
C1H3′–C1Q5′	3.342	A3H3′–A3H2′1	3.22	G5H8–C4H2′2	3.303
C1H3′–C1H1′	4.136	A3H8–G2H1′	3.587	G5H8–C4H2′1	3.018
C1H3′–C1H2′2	2.529	A3H8–G2H2′2	3.024	G5H8–C4H3′	3.808
C1H3′–C1H2′1	2.579	A3H8–G2H8	4.516	G5H8–C4H6	4.598
C1H6–C1Q5′	3.857	A3H8–A3Q5′	4.161	G5H8–G5H5′2	3.769
C1H6–C1H4′	3.877	A3H8–A3H4′	4.625	G5H8–G5H5′1	3.216
C1H6–C1H1′	3.431	A3H8–A3H1′	3.696	G5H8–G5H4′	2.749
C1H6–C1H2′2	3.102	A3H8–A3H2′2	2.448	G5H8–G5H1′	3.767
C1H6–C1H2′1	2.61	A3H8–A3H2′1	2.984	G5H8–G5H2′2	1.813
C1H6–C1H3′	3.597	A3H8–A3H3′	3.435	G5H22–G5H1	3.640
C1H5–C1H2′1	1.762	A3H2–G2H1	6.952	G5H21–G5H1	4.608
C1H5–C1H6	2.618	A3H2–A3H1′	4.607	C6H5′2–G5H1′	3.483
G2Q5′–C1H1′	4.499	C4H1′–A3H2	4.691	C6H4′–C6H5′2	2.481
G2H1′–G2H4′	3.305	C4H1′–C4Q5′	3.207	C6H4′–C6H5′1	2.615
G2H2′2–G2H4′	3.553	C4H2′2–C4Q5′	2.308	C6H1′–G5H1	6.229
G2H2′2–G2H1′	2.743	C4H2′2–C4H1′	2.723	C6H1′–C6H5′2	3.850
G2H2′1–G2H4′	3.383	C4H2′1–C4Q5′	3.365	C6H1′–C6H5′1	3.322
G2H2′1–G2H1′	2.78	C4H3′–C4Q5′	2.914	C6H1′–C6H4′	2.987

G2H3′–G2Q5′	4.142	C4H3′–C4H1′	3.793	C6H2′2–C6H1′	2.773
G2H3′–G2H4′	2.974	C4H3′–C4H2′2	2.342	C6H2′1–C6H4′	2.962
G2H3′–G2H1′	3.613	C4H3′–C4H2′1	2.785	C6H2′1–C6H1′	2.932
G2H3′–G2H2′1	1.803	C4H6–A3H1′	3.515	C6H2′1–C6H2′2	2.624
G2H8–C1H1′	4.267	C4H6–A3H2′2	3.300	C6H3′–C6H5′2	2.956
G2H8–C1H2′2	2.667	C4H6–A3H2′1	3.230	C6H3′–C6H5′1	3.277
G2H8–C1H2′1	3.217	C4H6–A3H3′	4.031	C6H3′–C6H4′	3.038
G2H8–C1H3′	3.255	C4H6–A3H8	4.492	C6H3′–C6H1′	3.476
G2H8–C1H6	4.889	C4H6–C4Q5′	3.841	C6H3′–C6H2′2	2.657
G2H8–G2Q5′	4.503	C4H6–C4H1′	3.596	C6H3′–C6H2′1	2.769
G2H8–G2H4′	4.672	C4H6–C4H2′2	2.345	C6H6–G5H1′	3.458
G2H8–G2H1′	3.825	C4H6–C4H2′1	2.844	C6H6–G5H2′2	3.433
G2H8–G2H2′2	2.936	C4H6–C4H3′	3.645	C6H6–G5H2′1	3.223
G2H8–G2H2′1	2.414	C4H5–A3H1′	3.453	C6H6–C6H5′2	3.426
G2H8–G2H3′	4.152	C4H5–A3H2′2	2.913	C6H6–C6H5′1	3.930
A3Q5′–G2H1′	3.654	C4H5–A3H2′1	3.109	C6H6–C6H4′	3.381
A3H1′–A3Q5′	3.000	C4H5–A3H3′	4.249	C6H6–C6H1′	3.402
A3H1′–A3H4′	3.102	C4H5–A3H8	4.114	C6H6–C6H2′2	3.063
A3H2′2–A3H4′	3.224	C4H5–C4H2′1	3.462	C6H6–C6H2′1	2.717
C6H6–C6H3′	3.228	T8H6–T8H2′1	2.502	G10H2′2–G10H4′	1.514
C6H5–G5H2′2	2.901	T8H6–T8H3′	3.379	G10H2′2–G10H1′	2.673
C6H5–G5H2′1	2.829	T8M5–G7H1′	4.417	G10H2′1–G10H4′	2.459
C6H5–C6H2′1	3.535	T8M5–G7H2′2	3.527	G10H2′1–G10H1′	2.509
C6H5–C6H6	2.755	T8M5–G7H2′1	3.146	G10H3′–G10Q5′	2.193
C6H41–C6H42	3.505	T8M5–G7H3′	4.368	G10H8–C9H1′	4.555
G7H1′–G7Q5′	4.235	T8M5–G7H8	3.689	G10H8–C9H2′1	1.679
G7H2′2–G7H1′	3.176	T8M5–T8H6	3.538	G10H8–G10Q5′	3.548
G7H2′1–G7H1′	2.393	C9H5′2–T8H1′	3.846	G10H8–G10H1′	3.954
G7H3′–G7H4′	3.187	C9H5′2–T8H2′2	3.397	G10H8–G10H2′2	2.260
G7H3′–G7H1′	3.604	C9H5′1–T8H1′	3.748	C11H1′–C11H4′	2.948
G7H3′–G7H2′2	2.978	C9H5′1–T8H2′2	3.2	C11H2′2–C11Q5′	4.197
G7H3′–G7H2′1	2.395	C9H4′–T8H1′	4.389	C11H2′2–C11H4′	2.403
G7H8–C6H1′	3.719	C9H1′–C9H4′	2.967	C11H2′2–C11H1′	2.485
G7H8–C6H2′2	2.926	C9H2′2–C9H1′	2.501	C11H2′1–C11Q5′	3.710
G7H8–C6H2′1	3.109	C9H2′1–C9H1′	2.879	C11H2′1–C11H4′	3.059
G7H8–C6H3′	3.634	C9H3′–C9H5′2	3.55	C11H2′1–C11H1′	2.948

