Structural studies of DNA complexes with minor groove-binding drugs

Deeksha Ganesh Munnur

Thesis submitted to University of London for the degree of Doctor of Philosophy

March 2012

The London school of Pharmacy,
University of London

This thesis describes research conducted in the School of Pharmacy, University of London between 2008 and 2011 under the supervision of Prof. Stephen Neidle. I certify that the research described is original and that any parts of the work that have been conducted by collaboration are clearly indicated. I also certify that I have written all the text herein and have clearly indicated by suitable citation any part of this dissertation that has already appeared in publication.

Signature .... Date ....28|03|12....

To my parents...

My father, who has always encouraged and supported my every dream and

my mother, who introduced me to science through her kitchen.

Abstract

Targeting the minor groove of DNA with small molecules is an important recognition strategy in biology. A wide range of minor groove binding ligands (MGBLs) with good sequence discrimination ability are of interest as potential therapeutic agents in a variety of human diseases such as cancer, along with anti-bacterial and/or anti-parasitic activities. Whilst the detailed mechanism of action of some of these MGBLs is still unproven, they are known to be effective inhibitors of a number of minor and major groove binding protein-DNA interactions.

This thesis reports on crystallographic studies to determine the molecular structure of MGBLs bound to DNA sequences, in order to better understand the details of their molecular recognition by DNA. Several interesting MGBLs differing in their structural features were crystallised with A-T rich oligonucleotides for neutron and high-resolution X-ray data collection. Phases for the X-ray crystal structures were determined using molecular replacement, with diffraction data up to 1.2 Å resolution. The crystal structure revealed the MGBLs bound in the central AATT or AAATTT rich region of the minor groove of the DNA. The ligands form hydrogen bonds with the bases of the DNA at the floor of the minor groove directly or mediated via water molecules depending on the shape of the ligand. Several oligonucleotide-MGBL complexes were crystallised in the presence of deuterium oxide (heavy water) with the aim of studying the water network around the minor groove in the presence of ligand using neutron crystallography.

In order to further our understanding of the biological mechanism of action of MGBLs, biophysical studies were undertaken with the DNA major groove binding transcription factor, NF-κB. This transcription factor binds to the continuous guanine and cytosine bases of the major groove leaving the minor groove exposed to other molecules. Surface plasmon resonance (SPR) and small angle X-ray scattering (SAXS) studies were undertaken to study the effects on MGBLs on NF-κB-DNA binding. It was revealed that MGBLs had significant effect on the protein-DNA interactions which was further dependent on the shape of the MGBLs.

Acknowledgements

I wish to express my sincere gratitude to Prof. Stephen Neidle, at The School of Pharmacy, London (LSOP) for his scientific support and to have given me the opportunity to be part of this valuable project. Without his invaluable support and help, this project would not have been at this stage.

I am deeply grateful to my supervisors Prof. Trevor Forsyth at Institut Laue-Langevin (ILL) and Dr. Edward Mitchell at European Synchrotron Radiation Facility (ESRF). It has been a pleasure to work and learn new things under their guidance. I thank them for their constructive criticism and suggestions that they have given me from time to time, which have helped me, push the boundaries continuously and realise my own potential.

I would also like to thank my supervisor Dr. Susana Teixeira (ILL) for introducing me to an entire new world of protein expression and purification; it has been an absolute pleasure to work with you.

I would like to thank Overseas Research Student (ORS) award, LSOP, ILL and ESRF for funding this work. Special thanks to Andrew Harrison, Director of ILL for providing the additional financial aid to extend my stay at ILL which has helped to add a new dimension to this work.

I am grateful to our collaborators Prof. David Boykin and Prof. David Wilson and their groups at Georgia State University, Atlanta for providing the ligands; Dr. Darren Hart at EMBL, Grenoble for providing the plasmid for the transcription factor. I would like to acknowledge Dr. Matthew Blakeley (ILL) for his help with neutron data collection on LADI-III; Dr. Nicole Thielens (IBS, Grenoble) for her help with initial SPR experiments; Dr. Louiza Zerrad (ESRF) and Dr. Adam Round (EMBL) for their help with SAXS beamtime. I would like to thank everyone at BMSG, LSOP and everyone at

Deuteration Lab, ILL for their help and guidance; especially to JB, Marie Therese, Martine and Michael.

I must mention thanks to Nancy, Silvia and Shozeb at LSOP for their friendship; to my girls Estelle, Salyha and Shona for making my time in Grenoble so full of fun and madness and also for patiently listening to all my laptop tantrums. Thanks to Lakshmi, Paras, Tejas, Kailash, Anton, Stu, Ricardo, Max and many others for all the fun time in Grenoble and also to my favourite people at 61 and 66: Shri, Sau, Vips, Vam, Navos, Mana and Asmi for their friendship and for practically feeding me through my time in London!

This acknowledgement would be incomplete without thanking my family; my *Amma & Appa* and my sisters (Nayana & Sneha) for their endless love and support. To Yash, for his constant love, support, patience and positive attitude for as long as I have known him. Thank you for counting my time until deadline in number of hours instead of number of days or weeks!

Table of Contents

……………………………………………………………………………………………………

Chapter 1: Introduction 1

……………………………………………………………………………………………………

……

List of Figures

List of Tables

List of Abbreviations

A	Adenine
A_2T_2	$\mathrm{d(CGCG\underline{AATT}CGCG)_2}$
A_3T_3	$\mathrm{d(CGC\underline{AAATTT}GCG)_2}$
BDC	Biotin drug control DNA
BME	beta mercaptoethanol
BM marker	Board range marker
bp	base pair
BSA	Bovine serum albumin
C	Cytosine
D_2O	Heavy water
DB75	Furamidine
DLS	Dynamic light scattering
D_{max}	Maximum dimension
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
E.coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetate
ESRF	European synchrotron radiation facility
Fc1	Flow cell 1
Fc2	Flow cell 2
FPLC	Fast protein liquid chromatography
FT	Flow through
G	Guanine
HCl	Hydrogen chloride
HEPES	(4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)
HPLC	High performance liquid chromatography
$I(0)$	Scattering intensity