G7H8−C6H6	3.725	C9H3′−C9H5′1	2.774	C11H3′−G10H2′2	2.495
G7H8−G7Q5′	4.651	C9H3′−C9H4′	2.421	C11H3′−C11Q5′	3.410
G7H8−G7H4′	4.253	C9H3′−C9H1′	3.765	C11H3′−C11H1′	4.145
G7H8−G7H1′	3.790	C9H3′−C9H2′2	2.557	C11H3′−C11H2′2	2.534
G7H8−G7H2′2	3.106	C9H3′−C9H2′1	2.83	C11H3′−C11H2′1	2.59
G7H8−G7H2′1	2.631	C9H6−T8H1′	3.786	C11H6−C11Q5′	3.844
G7H8−G7H3′	3.593	C9H6−T8H2′2	2.656	C11H6−C11H4′	3.889
G7H1−C6H42	5.895	C9H6−T8H2′1	3.305	C11H6−C11H1′	3.430
G7H1−C6H41	3.946	C9H6−T8H3′	4.018	C11H6−C11H2′2	3.102
T8Q5′−G7H1′	4.013	C9H6−C9H5′2	3.223	C11H6−C11H2′1	2.615
T8H1′−T8Q5′	4.279	C9H6−C9H5′1	3.898	C11H6−C11H3′	3.600
T8H1′−T8H4′	2.920	C9H6−C9H4′	3.045	C11H5−C11H2′1	1.771
T8H2′2−T8H1′	2.829	C9H6−C9H1′	2.244	C11H5−C11H6	2.620
T8H2′1−T8H4′	2.047	C9H6−C9H2′2	2.909	G12Q5′−C11H1′	4.457
T8H2′1−T8H1′	2.841	C9H6−C9H2′1	2.359	G12H1′−G12H4′	3.300
T8H3′−T8Q5′	3.373	C9H6−C9H3′	3.353	G12H2′2−G12H4′	3.551
T8H3′−T8H4′	3.195	C9H5−T8H1′	3.62	G12H2′2−G12H1′	2.741
T8H3′−T8H1′	3.610	C9H5−T8H2′2	3.46	G12H2′1−G12H4′	3.368
T8H3′−T8H2′2	2.973	C9H5−T8H2′1	3.206	G12H2′1−G12H1′	2.762
T8H3′−T8H2′1	2.516	C9H5−T8H3′	3.95	G12H3′−G12Q5′	4.130
T8H6−G7H1′	3.756	C9H5−T8H6	4.109	G12H3′−G12H4′	2.968
T8H6−G7H2′2	2.955	C9H5−T8M5	4.459	G12H3′−G12H1′	3.624
T8H6−G7H2′1	3.193	C9H5−C9H2′1	3.001	G12H3′−G12H2′1	1.790
T8H6−G7H3′	3.856	C9H5−C9H6	2.924	G12H8−C11H1′	4.266
T8H6−G7H8	3.767	C9H42−T8H3	4.98	G12H8−C11H2′2	2.668
T8H6−T8H1′	3.687	G10H1′−G10Q5′	3.824	G12H8−C11H2′1	3.223
T8H6−T8H2′2	3.601	G10H1′−G10H4′	3.566	G12H8−C11H3′	3.257
G12H8−C11H6	4.875	C14H6−A13H2′2	3.303	C16H2′1−C16H1′	2.915
G12H8−G12Q5′	4.497	C14H6−A13H2′1	3.229	C16H2′1−C16H2′2	2.634
G12H8−G12H4′	4.668	C14H6−A13H3′	4.034	C16H3′−C16H5′2	2.961
G12H8−G12H1′	3.822	C14H6−A13H8	4.489	C16H3′−C16H5′1	3.272
G12H8−G12H2′2	2.939	C14H6−C14Q5′	3.949	C16H3′−C16H4′	3.036
G12H8−G12H2′1	2.400	C14H6−C14H1′	3.591	C16H3′−C16H1′	3.443
G12H8−G12H3′	4.279	C14H6−C14H2′2	2.350	C16H3′−C16H2′2	2.657
G12H1−C9H5	5.816	C14H6−C14H2′1	2.861	C16H3′−C16H2′1	2.781
G12H1−C9H42	4.899	C14H6−C14H3′	3.644	C16H6−G15H1′	3.428

G12H1–C9H41	4.493	C14H5–A13H1′	3.480	C16H6–G15H2′2	3.438
A13Q5′–G12H1′	3.662	C14H5–A13H2′2	2.912	C16H6–G15H2′1	3.213
A13H1′–A13Q5′	3.031	C14H5–A13H2′1	3.100	C16H6–C16H5′2	3.43
A13H1′–A13H4′	3.147	C14H5–A13H3′	4.247	C16H6–C16H5′1	3.963
A13H2′2–A13H4′	3.256	C14H5–A13H8	4.107	C16H6–C16H4′	3.379
A13H2′2–A13H1′	2.666	C14H5–C14H2′1	3.477	C16H6–C16H1′	3.367
A13H2′1–A13H4′	3.622	C14H5–C14H6	2.760	C16H6–C16H2′2	3.062
A13H2′1–A13H1′	2.566	C14H42–G7H1	5.785	C16H6–C16H2′1	2.727
A13H3′–A13Q5′	3.347	C14H41–G7H1	4.353	C16H6–C16H3′	3.227
A13H3′–A13H4′	2.708	G15H4′–G15H5′2	2.379	C16H5–G15H2′2	2.881
A13H3′–A13H1′	3.398	G15H4′–G15H5′1	2.578	C16H5–G15H2′1	2.816
A13H3′–A13H2′2	3.084	G15H2′2–G15H1′	2.578	C16H5–C16H2′1	3.551
A13H3′–A13H2′1	3.212	G15H2′1–G15H1′	2.792	C16H5–C16H6	2.756
A13H8–G12H1′	3.596	G15H8–C14Q5′	3.773	C16H41–C16H42	3.497
A13H8–G12H2′2	3.148	G15H8–C14H1′	3.986	G17H1′–G17Q5′	4.366
A13H8–G12H8	4.882	G15H8–C14H2′2	3.293	G17H2′2–G17H1′	3.162
A13H8–A13Q5′	4.194	G15H8–C14H2′1	3.016	G17H2′1–G17H1′	2.359
A13H8–A13H4′	4.668	G15H8–C14H3′	3.799	G17H3′–G17H4′	3.182
A13H8–A13H1′	3.716	G15H8–C14H6	4.594	G17H3′–G17H1′	3.591
A13H8–A13H2′2	2.440	G15H8–G15H5′2	3.739	G17H3′–G17H2′2	2.984
A13H8–A13H2′1	2.966	G15H8–G15H5′1	3.199	G17H3′–G17H2′1	2.392
A13H8–A13H3′	3.427	G15H8–G15H4′	2.728	G17H8–C16H1′	3.685
A13H2–G7H1	3.940	G15H8–G15H1′	3.779	G17H8–C16H2′2	2.924
A13H2–T8H3	5.193	G15H8–G15H2′2	1.813	G17H8–C16H2′1	3.121
A13H2–A13H1′	4.463	G15H22–G15H1	3.607	G17H8–C16H3′	3.634
C14H1′–A13H2	4.671	G15H21–G15H1	4.605	G17H8–C16H6	3.723
C14H1′–C14Q5′	3.224	C16H5′2–G15H1′	3.407	G17H8–G17Q5′	4.619
C14H2′2–C14Q5′	2.285	C16H4′–C16H5′2	2.482	G17H8–G17H4′	4.280
C14H2′2–C14H1′	2.722	C16H4′–C16H5′1	2.606	G17H8–G17H1′	3.761
C14H2′1–C14Q5′	3.382	C16H1′–G15H1	6.039	G17H8–G17H2′2	3.108
C14H3′–C14Q5′	2.872	C16H1′–C16H5′2	3.828	G17H8–G17H2′1	2.623
C14H3′–C14H1′	3.812	C16H1′–C16H5′1	3.280	G17H8–G17H3′	3.598
C14H3′–C14H2′2	2.345	C16H1′–C16H4′	2.960	G17H1–A3H2	3.943
C14H3′–C14H2′1	2.800	C16H2′2–C16H1′	2.744	G17H1–C4H42	5.789
C14H6–A13H1′	3.546	C16H2′1–C16H4′	2.971	G17H1–C4H41	4.341
G17H1–C16H41	3.928	C19H6–T18H2′1	3.316	L21H2–G5H22	5.080

T18Q5′−G17H1′	4.158	C19H6−T18H3′	4.024	L21H2−G5H21	4.777
T18H1′−T18Q5′	4.103	C19H6−C19H5′2	3.247	L21H2−L21M1	4.656
T18H1′−T18H4′	2.873	C19H6−C19H5′1	3.757	L21H4−C16H5′2	4.266
T18H2′2−T18H1′	2.839	C19H6−C19H4′	3.033	L21H4−C16H5′1	2.757
T18H2′1−T18H4′	2.013	C19H6−C19H1′	2.252	L21H4−C16H4′	3.276
T18H2′1−T18H1′	2.850	C19H6−C19H2′2	2.921	L21H4−C16H3′	4.715
T18H3′−T18Q5′	3.360	C19H6−C19H2′1	2.377	L21M6−G17H1′	4.074
T18H3′−T18H4′	3.145	C19H6−C19H3′	3.354	L21M6−L22H4	3.010
T18H3′−T18H1′	3.615	C19H5−G2H1	5.847	L21H8−C4H5	6.480
T18H3′−T18H2′2	2.973	C19H5−T18H1′	3.621	L21H8−G5H5′2	5.897
T18H3′−T18H2′1	2.526	C19H5−T18H2′2	3.472	L21H8−G5H5′1	6.558
T18H6−G17H1′	3.705	C19H5−T18H2′1	3.218	L21H8−G5H8	5.485
T18H6−G17H2′2	2.963	C19H5−T18H3′	3.956	L21H8−G5H1	6.940
T18H6−G17H2′1	3.097	C19H5−T18H6	4.109	L21H8−G5H22	4.703
T18H6−G17H3′	3.858	C19H5−T18M5	4.555	L21H8−G5H21	4.599
T18H6−G17H8	3.775	C19H5−C19H2′1	3.019	L21H8−C6H1′	3.649
T18H6−T18H1′	3.686	C19H5−C19H6	2.928	L21H8−C6H2′2	6.258
T18H6−T18H2′2	3.434	C19H42−G2H1	5.022	L21H8−C6H2′1	6.326
T18H6−T18H2′1	2.515	C19H42−T18H3	4.941	L21H8−L21H2	6.509
T18H6−T18H3′	3.357	C19H41−G2H1	4.500	L21H10−C6H5	3.351
T18M5−G17H1′	4.340	G20H1′−C1H3′	3.624	L21H10−L21M6	3.771
T18M5−G17H2′2	3.590	G20H1′−G20Q5′	3.826	L21M5−C6H4′	3.596
T18M5−G17H2′1	3.073	G20H1′−G20H4′	3.318	L21M5−C6H1′	3.993
T18M5−G17H3′	4.359	G20H2′2−C1H3′	2.541	L21M5−G5H1′	3.843
T18M5−G17H8	3.784	G20H2′2−G20H4′	1.601	L21M5−G5H2′2	4.170
T18M5−T18H6	3.534	G20H2′2−G20H1′	2.703	L21M5−G5H2′1	4.054
T18H3−A3H2	5.169	G20H2′1−G20H4′	2.25	L21M2−C1H4′	3.376
C19H5′2−T18H1′	3.854	G20H2′1−G20H1′	2.503	L21M2−A3H4′	5.123
C19H5′2−T18H2′2	3.404	G20H3′−G20Q5′	2.193	L21M2−C9H5′1	3.470
C19H5′1−T18H1′	3.72	G20H8−C19H1′	4.524	L21M2−C9H4′	3.123
C19H5′1−T18H2′2	3.017	G20H8−C19H2′1	1.697	L21M2−C9H1′	3.982
C19H4′−T18H1′	4.366	G20H8−G20Q5′	3.532	L21M2−G10H1′	4.698
C19H1′−C19H4′	2.959	G20H8−G20H1′	3.929	L21M2−A13Q5′	3.588
C19H2′2−C19H1′	2.532	G20H8−G20H2′2	2.351	L21M2−A13H1′	4.181
C19H2′1−C19H1′	2.887	L21M1−A3H2	4.166	L21M2−A13H2	4.688
C19H3′−C19H5′2	3.544	L21M1−C4H1′	3.645	L21M2−C14H1′	4.446

C19H3′−C19H5′1	2.782	L21M1−C4H2′2	3.111	L21M3−C1H4′	3.116
C19H3′−C19H4′	2.411	L21M1−G5H1′	3.663	L21M3−A3H4′	4.750
C19H3′−C19H1′	3.762	L21M1−G7Q5′	3.314	L21M3−C9H5′1	3.747
C19H3′−C19H2′2	2.573	L21M1−G17H4′	2.913	L21M3−C9H4′	3.014
C19H3′−C19H2′1	2.836	L21M1−G17H1	7.475	L21M3−C9H1′	3.957
C19H6−T18H1′	3.782	L21H2−G5H1′	4.912	L21M3−G10H1′	4.725
C19H6−T18H2′2	2.672	L21H2−G5H1	6.488	L21M3−A13Q5′	3.844
L21M3−A13H1′	4.265	L21M8−L21M3	5.484	L22H2−L22M1	4.802
L21M3−A13H2	4.659	L21H20−L21M7	2.812	L22H4−C6H5′2	4.218
L21M3−C14H1′	4.021	L21H20−L21M8	2.751	L22H4−C6H5′1	2.765
L21Q25−C9H4′	3.507	L21H14−C6H2′2	5.32	L22H4−C6H4′	3.278
L21Q25−C9H1′	3.948	L21H14−G7H1′	4.229	L22H4−C6H3′	4.709
L21Q25−A13H2	3.688	L21H14−G7H2′2	6.752	L22H4−L21M5	3.219
L21Q25−C14H1′	3.905	L21H14−G7H2′1	5.878	L22H4−L21M7	4.178
L21H241−A13H2	2.837	L21H14−G7H1	6.943	L22M6−G7H1′	4.213
L21H241−L21M2	3.292	L21H14−G7H22	4.798	L22M6−L22H10	2.407
L21H241−L21M3	2.842	L21H14−T8Q5′	6.688	L22M6−L21H13	3.423
L21H241−L21Q25	2.713	L21H14−T8H1′	5.853	L22M6−L22H4	3.004
L21H242−A13H2	3.142	L21H14−T8H6	6.92	L22H8−C14H5	6.414
L21H242−L21M2	3.732	L21H14−G15H4′	5.519	L22H8−G15H5′2	5.881
L21H242−L21M3	3.586	L21H14−G17H21	4.406	L22H8−G15H8	5.417
L21H242−L21Q25	2.988	L21H14−L21H8	6.589	L22H8−G15H22	4.529
L21H231−A13H2	3.086	L21H13−C6H1′	4.238	L22H8−G15H21	4.48
L21H231−C19H1′	3.545	L21H13−G7H1′	3.134	L22H8−C16H1′	3.568
L21H231−L21H241	2.778	L21H13−G15H21	4.501	L22H8−C16H2′2	6.156
L21H231−L21H242	2.782	L21H13−G17H22	3.792	L22H8−C16H2′1	5.93
L21H232−G7H3′	2.831	L21H13−L21H8	4.310	L22H8−C16H6	6.955
L21H232−C9H4′	3.265	L21H13−L21H10	3.729	L22H8−L21H8	2.808
L21H232−A13H2	2.841	L21H13−L21H14	3.979	L22H8−L21M5	7.905
L21H232−C14H1′	3.753	L22M1−G7H4′	2.945	L22H8−L22H2	6.117
L21H232−C16H1′	3.297	L22M1−G7H1	7.752	L22H10−C16H5	3.344
L21H232−C19H1′	3.203	L22M1−A13H2	4.12	L22H10−L21M6	2.896
L21H232−L21M2	3.262	L22M1−C14H1′	3.463	L22H10−L22M6	3.76
L21H232−L21M3	3.536	L22M1−C14H2′2	2.605	L22M5−G15H1′	3.781
L21H232−L21H241	1.936	L22M1−G15H1′	3.238	L22M5−G15H2′2	4.183
L21H22−T8H2′2	6.055	L22M1−G17Q5′	3.541	L22M5−G15H2′1	4.054