IBS	Institut de Biologie Structurale
ILL	Institut Laue Langevin
KCl	Potassium chloride
LB	Luria-Bertani medium
MGBLs	Minor groove binding ligands
MM	Molar mass
MPD	2-methyl-2,4-pentanediol
NaCl	Sodium chloride
NaOH	Sodium hydroxide
NDB	Nucleic acid database
NF-κB	Nuclear factor kappa-light-chain-enhancer of activated B cells
OD	Optical density
p50	NF-κB p50 protein
PAGE	Polyacrylamide gel electrophoresis
PDB	Protein data bank
PdI	Poly dispersity index
PM marker	Precision marker
$p(\mathrm{R})$	Pair distance distribution function
R_g	Radius of gyration
RNA	Ribonucleic acid
rpm	rotations per minute
RU	Response unit
SAXS	Small angle X-ray scattering
SE	Standard error
SDS	Sodium dodecyl sulphate
SPR	Surface Plasmon Resonance
T	Thymine
UV	Ultraviolet
V_p	Porod volume

Chapter 1

Introduction

DNA (deoxyribonucleic acid) is one of the most important molecules in living systems. Since the 1940's various studies have been performed to study its structural features. In 1950, Chargaff proposed that DNA have similar amounts of adenine (A) and thymine (T) and the amount of guanine (G) is usually similar to amount of cytosine (C). In other words the total amount of purine (A+G) is usually equal to the amount of pyrimidine (T+C) (Chargaff. 1950), also known as "Chargaff's rule". The Chargaff's rule along with X-ray diffraction image from Rosalind Franklin, contributed to Watson and Crick's derivation of the double helical DNA model in 1953 (Watson & Crick. 1953). With the rapidly increasing understanding of DNA structures and functions, nucleic acid targeting by small molecules in a sequence-selective manner became a forefront drug design strategy.

This thesis describes structural and biophysical studies performed on DNA and a series of small molecules that interact with DNA, and also attempts to add understanding to the biological relevance of these small DNA interacting molecules.

1.1. DNA

1.1.1. Structure of DNA

DNA is a polymer which contains repeating monomer units of nucleotides. Each nucleotide of the DNA consists of a 2'-deoxyribose sugar moiety, nitrogen bases attached to the sugar moiety and a phosphate group which forms part of the DNA backbone (Figure 1.1.a). The nitrogenous bases attach to the C1' position of the deoxyribose sugar and the phosphate group can attach to the 3' or 5' hydroxyl group of the deoxyribose sugar to form a single nucleotide. The DNA bases which determine the sequence of DNA consists of the purines - guanine (G) and adenine (A) - and the pyrimidines - thymine (T) and cytosine (C). The nucleotides are linked to each other via the phosphate groups which bind the 3' or 5' hydroxyl group of the deoxyribose sugar providing directionality to the DNA chain.

One of the greatest milestones in the field of nucleic acid research was when James Watson and Francis Crick proposed the model of a double helix DNA (Watson & Crick. 1953). According to the model they proposed DNA consists of two helical chains each coiled around the same axis with the chains running in opposite directions (Figure 1.1.b). The distance between each base in the chain was proposed to be 3.4 Å and the complete helix turn is achieved after 10 bases on each chain giving a pitch of 34 Å. The plane of the bases is perpendicular to the helix axis. The bases are present on the inside of the helix and the phosphate groups are present on the outside forming the backbone of the DNA. The two chains are held together by hydrogen bonds between the purine and pyrimidine bases of each chain. Based on Chargaff's rules it was also suggested that the A would base-pair with T, and G would pair with C (Chargaff. 1950). The A-T base-pair forms two hydrogen bonds whilst the G-C base pair forms three hydrogen bonds. The base pairing arrangement is such that the sugar moieties linked to the individual pairs are on the same side. So when the base pairs are stacked on each other in the DNA helix, the gap between the sugars forms continuous indentations in the surface that wind along, parallel to the sugar phospodiester chain (Neidle. 2007). These indentations are termed as the grooves of DNA (Figure 1.1.b). The sugar moieties of the base pair are on the minor groove of the DNA (Figure 1.5).

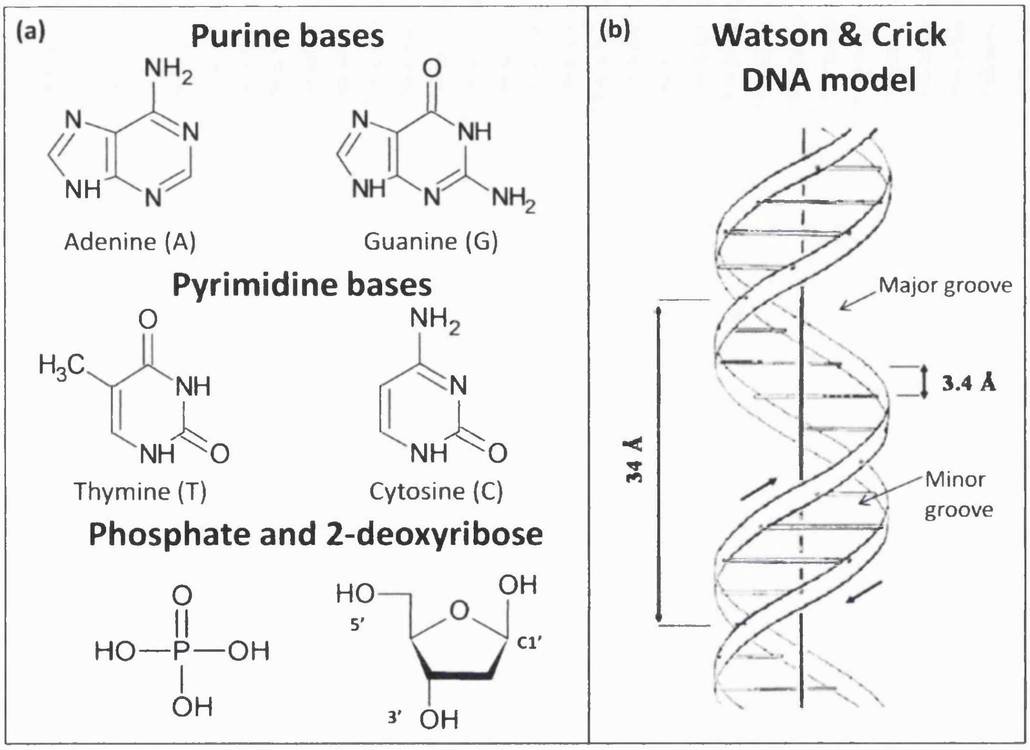


Figure 1.1. (a) Schematic representation of the four DNA bases, phosphate group and 2- deoxyribose sugar moiety. (b) DNA double helix model proposed by Watson and Crick (Watson & Crick. 1953).