L21H22−T18H1′	4.145	L22M1−L22M5	4.243	L22M5−C16H4′	3.485
L21H22−L21H231	5.326	L22M1−L21M2	4.702	L22M5−C16H1′	3.919
L21H22−L21H232	4.233	L22M1−L21M3	4.825	L22M5−L22M1	4.386
L21M7−G7Q5′	3.789	L22M1−L21Q25	3.809	L21M5−L21M1	4.386
L21M7−T8Q5′	4.774	L22M1−L21H231	3.909	L22M5−L22H4	2.373
L21M7−T8H2′2	4.225	L22M1−L21H232	3.726	L22M2−A3Q5′	3.597
L21M7−C9H5′2	3.615	L22M1−L21H22	6.796	L22M2−A3H1′	4.246
L21M7−C14Q5′	3.524	L22H2−G15H1′	4.858	L22M2−A3H2	4.500
L21M7−T18Q5′	4.880	L22H2−G15H1	6.36	L22M2−C4H1′	4.247
L21M7−T18H4′	3.891	L22H2−G15H22	5.018	L22M2−C11H4′	3.298
L21M6−L21H10	2.407	L22H2−G15H21	4.746	L22M2−A13H4′	5.009
L21M7−L21M2	5.210	L22H2−L21H8	5.400	L22M2−C19H5′1	3.748
L21M7−L21M3	5.615	L22H2−L22M5	5.825	L22M2−C19H4′	3.406
L21M8−L21M5	4.812	L22H2−L21H14	3.891	L22M2−C19H1′	3.767
L21M8−L21M2	5.242	L22H2−L21H13	4.886	L22M2−G20H1′	4.559
L22M2−L21M1	4.903	L22H242−L22M3	3.842	L22H232−L22H241	1.891
L22M3−A3Q5′	3.800	L22H242−L22Q25	3.12	L22H22−T8H1′	4.181
L22M3−A3H1′	4.410	L22H231−A3H2	3.064	L22H22−T18H2′2	5.819
L22M3−A3H2	4.665	L22H231−C9H1′	3.527	L22H22−T18H2′1	6.871
L22M3−C4H1′	4.138	L22H231−L21M1	3.836	L22H22−L21M1	6.435
L22M3−A13H4′	5.059	L22H231−L22H241	2.757	L22H22−L22H242	5.749
L22M3−C19H5′1	3.567	L22H231−L22H242	2.779	L22H22−L22H231	4.532
L22M3−C19H4′	3.144	L22H232−A3H2	2.761	L22M7−T18Q5′	4.415
L22M3−C19H1′	3.829	L22H232−C4H1′	3.719	L22M7−T18H2′2	4.011
L22M3−G20H1′	4.535	L22H232−C6H1′	3.28	L22M7−C19H5′2	3.291
L22M3−L21M1	4.632	L22H232−C9H1′	3.135	L22M7−L21H4	3.823
L22Q25−A3H2	3.712	L22H232−G17H3′	2.79	L22M7−L22M2	5.94
L22Q25−C4H1′	3.562	L22H232−C19H4′	3.167	L22M7−L22M3	5.638
L22Q25−C19H4′	3.75	L22H232−L21M1	3.585	L22M8−L22M2	5.249
L22Q25−C19H1′	4.225	L22H232−L22M2	3.195	L22M8−L22M3	4.886
L22Q25−L21M1	4.176	L22H232−L22M3	3.532	L22H20−L22M7	2.803
L22H241−A3H2	2.903	L22H22−L22H232	4.775	L22H20−L22M8	2.731
L22H241−L22M2	2.965	L22M7−C4Q5′	3.844	L22H14−G5H4′	5.577
L22H241−L22M3	3.110	L22M7−T8Q5′	4.613	L22H14−G7H21	4.378
L22H241−L22Q25	2.615	L22M7−T8H4′	3.747	L22H14−C16H2′2	5.265
L22H242−L22M2	3.522	L22M7−G17Q5′	3.553	L22H14−G17H1′	4.149

L22H14−T18Q5′	6.301	L22H14−L21H2	6.856	L22H14−G17H2′2	6.350
L22H14−T18H1′	6.236	L22H14−L22H8	6.133	L22H14−G17H2′1	5.726
L22H14−L21M1	7.705	L22H13−G5H21	5.013	L22H14−G17H22	4.949
L22H13−G7H22	3.666	L22H13−L22H8	4.583	L21M5−L21H4	2.373
L22H13−C16H1′	4.128	L22H13−L22H10	3.672	L22H13−L22H14	3.917
L22H13−G17H1′	2.659	L22H13−L21M6	3.270		

7.2 Appendix 2

List of NOE assignments together with restraints calculated using MARDIGRAS for the 2:1 complex between AIK18-51 and d(CGACTAGTCG)$_2$. DNA bases are represented by letter codes A, T, C and G. The ligands are represented by L21 and L22. Q indicates a pseudo atom whilst M corresponds to a methyl group.

[H–H] NOEs	Distance (Å)	[H–H] NOEs	Distance (Å)	[H–H] NOEs	Distance (Å)
C1H1′–C1H4′	3.277	A3H8–G2H1′	3.735	T5H3′–T5H4′	2.816
C1H6–C1Q5′	3.701	A3H8–G2H8	4.55	T5H2′2–T5H3′	2.215
C1H6–C1H4′	4.174	A3H8–G2H2′2	3.024	T5H2′1–T5H5′1	2.943
C1H6–C1H3′	3.115	A3H8–G2H2′1	3.99	T5H2′1–T5H4′	3.881
C1H6–C1H1′	3.147	A3H8–A3H5′2	4.56	T5H2′1–T5H3′	2.508
C1H6–C1H5	2.409	A3H8–A3H5′1	4.794	T5H1′–T5H5′2	2.867
C1H2′2–C1H3′	2.512	A3H8–A3H4′	5.42	T5H1′–T5H5′1	2.501
C1H2′2–C1H1′	2.785	A3H8–A3H3′	4.018	T5H1′–T5H4′	2.748
C1H2′2–C1H6	2.971	A3H8–A3H1′	3.345	T5H1′–T5H2′2	1.408
C1H2′1–C1H4′	4.331	A3H2–A3H1′	4.087	T5H1′–T5H2′1	2.56
C1H2′1–C1H1′	3.151	A3H2′2–A3H4′	2.785	T5H3–C4H42	5.388
C1H2′1–C1H5	2.994	A3H2′2–A3H3′	2.54	T5H3–C4H41	3.647
C1H2′1–C1H6	2.637	A3H2′2–A3H1′	2.229	T5M5–C4H5′1	3.959
C1H2′1–C1H2′2	2.348	A3H2′2–A3H8	2.455	T5M5–C4H3′	3.704
G2H3′–G2H5′2	3.382	A3H2′1–A3H3′	2.934	T5M5–C4H1′	3.688
G2H3′–G2H5′1	2.882	A3H2′1–A3H1′	2.291	T5M5–C4H6	3.55
G2H3′–G2H4′	2.939	A3H2′1–A3H8	3.303	T5M5–C4H2′2	3.594
G2H1′–G2H4′	3.299	C4H1′–A3H5′1	3.001	T5H6–C4H3′	5.138
G2H1′–G2H3′	3.637	C4H1′–A3H1′	3.102	T5H6–C4H1′	2.424
G2H8–C1H3′	4.562	C4H1′–A3H2	4.285	T5H6–C4H6	3.462
G2H8–C1H2′2	3.475	C4H1′–C4H3′	3.688	T5H6–C4H2′1	2.107
G2H8–C1H2′1	3.638	C4H41–C4H42	2.738	T5H6–T5H5′2	3.694
G2H8–G2H5′2	4.823	C4H5–A3H1′	1.524	T5H6–T5H5′1	3.558
G2H8–G2H5′1	3.789	C4H5–A3H2′2	1.796	T5H6–T5H4′	4.029
G2H8–G2H4′	3.409	C4H5–A3H2′1	2.878	T5H6–T5H3′	4.375
G2H8–G2H3′	3.901	C4H5–C4H42	3.122	T5H6–T5H2′2	2.617
G2H8–G2H1′	3.935	C4H5–C4H41	3.814	T5H6–T5H2′1	2.489
G2H1–G2H8	5.033	C4H6–A3H8	5.019	T5H6–T5H1′	3.664
G2H2′2–G2H4′	3.978	C4H6–A3H2′2	3.348	T5H6–T5M5	3.048

G2H2′2−G2H3′	2.825	C4H6−A3H2′1	3.422	A6H4′−A6H5′2	2.680
G2H2′2−G2H1′	2.604	C4H6−C4H4′	2.975	A6H4′−A6H5′1	2.617
G2H2′2−G2H8	3.418	C4H6−C4H3′	3.146	A6H3′−A6H5′2	3.137
G2H2′1−G2H4′	3.665	C4H6−C4H1′	3.217	A6H3′−A6H4′	2.458
G2H2′1−G2H3′	2.597	C4H6−C4H5	2.357	A6H1′−T5H5′2	2.95
G2H2′1−G2H1′	3.006	C4H2′2−C4H4′	3.633	A6H1′−A6H4′	3.131
G2H2′1−G2H8	2.487	C4H2′2−C4H3′	3.107	A6H8−T5H2′2	2.289
A3H5′2−G2H1′	3.174	C4H2′2−C4H1′	2.389	A6H8−T5H2′1	3.439
A3H3′−A3H5′2	3.197	C4H2′2−C4H6	3.675	A6H8−T5H3	2.719
A3H3′−A3H5′1	2.828	C4H2′1−C4H4′	2.609	A6H8−T5H6	4.415
A3H3′−A3H4′	2.894	C4H2′1−C4H3′	2.589	A6H8−A6H5′2	3.576
A3H1′−A3H5′1	3.541	C4H2′1−C4H1′	3.276	A6H8−A6H3′	3.515
A3H1′−A3H4′	3.095	C4H2′1−C4H6	2.76	A6H8−A6H1′	4.071
A3H1′−A3H3′	3.225	T5H4′−T5H5′2	2.605	A6H2−A6H1′	4.454
A3H8−G2H3′	4.507	T5H4′−T5H5′1	2.666	A6H2′2−A6H3′	4.007
A6H2′2−A6H1′	2.722	T8H1′−T8H4′	2.771	G10H8−G10Q5′	4.239
A6H2′2−A6H8	2.97	T8H1′−T8H3′	3.369	G10H8−G10H4′	4.446
A6H2′1−A6H3′	2.814	T8H1′−T8H2′2	2.455	G10H8−G10H3′	3.878
A6H2′1−A6H1′	3.028	T8H1′−T8H2′1	2.882	G10H8−G10H1′	3.865
A6H2′1−A6H8	2.757	T8M5−G7H3′	5.999	G10H2′2−G10H4′	3.986
A6H2′1−A6H2′2	2.236	T8M5−G7H8	3.633	G10H2′2−G10H1′	2.292
G7H5′2−A6H8	4.436	T8M5−G7H2′2	3.691	G10H2′2−G10H8	3.162
G7H5′1−A6H1′	3.093	T8M5−G7H2′1	3.661	G10H2′1−G10H4′	3.776
G7H4′−G7H5′2	2.545	T8H6−T5H6	3.611	G10H2′1−G10H1′	3.612
G7H4′−G7H5′1	2.567	T8H6−G7H3′	4.847	G10H2′1−G10H8	2.562
G7H3′−G7H5′2	4.002	T8H6−G7H1′	3.892	C11H1′−C11H4′	3.276
G7H3′−G7H4′	2.074	T8H6−G7H2′2	2.743	C11H6−C11Q5′	3.724
G7H1′−A6H2	4.437	T8H6−G7H2′1	3.78	C11H6−C11H4′	4.177
G7H1′−G7H4′	3.061	T8H6−T8H4′	2.156	C11H6−C11H3′	3.113
G7H1′−G7H3′	3.503	T8H6−T8H3′	3.723	C11H6−C11H1′	3.145
G7H8−A6H1′	4.14	T8H6−T8H2′2	3.392	C11H6−C11H5	2.41
G7H8−A6H8	4.249	T8H6−T8H2′1	2.655	C11H2′2−C11H3′	2.51
G7H8−A6H2′2	2.81	T8H6−T8H1′	4.208	C11H2′2−C11H1′	2.783
G7H8−A6H2′1	3.116	T8H6−T8H3	3.497	C11H2′2−C11H6	2.97
G7H8−G7H5′2	3.744	T8H6−T8M5	3.113	C11H2′1−C11H4′	4.343
G7H8−G7H3′	3.77	C9H1′−C9H4′	3.11	C11H2′1−C11H1′	3.15