1.1.2. Different forms of DNA

DNA is a flexible macromolecule which is one of the major reasons for the wide variety of DNA species available in nature. DNA can exist as double helices or it can exist as multi-strand structures. The different known double helical forms of DNA are the A, B, C, D and the Z form of DNA; multi-stranded DNA can exist as the triple (Felsenfeld *et al.* 1957), quadruplex (Morgan. 1970), junction (Beerman & Lebowitz. 1973) and parallel helix structure (van de Sande *et al.* 1988). This section describes the different forms of duplex DNA with emphasis on the B-form of DNA.

Five different forms of double helix DNA (A, B, C, D and Z) have been studied for their structural features. The A, B, C and D forms of DNA are right handed helices, whilst the Z-form is the only known left handed helix structure (Figure 1.2). The model proposed by Watson and Crick is for the B-form of DNA. Fibre X-ray diffraction techniques first analysed the duplex DNA structures of the A (Fuller *et al.* 1965), B (Langridge *et al.* 1960), C (Marvin *et al.* 1958), and D (Davies & Baldwin. 1963) forms of DNA (Figure 1.2). The evidence for the presence of the Z form of DNA was first observed from a circular dichroism study, which showed a major change in the standard B-DNA spectra at high ionic (salt) concentration (Pohl & Jovin 1972). The structural evidence for Z-DNA was obtained from the X-ray crystal structure (Wang *et al.* 1979) and fibre diffraction studies (Arnott *et al.* 1980). The B-form is the major form of DNA found in cells, although under a number of biologically important circumstances, such as change in hydration state and/or ionic strength, the DNA conformation can be readily deformed either slightly or completely to one of the other DNA forms (Fuller *et al.* 2004).

These five forms of DNA double helices show significant differences in their pitch, rise per base pair, and number of base-pair per helix per turn (Figure 1.2), which further causes a difference in their major and minor groove parameters. For example, the major groove of the A-form of DNA is narrow and very deep, the B-form has a wide and deep major groove, whilst the Z-form of DNA has a flat major groove. The minor grooves of these DNA forms are very broad and shallow for A-form, narrow and comparatively deep for B-form, and very narrow and deep for the Z-form of DNA (Berg *et al.* 2002). These differences in the nature of the DNA grooves provide important and useful recognition ability to the DNA binding protein and also small molecules.

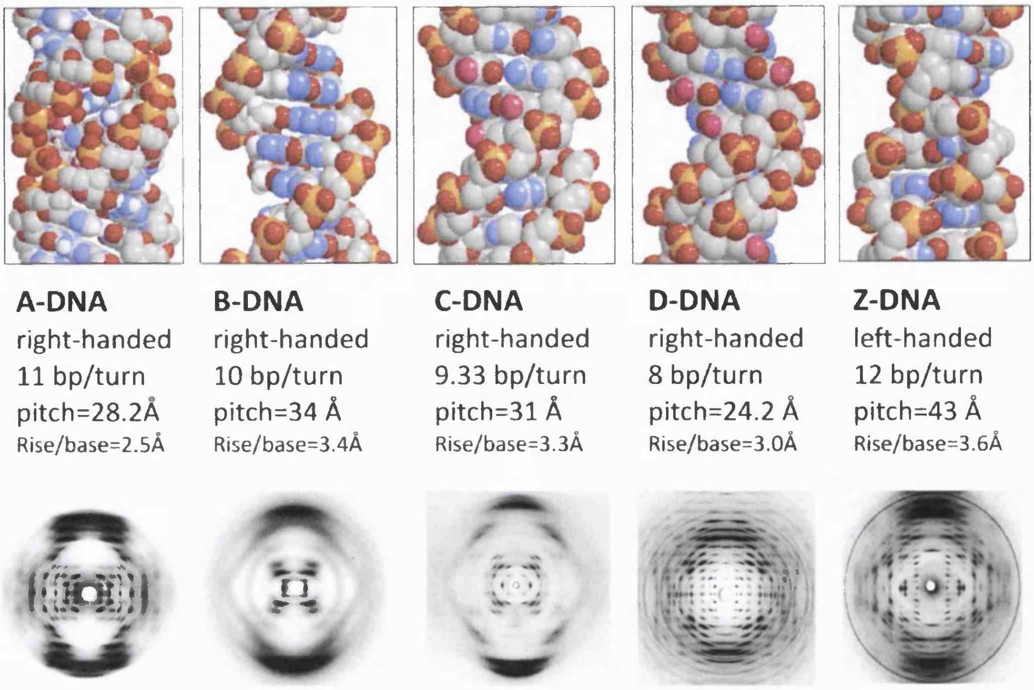


Figure 1.2. The five major conformational variants of the DNA double helix, their helix parameters and X-ray diffraction pattern (Fuller *et al.* 2004).

Most of the structural information on the B-form of DNA was obtained from the single crystal X-ray structure of the Dickerson-Drew dodecamer oligonucleotide sequence CGCGAATTCGCG (Wing *et al.* 1980). This structure demonstrated the presence of the anti-parallel right handed double helix B-DNA and thus proving Watson and Crick's DNA model to be correct. The crystallographic structure revealed many sequence dependent morphological features of duplex DNA which could not be observed in fibre diffraction studies. From the crystal structure it was observed that the 5'-AATT minor groove region was narrow with a width of 3.2 Å whilst the major groove region is 12.7 Å wide. The crystal structure also revealed the presence of a structured water network in the minor groove. There are primary and secondary shells of hydration. The primary shells of hydration forms direct hydrogen bonds with the O2 and N3 atoms of thymine and adenine bases respectively, while these water molecules are connected to each other via the secondary shell of hydration. The crystal structure showed the DNA helix pitch to be 35.9 Å and that there are 10.1 bases per helix turn compared to the lower resolution fibre diffraction study which revealed the pitch and number of bases per turn being 34 Å and 10, respectively. The crystal structure also shows that the helix axis is not straight, but bent by 19° towards the major groove of the DNA.

Many crystallographic studies have since then been carried on different duplex DNA sequences, which have led to the conclusion that many morphological features of duplex DNA are sequence dependent, such as groove width and distribution of electrostatic potential (Neidle. 2001). For example, the synthetic polymer poly [d(A-T)].[d(A-T)] has a narrower minor groove and greater negative potential in comparison to poly[d(G-C)].[d(G-C)] (Figure 1.3).

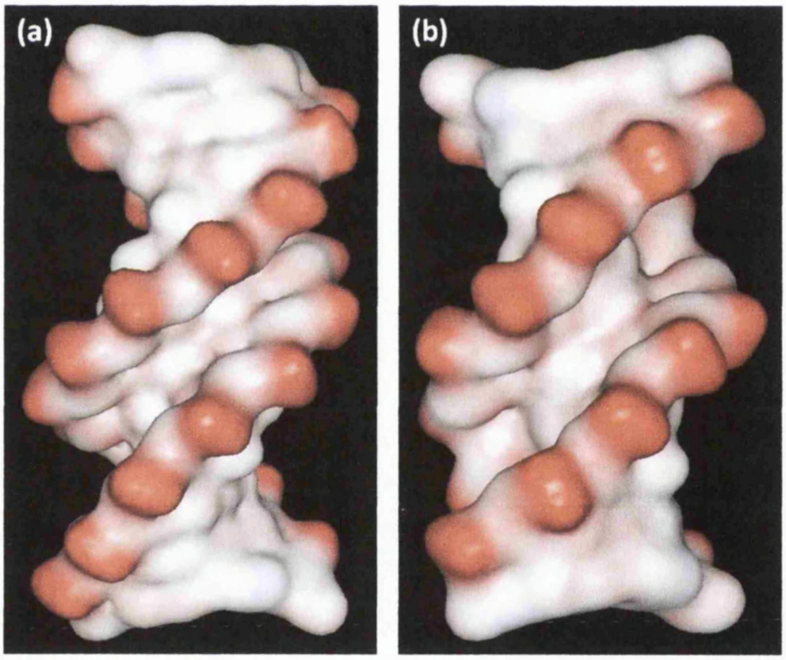


Figure 1.3. Views of electrostatic potential surface and groove width difference of (a) A-T minor groove in the crystal structure of d(CGCAAATTTGCG)$_2$ (PDB code 1D65) and (b) G-C minor groove in the crystal structure of d(CCAGGCCTGG)$_2$ (PDB code 1BD1) (Neidle. 2001).

1.1.3. Function of DNA

The DNA molecule has great biological significance as it is the information bank which governs all life processes of an organism. The main functions of DNA is coding for proteins, replication and genetic information. DNA holds the information for the development and reproduction of an organism and hence the discovery and understanding of DNA has been vital in treating diseases.

Transcription and translation

DNA holds the genetic code for protein synthesis. A "unit" of three DNA nucleotides (DNA code) is first transcribed from the DNA into a RNA code by the mRNA. The adenine on the DNA pairs with uracil on the RNA, thymine with adenine, guanine with cytosine and cytosine with guanine on the RNA. This RNA code is then in turn translated to protein code (an amino acid) by the tRNA. The amino acids are organised in a sequence to build a protein. The ribosomes present in every cell work with the mRNA and tRNA to combine amino acids together in the correct sequence to form a protein.

DNA replication

Replication of DNA is essential for the reproduction, maintenance and growth of cells, tissues and body systems. The replication process begins with one duplex DNA which essentially "unzips" with the aid of helicase enzymes which break the hydrogen bonds holding the two DNA strands together, giving two branching "prongs", each of which are single strands of DNA. These single strands of DNA act as template for production of the complementary strands by the DNA polymerase enzyme. By the end of the replication process, the new DNA strands created are a perfect copy of the original DNA strand.

Genetic code

DNA sequences that do not code for protein synthesis are termed as non-coding DNA. This non-coding DNA encodes functional non-coding RNAs which are involved in regulation of gene expression. Some non-coding DNA sequences play roles in chromosomes. These sequences serve as genetic material for the creation of new genes through gene duplication and divergence.

1.2. DNA interacting agents

As mentioned earlier, as understanding of DNA structures and functions has increased, potential for nucleic acid targeting by small molecules in a sequence-selective manner has become a strong development line for new therapeutic molecules. There is increasing evidence that the interference of naturally-occurring small drug-like molecules with particular protein-DNA interactions is critical for biological activity and selectivity (Neidle & M. Nunn. 1998). These small molecules are potential anti-cancer, anti-bacterial, anti-viral and anti-parasitic agents (Neidle *et al.* 1994; Neidle & Thurston. 1994). A DNA interactive agent, carmustine, formulated as a slow release "wafer" dosage form (Gliadel Wafer TM) is being studied in phase II clinical trials for local treatment of brain tumours (Thurston. 2006). Berenil, a minor groove binding ligand (MGBL) is being currently used in West Africa for curative treatment of trypanosomiasis and babesiosis in animals. Topical use of berenil to treat human cutaneous leishmaniasis has also been reported (Seitz. 2003). Furamidine (also known as DB75) is a drug most active against *Pneumocystis carinii*, a widespread pathogen responsible for pneumonia that often afflicts the majority of AIDS patients. In addition, this drug is also active against highly infectious parasites such as *Giardia lamblia*, *Plasmodium falciparum* and *Trypansomarhodesience* (Lansiaux *et al.* 2002).

1.2.1. Different types of DNA binding ligands

Drugs binding to nucleic acid can be classified into two main groups which are described below.

1.2.1.1. Covalently bound ligands

Covalent binding of a drug to DNA is an irreversible process leading to complete inhibition of DNA processes and to cell death, usually through invoking apoptosis (Thurston. 2006). This class of DNA binders include alkylators (e.g. dacarbazine) and cross-linking alkylators (e.g. cisplatin). They interfere with the replication and transcription processes of DNA.

1.2.1.2. Non-covalently bound ligands

This class has two basic types of molecules:

Intercalating agents

Drugs belonging to this class are normally planar heterocyclic molecules which insert between the base pairs of DNA, perpendicular to the axis of the helix. These ligands are held in position by non-covalent interactions, such as hydrogen bonding and van der Waals contacts. These agents (e.g. doxorubicin, actinomycin D) can block DNA transcription and interfere with some DNA processing enzymes, although they show only modest selectivity.