G7H8–G7H1′	3.883	C9H41–C9H42	2.779	C11H2′1–C11H5	2.995
G7H1–A6H2	3.762	C9H5–T8H2′2	2.855	C11H2′1–C11H6	2.636
G7H22–G7H1	3.784	C9H5–T8H2′1	3.608	C11H2′1–C11H2′2	2.348
G7H21–G7H1	3.668	C9H5–T8M5	4.022	G12H3′–G12H5′2	3.38
G7H21–G7H22	2.617	C9H5–T8H6	3.483	G12H3′–G12H5′1	2.881
G7H2′2–G7H3′	2.542	C9H5–C9H42	3.415	G12H3′–G12H4′	2.938
G7H2′2–G7H1′	2.608	C9H5–C9H41	4.012	G12H1′–G12H4′	3.305
G7H2′2–G7H8	2.875	C9H6–T8H2′1	2.581	G12H1′–G12H3′	3.645
G7H2′1–G7H3′	2.667	C9H6–T8H1′	2.399	G12H8–C11H3′	4.56
G7H2′1–G7H1′	2.969	C9H6–C9H1′	3.629	G12H8–C11H2′2	3.474
G7H2′1–G7H8	2.653	C9H6–C9H5	2.71	G12H8–C11H2′1	3.64
T8H5′2–G7H1′	2.784	C9H2′2–C9H4′	4.17	G12H8–G12H5′2	4.809
T8H4′–T8H5′2	2.58	C9H2′2–C9H1′	2.706	G12H8–G12H5′1	3.79
T8H4′–T8H5′1	2.651	C9H2′2–C9H6	3.064	G12H8–G12H4′	3.405
T8H3′–T8H5′2	1.729	C9H2′1–C9H1′	3.226	G12H8–G12H3′	3.901
T8H3′–T8H5′1	1.706	C9H2′1–C9H5	3.352	G12H8–G12H1′	3.943
T8H3′–T8H4′	3.066	C9H2′1–C9H6	2.415	G12H1–C9H42	4.183
T8H2′2–T8H4′	3.242	G10H3′–G10H4′	3.455	G12H1–C9H41	3.915
T8H2′2–T8H3′	2.968	G10H1′–G10H4′	3.425	G12H1–G12H8	5.017
T8H2′1–T8H4′	3.632	G10H1′–G10H3′	3.458	G12H2′2–G12H4′	3.986
T8H2′1–T8H3′	2.627	G10H8–C9H1′	3.987	G12H2′2–G12H3′	2.83
T8H1′–T5M5	3.108	G10H8–C9H2′2	3.608	G12H2′2–G12H1′	2.617
T8H1′–T8H5′1	2.783	G10H8–C9H2′1	3.514	G12H2′2–G12H8	3.417
G12H2′1–G12H4′	3.662	C14H5–C14H41	3.769	G15H6–G15H4′	3.995
G12H2′1–G12H3′	2.599	C14H6–T8H6	4.087	G15H6–G15H3′	4.35
G12H2′1–G12H1′	3.002	C14H6–A13H8	4.936	G15H6–G15H2′2	2.59
G12H2′1–G12H8	2.488	C14H6–A13H2′2	3.422	G15H6–G15H2′1	2.476
A13H5′2–G12H1′	3.188	C14H6–A13H2′1	3.372	G15H6–G15H1′	3.53
A13H3′–A13H5′2	3.202	C14H6–C14H4′	2.445	G15H6–G15M5	3.09
A13H3′–A13H5′1	2.797	C14H6–C14H3′	3.135	C16H4′–C16H5′2	2.674
A13H3′–A13H4′	2.855	C14H6–C14H1′	3.162	C16H4′–C16H5′1	2.601
A13H1′–A13H5′1	3.518	C14H6–C14H5	2.315	C16H3′–C16H5′2	3.036
A13H1′–A13H4′	3.054	C14H2′2–G7H5′2	2.649	C16H3′–C16H4′	2.468
A13H1′–A13H3′	3.157	C14H2′2–C14H3′	3.126	C16H1′–G15H5′2	2.94
A13H8–G12H3′	4.502	C14H2′2–C14H1′	2.301	C16H1′–C16H4′	3.145
A13H8–G12H1′	3.747	C14H2′2–C14H6	3.71	C16H8–G15H2′2	2.298

A13H8–G12H8	4.541	C14H2′1–C14H4′	1.966	C16H8–G15H2′1	3.40
A13H8–G12H2′2	3.024	C14H2′1–C14H3′	2.532	C16H8–G15H3	2.709
A13H8–G12H2′1	4.071	C14H2′1–C14H1′	3.375	C16H8–G15H6	4.394
A13H8–A13H5′2	4.479	C14H2′1–C14H6	2.717	C16H8–C16H5′2	3.615
A13H8–A13H5′1	4.768	G15H4′–G15H5′2	2.579	C16H8–C16H4′	6.516
A13H8–A13H4′	5.408	G15H4′–G15H5′1	2.671	C16H8–C16H3′	3.521
A13H8–A13H3′	4.021	G15H3′–G15H4′	2.806	C16H8–C16H1′	4.082
A13H8–A13H1′	3.286	G15H2′2–G15H3′	2.208	C16H2–C16H1′	4.439
A13H2–T8H3	3.862	G15H2′1–G15H5′1	2.987	C16H2′2–C16H3′	4.019
A13H2–A13H1′	4.143	G15H2′1–G15H4′	3.889	C16H2′2–C16H1′	2.728
A13H2′2–A13H4′	2.761	G15H2′1–G15H3′	2.495	C16H2′2–C16H8	2.967
A13H2′2–A13H3′	2.513	G15H1′–G15H5′2	2.811	C16H2′1–C16H3′	2.839
A13H2′2–A13H1′	2.29	G15H1′–G15H5′1	2.513	C16H2′1–C16H1′	3.056
A13H2′2–A13H8	2.449	G15H1′–G15H4′	2.74	C16H2′1–C16H8	2.783
A13H2′1–A13H3′	2.903	G15H1′–G15H2′2	1.408	C16H2′1–C16H2′2	2.255
A13H2′1–A13H1′	2.244	G15H1′–G15H2′1	2.535	G17H5′2–C4H2′2	2.682
A13H2′1–A13H8	3.297	G15H3–C14H42	5.101	G17H5′2–C16H8	4.218
C14H1′–A13H5′1	2.877	G15H3–C14H41	3.724	G17H5′1–C16H1′	3.212
C14H1′–A13H1′	2.962	G15M5–G7H1	5.939	G17H4′–G17H5′2	2.563
C14H1′–A13H2	4.21	G15M5–C14H5′1	3.748	G17H4′–G17H5′1	2.564
C14H1′–C14H3′	3.603	G15M5–C14H3′	3.41	G17H3′–G17H5′2	3.709
C14H42–G7H22	3.992	G15M5–C14H1′	3.445	G17H3′–G17H4′	2.075
C14H42–G7H21	5.033	G15M5–C14H6	3.72	G17H1′–C16H2	4.466
C14H41–G7H22	2.732	G15M5–C14H2′2	3.328	G17H1′–G17H4′	3.064
C14H41–G7H21	2.174	G15H6–C14H3′	6.528	G17H1′–G17H3′	3.508
C14H41–C14H42	2.743	G15H6–C14H1′	2.319	G17H8–C16H1′	4.169
C14H5–A13H1′	1.494	G15H6–C14H6	3.475	G17H8–C16H8	4.246
C14H5–A13H2′2	1.766	G15H6–C14H2′1	2.043	G17H8–C16H2′2	2.807
C14H5–A13H2′1	2.973	G15H6–G15H5′2	3.597	G17H8–C16H2′1	3.14
C14H5–C14H42	3.09	G15H6–G15H5′1	3.555	G17H8–G17H5′2	3.711
G17H8–G17H3′	3.774	T18H6–T18H4′	2.151	L21H16–A6H2	2.749
G17H8–G17H1′	3.885	T18H6–T18H3′	3.75	L21H16–G7H5′1	4.129
G17H1–T5M5	5.952	T18H6–T18H2′2	3.396	L21H16–G7H4′	3.587
G17H1–C16H2	3.753	T18H6–T18H2′1	2.651	L21H16–G7H1′	2.847
G17H22–C4H42	3.963	T18H6–T18H1′	4.227	L21H16–G7H1	5.172
G17H22–C4H41	2.715	T18H6–T18H3	3.498	L21H16–G17H4′	3.434