Groove binders

DNA groove binding ligands compete with transcription factors and so interfere with gene regulation. MGBLs have better sequence discrimination ability in comparison to intercalators and hence are of interest as a potential therapeutic target in many human diseases. Most of the MGBLs interact with between three and five A-T base pairs.

However, guanine interactions are not favoured due to steric clashes with the 2-amino group of guanine as shown in Figure 1.5. The sections below focus on this class of DNA binding molecules.

1.2.2. Minor groove binders

This class of ligands binds to duplex DNA, favouring the B-form, without disturbing the DNA conformation during the binding process. Several DNA foot-printing studies have been performed which reveals that AATT and AAAA are much better binding sites than TTAA and TATA (Abu-Daya *et al.* 1995). The Hoechst 33258 molecule shows better discrimination ability between these sites with a 50 fold difference in affinity between AATT and TATA (Abu-Daya *et al.* 1995). This class of drugs can be sub-classified into three groups with parent compounds as follows (Figure 1.4).

a) Netropsin and distamycin based motifs, which are naturally occurring.
b) Benzimidazole-based ligands, which are synthetic molecules, e.g. Hoechst-33258.
c) Aromatic diamidine ligands, which are also synthetic molecules, e.g. berenil.

The ligands belonging to the first two groups have a number of charged and neutral hydrogen-bond donor groups dispersed throughout the molecules. These ligands have the potential to form many hydrogen bonds with DNA base-pair edges. The aromatic diamidine or the bis-phenylamidinium ligands have a central aromatic ring structures which is flanked with charged amidinium groups at the end, unlike the parent compound berenil (Neidle. 1997).

In 1984, the first minor groove drug complex structure of netropsin with d(CGCGAATTCGCG)$_2$ was determined to 2.2 Å resolution; PDB Code 6BNA (Kopka *et al.* 1985). Since then, X-ray crystallography has been used as the major tool to obtain structural information on the complexes of these ligands with DNA. The Nucleic Acid Database now cites over 83 X-ray crystallographic structures of MGBLs with the d(CGCGAATTCGCG)$_2$ and d(CGCAAATTTGCG)$_2$ sequences. Several MGBLs have been extensively studied to obtain their molecular structures (by X-ray crystallography), to identify their DNA binding sites (by DNase I foot-printing) and mode of DNA binding (by circular dichromism or scanning-force microscopy), and to obtain their relative binding affinities (by thermal melting study and electrospray ionisation mass spectroscopy) (Mazur *et al.* 2000; Neidle. 2001).

Crystallographic studies on a number of diamidine derivatives have revealed that they bind to the central A-T rich region in the minor groove of the DNA. The amidinium groups can participate in hydrogen bonding with acceptor base atoms on the floor of the minor groove. Most of the MGBLs that have been studied are planar molecules with a narrow cross section, which is a prerequisite for the ligand to bind in the narrow minor groove in an A-T region via van der Waals interactions. The classic view of MGBLs is that they have an optimum curvature matching the curvature of the DNA, greatly facilitating the recognition process, as seen with berenil and DB75 (Figure 1.4). Earlier studies have shown that increasing or decreasing the curvature of the ligands can affect their binding affinity towards the groove. This has been noted for DB359 (Figure 1.4) which has meta-substituted amidinium groups with increased curvature binds weakly to DNA minor groove in comparison to DB75. DB75 has a high binding affinity for A-T sequence ($K_a > 2 \times 10^7 M^{-1}$) whilst DB359 has a weaker interaction with both A-T and G-C DNAs ($K_a < 1 \times 10^7 M^{-1}$). There is evidence which suggests that DB359 has a predominant intercalation mode in A-T and G-C DNA (Nguyen *et al.* 2002). However, the hydrogen bonding characteristics of a particular ligand may relax the iso-helicity requirements for effective minor groove binding. MGBLs such as pentamidine, CGP 40215A, and DB921 (Figure 1.4) have a near-linear structure but have shown better affinity and more biological significance (Miao *et al.* 2005; Nguyen *et al.* 2004). Pentamidine is considered as one of the most successful candidates against eukaryotic parasites and it has been used clinically against trypanosomiasis, leishmaniasis and *P. Carnii* pneumonia for over seven decades (Bakshi & Shapiro. 2003; Bouteille *et al.* 2003; Seitz. 2003). Due to this success, two new classes of aromatic diamidine derivates have emerged: the classical iso-helical ligands, which match the curvature of the minor groove of the DNA, and the new class of linear ligands.

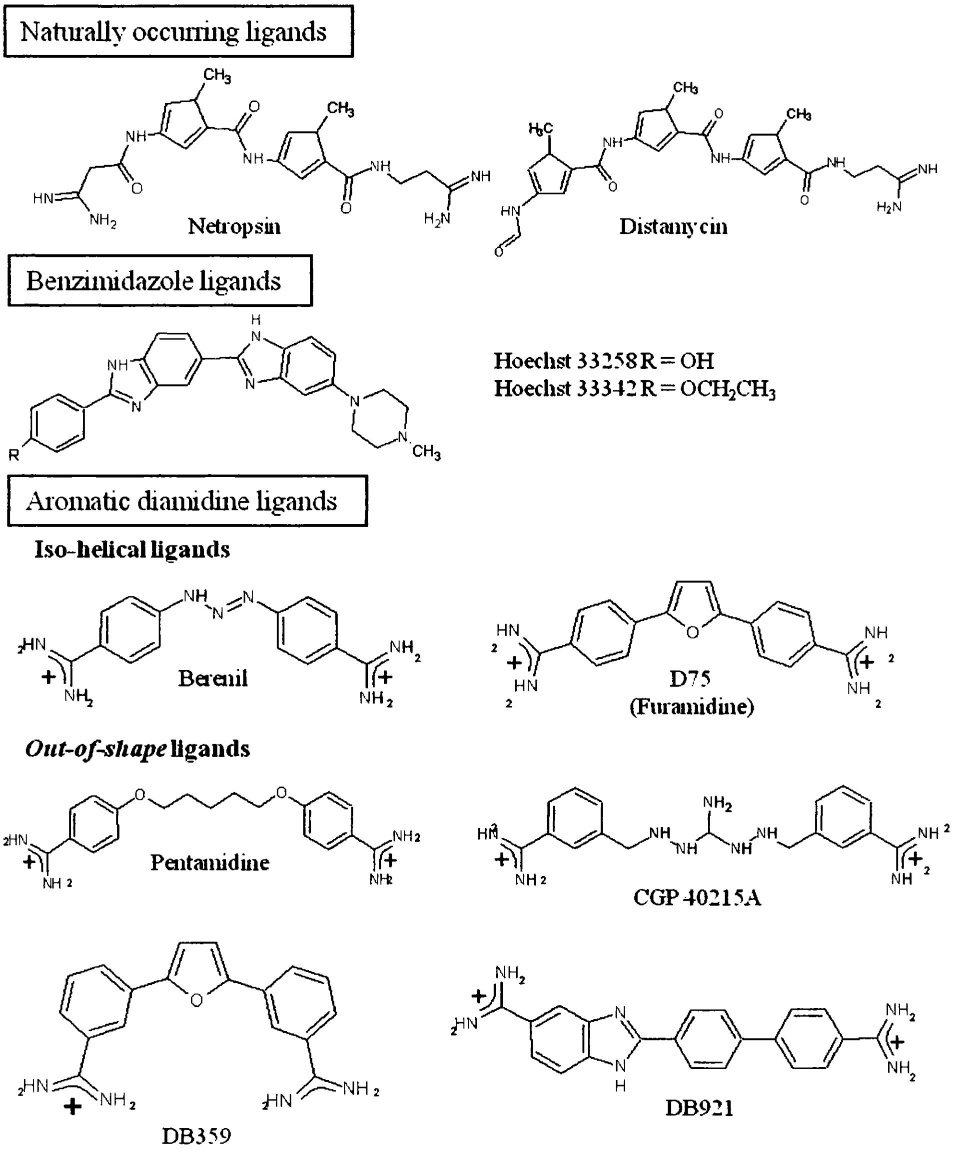


Figure 1.4. Chemical schematics of the benzimidazole and the aromatic diamidine analogues that bind to the minor groove of the DNA.

1.2.3. Factors that affect the minor groove recognition

As mentioned earlier, DNA groove features such as groove width, distribution of the electrostatic potential, and also the shape of the groove are sequence dependent. The A-T rich minor groove region is narrower (groove width 3-4Å) compared to the minor groove of the G-C rich region (groove width 5-6Å) (Figure 1.3). The sequence selectivity of the MGBLs depends upon structural and electronic factors, along with electrostatic, van der Waals and hydrophobic factors (Neidle. 2001). The electrostatic distribution along DNA sequences is a key factor involved in the minor-groove recognition of the ligands. Almost all known MGBLs have positive charges complementary to the strong negative potentials at the floor of the minor groove in A-T rich regions.

The floor of the minor groove formed by the base edges also influences the recognition process. Both NMR and X-ray crystallographic studies have shown that minor groove floor of A-T tract is a single smooth curve, whilst that of a G-C tract would be discontinuous due to the presence of the exocyclic 2-amino groups of guanine and cytosine (Figure 1.5). The presence of these irregularities acts as a barrier to prevent the ligands from attaining close contact with floor of a G-C sequence. There is also a difference in the numbers of hydrogen-bond donors and acceptors in the bases (Figure 1.5) which is an additional differentiation between A-T and G-C base pairs in the groove.

Water molecules present in the minor groove also play an important role in recognition. When binding to the minor groove the ligands are believed to displace the water network, thereby causing a positive entropy change which could constitute a significant driving force for the interactions. Water molecules provide alternative groove recognition ability, especially in the case of linear ligands, by mediating between the linear ligands and minor groove with loss of binding affinity (Miao *et al.* 2005; Nguyen *et al.* 2008). However, the role of water molecules in recognition has not yet been studied extensively.

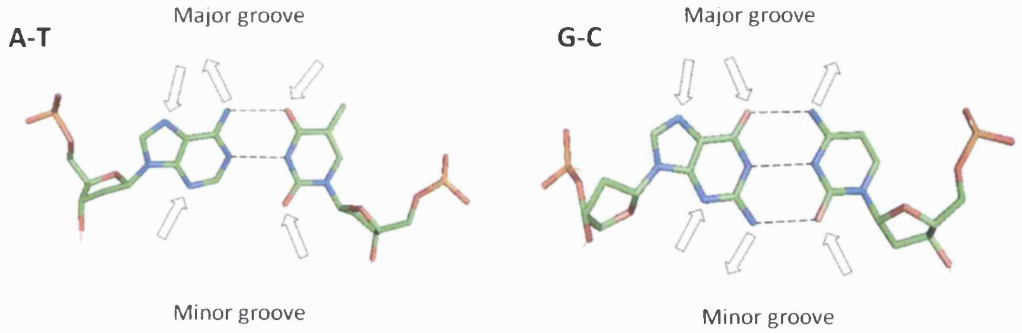


Figure 1.5. The potential hydrogen bond acceptors and donors of A-T and G-C base pairs are shown by the arrows. The figure also shows the major and minor groove of the DNA.

1.2.4. Physiological activity of minor groove binders

The minor groove of DNA is an important target site for small molecules with therapeutic and sequence-selective activities. Ligand interactions in the minor groove are distinct from those in the major groove, often having increased surface contact. The exact mechanism of action of MGBLs is not known. It is believed that some may act through inhibition of enzymes associated with DNA replication, such as DNA topoisomerase I and/or II, or by direct inhibition of transcriptional control proteins (Simpson *et al.* 2000).

Topoisomerase are essential enzymes for viability even in non-replicating cells. Topoisomerase work via a three-step catalytic mechanism involving cleavage, strand-passage and relegation of DNA strands (Jean-Moreno *et al.* 2006; Wang. 1985). They can be classified as topoisomerase I and topoisomerase II based on their specific mode of action on one or both strands of DNA double helix (Wang. 1985). In a study performed by Chen *et. al.* on mammalian DNA topoisomerase it was observed that MGBLs such as bis-benzimidazole dyes, Hoechst 33342 and Hoechst 33258 (Figure 1.4) induce limited but highly specific single strand DNA breaks in the presence of topoisomerase I at extremely low drug concentration. These ligands bind to the A-T rich regions of the DNA, interfering with the enzyme catalysis and suggesting possible involvement of a topoisomerase I-drug-DNA ternary complex (Chen *et al.* 1993). Some MGBLs such as distamycin A, Hoechst 33258 and DAPI, modulate the enzymatic activities of both topoisomerase I and II purified from L1210 cells (mammalian cell) (McHugh *et al.* 1989; Woynarowski *et al.* 1989). DNA binding ligands that significantly alter the binding affinity, the binding site(s), and/or the subsequent enzymatic activity of topoisomerase may be agents of cell death, either directly or by initiating a cascade of events leading to cell death (Dykstra *et al.* 1994).