G17H22–G17H1	3.762	T18H6–T18M5	3.083	L21H16–G17H1′	2.732
G17H21–C4H42	5.015	C19H1′–C19H4′	3.076	L21H21–A6H2	6.777
G17H21–C4H41	2.164	C19H42–G2H1	4.209	L21H21–G15H3	5.115
G17H21–G17H1	3.616	C19H41–G2H1	3.912	L21H21–G17H4′	5.41
G17H21–G17H22	2.616	C19H41–C19H42	2.767	L21H21–G17H1′	3.591
G17H2′2–G17H3′	2.547	C19H5–T18H2′2	2.846	L21H21–G17H2′2	4.66
G17H2′2–G17H1′	2.613	C19H5–T18H2′1	3.621	L21H21–T18H5′2	4.28
G17H2′2–G17H8	2.875	C19H5–T18M5	4.052	L21H21–T18H4′	4.262
G17H2′1–G17H3′	2.662	C19H5–T18H6	3.482	L21H21–T18H1′	4.715
G17H2′1–G17H1′	2.959	C19H5–C19H42	3.406	L21H21–L21H16	3.321
G17H2′1–G17H8	2.647	C19H5–C19H41	4.003	L21M4–C9H4′	4.682
T18H5′2–G17H1′	2.788	C19H6–T18H2′1	2.573	L21M4–C14H5′1	3.648
T18H4′–T18H5′2	2.577	C19H6–T18H1′	2.392	L21M4–T18H4′	4.727
T18H4′–T18H5′1	2.655	C19H6–C19H1′	3.601	L21M4–L21H25	2.972
T18H3′–T18H5′2	1.733	C19H6–C19H5	2.706	L21M3–L21H25	2.843
T18H3′–T18H5′1	1.719	C19H2′2–C19H4′	4.126	L21H29–T8H1′	3.326
T18H3′–T18H4′	3.025	C19H2′2–C19H1′	2.693	L21Q30–C9H1′	2.998
T18H2′2–T18H4′	3.228	C19H2′2–C19H6	3.049	L21Q30–A13H2	3.703
T18H2′2–T18H3′	2.967	C19H2′1–C19H1′	3.202	L21Q30–C14H1′	3.772
T18H2′1–T18H4′	3.646	C19H2′1–C19H5	3.359	L21Q30–L21H29	3.938
T18H2′1–T18H3′	2.636	C19H2′1–C19H6	2.418	L21Q31–T8H6	4.686
T18H1′–G15M5	3.191	G20H3′–G20H4′	3.486	L21Q31–C9H4′	3.985
T18H1′–T18H5′1	2.778	G20H1′–G20H4′	3.46	L21Q31–A13H2	2.971
T18H1′–T18H4′	2.764	G20H1′–G20H3′	3.436	L21Q31–C14H1′	3.37
T18H1′–T18H3′	3.335	G20H8–C19H1′	3.966	L21Q31–G20H2′2	2.242
T18H1′–T18H2′2	2.444	G20H8–C19H2′2	3.613	L21Q31–L21H29	4.36
T18H1′–T18H2′1	2.885	G20H8–C19H2′1	3.513	L21Q31–L21Q30	3.145
T18H3–A3H2	3.813	G20H8–G20Q5′	4.123	L21Q32–C9H1′	3.458
T18M5–G17H3′	6.009	G20H8–G20H4′	4.461	L21Q32–A13H2	2.756
T18M5–G17H8	3.672	G20H8–G20H3′	3.862	L21Q32–C14H1′	4.506
T18M5–G17H2′2	3.73	G20H8–G20H1′	3.861	L21Q32–L21H29	3.656
T18M5–G17H2′1	3.629	G20H2′2–G20H4′	4.24	L21Q32–L21Q31	2.395
T18H6–C4H6	4.104	G20H2′2–G20H1′	2.426	L21H33–T8H3	3.819
T18H6–G15H6	3.588	G20H2′2–G20H8	3.396	L21H33–G12H1	4.141
T18H6–G17H3′	4.856	G20H2′1–G20H4′	3.812	L21M6–C9H4′	4.172
T18H6–G17H1′	3.893	G20H2′1–G20H1′	3.536	L21M6–C9H1′	3.916

T18H6–G17H2′2	2.747	G20H2′1–G20H8	2.57	L21M6–C11H4′	3.487
T18H6–G17H2′1	3.745	L21H16–A6H1′	3.893	L21M6–A13H1′	3.089
L21M6–A13H2	3.726	L21H1–A13H5′1	2.949	L22H21–C16H2	6.153
L21M6–C14H1′	3.181	L21H1–C14H1′	2.346	L22H21–L21H2	4.313
L21M6–L21Q31	3.076	L21H1–G15H5′2	2.56	L22H21–L21H3	4.153
L21M7–C9H4′	3.759	L21H1–G15H5′1	2.928	L22H21–L22H16	3.314
L21M7–C9H1′	3.839	L21H1–G15H4′	2.881	L22H25–L21H5	3.277
L21M7–C11H4′	3.285	L21H1–L21H5	2.766	L22H25–L21H1	3.51
L21M7–A13H1′	3.648	L21H2–C4H1′	2.633	L22M4–C4H5′1	3.744
L21M7–A13H2	4.003	L21H2–T5H3	5.131	L22M4–T8H4′	4.794
L21M7–C14H1′	3.513	L21H2–C14H1′	2.631	L22M4–C19H4′	4.079
L21M7–L21Q31	3.834	L21H2–C14H2′2	2.782	L22M4–L21H5	6.049
L21M2–L21H16	4.953	L21H2–C14H2′1	2.78	L22M4–L22H25	2.857
L21M2–L21M4	4.229	L21H2–L21H7	2.085	L22M3–L21H5	3.395
L21H18–G7H5′2	3.706	L21H2–L21H1	1.684	L22M3–L21H1	3.722
L21H18–G7H4′	3.362	L21H3–C4H1′	2.92	L22M3–L22H25	2.841
L21H18–G17H4′	3.334	L21H3–T5H3	4.655	L22H29–T18H1′	3.3
L21H18–L21H16	3.846	L21H3–G15H3′	2.827	L22H29–L21H2	3.862
L21H18–L21M2	2.871	L21H3–C16H2	3.753	L22Q30–A3H2	3.643
L21H14–A6H2	3.701	L21H3–G17H1	3.18	L22Q30–C4H1′	3.838
L21H14–G15H5′2	4.514	L21H3–L21H7	2.562	L22Q30–C19H1′	2.854
L21H14–C16H4′	5.346	L21H3–L21H5	3.232	L22Q30–L21H1	3.073
L21H14–C16H1′	3.444	L21H3–L21H1	2.372	L22Q30–L22H29	3.896
L21H14–C16H2′2	5.59	L21H9–A6H4′	3.025	L22Q31–A3H2	2.831
L21H14–G17H4′	5.384	L21H9–A6H1′	2.994	L22Q31–C4H1′	3.117
L21H14–L21H16	3.705	L21H9–A6H2	2.797	L22Q31–G10H2′2	2.067
L21H14–L21M2	7.193	L21H9–L21H14	3.227	L22Q31–T18H6	4.264
L21M1–A6H4′	3.978	L21H9–L21H11	5.566	L22Q31–C19H4′	3.625
L21M1–L21H14	5.997	L21H9–L21H7	3.533	L22Q31–L21H1	1.869
L21H11–A6H4′	3.204	L21H9–L21H3	3.332	L22Q31–L21H2	2.847
L21H11–C16H5′2	3.942	L22H16–G7H4′	3.436	L22Q31–L22H29	4.507
L21H11–L21M1	3.012	L22H16–G7H1′	2.732	L22Q31–L22Q30	2.622
L21H7–T5H3′	3.259	L22H16–C16H1′	3.896	L22Q32–A3H2	2.221
L21H7–T5H2′2	4.066	L22H16–C16H2	2.754	L22Q32–C4H1′	2.853
L21H7–C16H5′2	4.555	L22H16–G17H5′1	4.145	L22Q32–C19H1′	2.709
L21H7–C16H4′	5.884	L22H16–G17H4′	3.589	L22Q32–L21H5	4.619