A recent report suggests that a part of antileishmanial activity of some MGBLs could be related to type I DNA topoisomerase inhibition (Jean-Moreno *et al.* 2006). The topoisomerase of the parasite exhibits structural and biochemical variation from mammalian topoisomerase and has an important function in organising the kinetoplast DNA (kDNA) network unique to these parasites(Jean-Moreno *et al.* 2006).

Kinetoplast contains mitochondrial DNA made up of thousands of circular DNAs of two types: maxicircles and minicircles (Klingbeil *et al.* 2001; Shapiro & Englund. 1995; Wilson *et al.* 2008). Fluorescence microscopy studies on the aromatic diamidine DB75 indicate specific and rapid binding of DB75 to mitochondrial k-DNA of parasites (Mathis *et al.* 2006; Wilson *et al.* 2008). The kinetoplast begins to change shape, breaks down and subsequently disappears leading to eventual cell death in 24-48 h. This indicates that DB75 exerts specific effects on kDNA replication and/or transcription which eventually lead to parasite death (Mathis *et al.* 2006; Wilson *et al.* 2008). The kDNA minicircles have extensive, closely spaced A-T sequences that provide a potential cellular target for diamidines, such as DB75, that have A-T sequence binding specificity (Wilson *et al.* 2008). The exact step involved after binding to DNA by MGBLs, which leads to cell death, is not clear. There are hypotheses that suggest diamidines induce topological change that results in increased global bending of the A-tract. These conformational changes could interfere with the ability of topoisomerase II to effectively close the minicircle DNA when it is opened for transcription or replication (Wilson *et al.* 2008).

In an enzyme inhibition assay study it was observed that the bis-benzimidazole dyes, Hoechst 33342 and Hoechst 33258, were equally potent inhibitors of both parasite (*Leishmaniadonovani* and *Pneumocystis carnii*) and mammalian topoisomerase I (Jean-Moreno *et al.* 2006). The MGBLs belonging to the aromatic diamidine derivative family, such as pentamidine and berenil, have selectivity for the parasite enzyme and low level of toxicity to mammalian topoisomerase (Jean-Moreno *et al.* 2006). These compounds hold promise as effective therapeutic agents due to their improved selectivity and also making parasite topoisomerase a potential target for further drug development.

MGBLs have also been shown to inhibit the binding of transcription factors to DNA. The transcription factor HMGA2 plays a critical role in disease processes from cancer to obesity. It was observed from a SPR study that the binding of this transcription factor to biologically important sites on DNA was efficiently inhibited in the presence of the MGBL netropsin (Miao *et al.* 2008). NF-κB, another transcription factor, also showed changes in protein-DNA interaction in the presence of the MGBL distamycin (Speight *et al.* 2002). Studies on NF-κB have been used for the present research to further our understanding of the biological mechanism of action of the MGBLs. Therefore, this protein is discussed in detail in the section to follow.

Diamidine derivatives have widespread uses, but they lack oral bioavailability, and some can cause severe toxicity (Delespaux & de Koning. 2007; Seitz. 2003). Regardless of these drawbacks, diamidines are indispensible for their therapeutic applications. But it is essential to develop and design improved ligands to combat drug resistance and to overcome toxicity, and to develop orally active drugs. Several prodrugs of DB75 have shown promising therapeutic responses (Wilson *et al.* 2005). A diamidoxime prodrug, upon oral administration, rapidly metabolises to active furamidine, considerably reducing toxicity. This compound is currently in Phase II clinical trials against malaria and *Pneumocystis carinii* pneumonia (PCP) and has recently been in Phase III clinical trials against human African trypanosomiasis (Wilson *et al.* 2005).

1.3. NF-κB

The NF-κB transcription factor was first discovered as the nuclear factor in B cells that bound to the κB DNA target site (Sen & Baltimore. 1986). The NF-κB family of transcription factors consists of five members: NF-κB1 (p50 and its precursor p105), NF-κB2 (p52 and its precursor p100), RelA (p65), RelB and c-Rel. All NF-κB subunits contain a conserved Rel homology domain which is responsible for nuclear localisation and DNA binding. Each subunit has distinct biological functions; therefore different dimeric combinations exert diverse effects on cell functions. The inactive precursors, such as p100 and p105, and inhibitory proteins (IκB α, β or γ) bind to the NF-κB dimers and sequester them in inactive form in the cytoplasm (Perkins. 2006). The NF-κB signalling pathway activates the inhibitory proteins which releases NF-κB and allows its nuclear translocation (Perkins. 2006).

The full length transcript encoding p105 spans 3452 base pairs with the p50 subunit spanning residues 1-430 at the N-terminal end and the IκB γ spanning residues 365-971 at the C-terminal end of p105. p105 is composed of the N-terminal domain (NTD), dimerisation domain (DimD), glycine-rich region (GRR), ankyrin repeat domain (ARD) and the death domain (DD) (Perkins. 2006). The NTD, DimD and nuclear localisation sequence (NLS) form the Rel homology domain of p50 which controls dimerisation, nuclear localisation and DNA binding.

The NF-κB transcription factor plays a vital role in processes such as growth, development, inflammatory responses and proliferation of the HIV and Herpes virus and pathogenesis of many diseases including cancer, arthritis, live and gastrointestinal disease (Baeuerle & Henkel. 1994; Baldwin. 1996; Pereira & Oakley. 2008; Siebenlist *et al.* 1994). Different subunits of NF-κB play distinct roles in the inflammatory response. Studies have suggested that the p50 homodimer has an anti-inflammatory function (Cao *et al.* 2006). p50 also plays an important protecting role in neural degeneration and brain injury (Pereira & Oakley. 2008; Yu *et al.* 2000). Targeting the pro-inflammatory NF-κB signalling whilst enhancing the protective effects of the p50 homodimer may allow the development of novel therapeutics for the treatment of inflammatory diseases (Pereira & Oakley. 2008).