L21H7–C16H2	4.49	L22H16–G17H1′	2.849	L22Q32–L21H2	2.566
L21H7–L21H14	4.452	L22H16–G17H1	5.174	L22Q32–L21H3	4.26
L21H7–L21H11	6.411	L22H16–L21H9	3.582	L22Q32–L22H29	3.176
L21H5–C4H5′2	3.765	L22H21–T5H3	5.026	L22Q32–L22Q31	1.502
L21H5–G15H4′	3.262	L22H21–G7H4′	5.401	L22H33–G2H1	3.946
L21H5–G15H3′	2.558	L22H21–G7H1′	3.588	L22H33–T18H3	3.853
L21H5–C16H5′2	3.213	L22H21–G7H2′2	4.655	L22M6–C1H4′	3.306
L21H5–L21H11	4.516	L22H21–T8H5′2	4.275	L22M6–A3H1′	3.423
L21H1–C4H2′2	2.033	L22H21–T8H4′	4.267	L22M6–A3H2	3.961
L21H1–C4H2′1	2.627	L22H21–T8H1′	4.72	L22M6–C4H1′	3.536
L22M6–C19H4′	3.999	L22H7–G15H2′1	5.261	L22H3–L22H1	3.085
L22M6–C19H1′	3.98	L22H7–L22H14	4.42	L22H9–C16H4′	3.03
L22M6–L21H1	6.269	L22H7–L22H11	6.589	L22H9–C16H1′	2.993
L22M6–L22Q31	3.262	L22H5–T5H4′	3.245	L22H9–C16H2	2.805
L22M7–C1H4′	3.107	L22H5–T5H3′	2.556	L22H9–L21H16	3.603
L22M7–A3H1′	3.264	L22H5–A6H5′2	3.219	L22H9–L21M2	3.668
L22M7–A3H2	3.585	L22H5–C14H5′2	3.739	L22H9–L21H18	4.543
L22M7–C4H1′	3.491	L22H5–L21H25	3.257	L22H9–L22H14	3.223
L22M7–C19H4′	3.943	L22H5–L21M4	4.772	L22H9–L22H11	5.544
L22M7–C19H1′	3.703	L22H5–L21M3	2.992	L22H9–L22H7	3.484
L22M7–L22Q31	2.671	L22H5–L21Q32	5.179	L22H9–L22H3	3.625
L22M2–L21H11	3.777	L22H5–L22H11	4.521	L22M1–L22H14	5.931
L22M2–L21H9	3.717	L22H1–A3H5′1	3.132	L22H11–A6H5′2	3.974
L22M2–L22H16	4.944	L22H1–C4H1′	2.613	L22H11–C16H4′	3.224
L22M2–L22M4	4.417	L22H1–T5H5′2	2.734	L22H11–L21M2	3.647
L22H18–G7H4′	3.335	L22H1–T5H5′1	3.059	L22H11–L21H18	4.76
L22H18–G17H5′2	3.76	L22H1–T5H4′	3.028	L22H11–L22M1	3.009
L22H18–G17H4′	3.361	L22H1–C14H2′2	2.077	L22H7–A6H5′2	4.565
L22H18–L21M1	4.009	L22H1–L21H25	3.706	L22H7–A6H4′	5.849
L22H18–L21H11	4.762	L22H1–L21M3	3.273	L22H7–A6H2	4.579
L22H18–L21H9	4.557	L22H1–L21Q30	3.221	L22H7–G15H3′	3.188
L22H18–L22H16	3.849	L22H1–L21Q31	2.669	L22H7–G15H2′2	4.019
L22H18–L22M2	2.937	L22H1–L21M6	5.438	L22H2–L22H7	2.21
L22H14–T5H5′2	4.566	L22H1–L22H5	2.923	L22H2–L22H1	1.977
L22H14–A6H4′	5.368	L22H2–C4H1′	2.883	L22H3–T5H3′	3.113
L22H14–A6H1′	3.457	L22H2–C4H2′2	2.977	L22H3–A6H2	4.097

L22H14−A6H2′2	5.657	L22H2−C4H2′1	3.038	L22H3−G7H1	3.462
L22H14−G7H4′	5.386	L22H2−C14H1′	2.995	L22H3−C14H1′	3.13
L22H14−C16H2	3.707	L22H2−G15H3	6.417	L22H3−G15H3	5.193
L22H14−L22H16	3.718	L22H2−L21H21	4.603	L22H3−L21H21	4.523
L22H14−L22M2	7.286	L22H2−L21H29	3.991	L22H3−L21Q32	6.715
L22M1−C16H4′	3.999	L22H2−L21Q31	3.227	L22H3−L22H7	2.807
L22M1−L21H18	4.064	L22H2−L21Q32	3.575	L22H3−L22H5	3.58

7.3 Appendix 3

Input files used in the NMR refinement of thiazotropsin B-d(CGACGCGTCG)$_2$ and AIK18-51-d(CGACTAGTCG)$_2$ complexes in vacuum.

Initial minimisation

```
10-mer DNA-ligand complex: initial minimization prior to MD
 &cntrl
  imin   = 1,
  maxcyc = 500,
  ncyc   = 250,
  ntb    = 0,
  igb    = 0,
  cut    = 12
 /
```

Molecular dynamics production run

```
10-mer DNA-ligand complex: MD in-vacuo, 12 angstrom cut off
 &cntrl
  imin = 0, nmropt=1, ntb = 0,
  igb = 0, ntpr = 100, ntwx = 100,
  ntt = 3, gamma_ln = 1.0,
  tempi = 300.0, temp0 = 300.0
  nstlim = 50000, dt = 0.001,
  cut = 12.0
 &end
 &wt type='REST', istep1=0,istep2=1000,value1=1.0,
           value2=1.0,  &end
 &wt type='END' &end
DISANG=filename.rest
LISTOUT=POUT
```

7.4 Appendix 4

Input files used in the NMR refinement of thiazotropsin B-d(CGACGCGTCG)$_2$ and AIK18-51-d(CGACTAGTCG)$_2$ complexes in explicit solvent.

Minimise water and ions

```
Initial minimization fix solute minimized solvent+ions 12.0 cut
 &cntrl
  imin =1,
  maxcyc =1000,
  ncyc =500,
  ntb = 1,
  ntr = 1,
  cut = 10
/
Hold The solute fixed
500.0
RES 1 20
END
END
```

Minimise the whole system

```
initial minimization whole system
 &cntrl
  imin   = 1,
  maxcyc = 2500,
  ncyc   = 1000,
  ntb    = 1,
  ntr    = 0,
  cut    = 10.0
 /
```

Equilibration

```
10-mer DNA-ligand complex: 20ps MD with res on DNA
 &cntrl
  imin   = 0,
  irest  = 0,
  ntx    = 1,
  ntb    = 1,
  cut    = 10,
  ntr    = 1,
  ntc    = 2,
  ntf    = 2,
  tempi  = 0.0,
  temp0  = 300.0,
  ntt    = 3,
  gamma_ln = 1.0,
  nstlim = 10000, dt = 0.002
  ntpr = 100, ntwx = 100, ntwr = 1000
 /
Keep DNA fixed with weak restraints
10.0
RES 1 20
END
END
```

Molecular dynamics production run

```
10-mer DNA-ligand complex: Production run
 &cntrl
  imin = 0, nmropt=1, irest = 1, ntx = 7,
  ntb = 2, pres0 = 1.0, ntp = 1,
  taup = 2.0,
  cut = 10.0, ntr = 0,
  ntc = 2, ntf = 2,
  tempi = 300.0, temp0 = 300.0,
  ntt = 3, gamma_ln = 1.0,
  nstlim = 50000, dt = 0.002,
  ntpr = 100, ntwx = 100, ntwr = 1000
 &end
 &wt type='REST', istep1=0,istep2=1000,value1=1.0,
            value2=1.0,  &end
 &wt type='END' &end
DISANG=ACTAGT.rest
LISTOUT=POUT
```