Two X-ray crystal structure of the p50 homodimer bound to DNA were obtained from two groups at 2.3 Å and 2.6 Å resolution (Ghosh *et al.* 1995; Müller *et al.* 1995). The first (Müller *et al.* 1995) complex structure contains a recombinant human p50 fragment (residues 2-366) and a 19 base oligonucleotides (5'-AGATGGGGAATCCCCTAGA-3') which forms a central 11 base pair duplex. The second (Ghosh *et al.* 1995) complex structure contains recombinant murine p50 (residues 39-364) and an 11 base oligonucleotide (5'-TGGGAATTCCC-3') which form a 10 base pair duplex. The Rel homology region which is clearly defined in the crystal structures of both human and murine p50, is almost identical. However, due to the different sequences of the DNA binding sites there is a difference in the binding modes of the two structures. The key difference between the two structures is the half-site spacing.

The p50 Rel homology region folds into two domains, the dimerisation domain and the specificity domain (Figure 1.6), which are connected by a short linker. Both domains contain a β-barrel. The p50 homodimer wraps into the major groove, so that it nearly encloses the DNA. The loop that connects the β-barrel of both domains contacts the DNA bases and backbone. This linker must be flexible to allow the dimer to open and close its N-terminal 'jaws' around the DNA. According to the crystal structure (Müller *et al.* 1995) base-specific contacts of the protein residues are restricted to the 'specificity domain'. The DNA sequence used for this structure has four continuous guanine bases of which the central two guanine bases form contacts with arginines 57 and 59, the outer guanine is in close contact with histidine 67 and the innermost guanine with lysine 244, which is also the last residue of the specificity domain. Two residues from the C-terminal domain of the second (Ghosh *et al.* 1995) structure extend sufficiently far into the major groove to contact bases; these same residues make DNA backbone contact in the (Müller *et al.* 1995) structure.

Crystal structures of other NF-κB subunits have also shown a similar binding pattern, where the protein wraps around the major groove of the DNA and the protein residues contact with the G-C base-pairs (Chen *et al.* 1998; Cramer *et al.* 1997). A crystal structure of the native DNA sequence that binds to NF-κB has been determined (Huang *et al.,* 2005). This shows significant differences compared to the identical DNA sequence when bound to the NF-κB protein (Chen *et al.* 1998). The G-C region of the DNA which contacts with the protein is buckled and opened in the complex structure but not in the native DNA structure. The central A-T region of the DNA has a large roll angle in the complex structure but not in the native structure. The binding of the NF-κB to DNA causes a significant distortion of the canonical B-form of duplex DNA geometry.

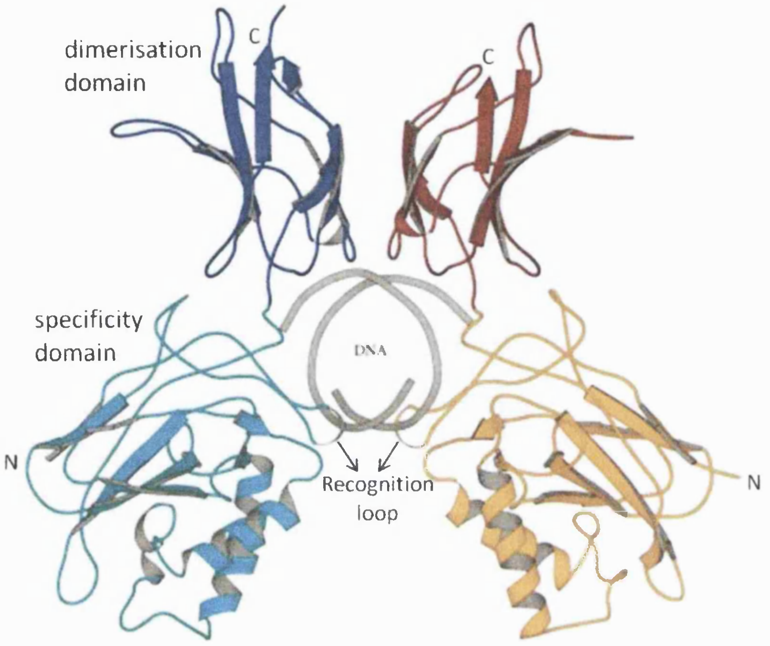


Figure 1.6. View along the DNA axis showing the NF-κB p50 homodimer bound to DNA (Müller *et al.* 1995).

1.4. Aims of this project

This research work describes a series of high resolution X-ray crystallographic structures of DNA in complex with several MGBLs. The aim of this work is to better understand the details of the molecular recognition of groove binding ligands by DNA, as water molecules play an important role in maintaining the stability and flexibility of DNA and also for DNA recognition by ligands. The DNA-ligand complex crystals were also considered for neutron crystallography studies. Finally, to further our understanding of the biological mechanism of action of MGBLs, biophysical studies were undertaken with the DNA major groove binding transcription factor, NF-κB p50 in the presence of MGBLs.

Chapter 2

Experimental Methods

This chapter describes the different techniques, used in this thesis. It describes the principles of crystallography and basics of surface plasmon resonance (SPR) and small angle X-ray scattering (SAXS) used for the studies. X-ray crystallography is one of the most popular techniques to determine structure of macromolecules, whilst neutron crystallography is a key technique to determine information regarding hydration and protonation states. SAXS is an important to tool to determine the low resolution structure of macromolecules in solution. SPR is a biophysical technique which helps to measure the dynamics of biomolecular interactions.

2.1. General Principles of Crystallography

Crystallography is one of the most precise methods to establish detailed features of the molecular structure of every kind of stable chemical species, from the simplest to those containing thousands of atoms. By revealing high resolution structures, it can also be used to understand biological processes at their most basic level. For example, the three-dimensional structure determination of macromolecules is now an integral part of the development process of new drugs.

2.1.1. Crystallisation

Structure determination by X-ray crystallography requires the growth of a single crystal. The basic principle of crystallisation involves three main steps which are nucleation, growth and cessation of growth. The nuclei formation and crystal growth is shown by a two dimensional phase diagram (Figure 2.1.a). The solubility curve divides the concentration areas into an undersaturated and a supersaturated zone. To initiate crystal growth, the solution of macromolecules is slowly brought to supersaturation. The supersaturation zone is further subdivided into the precipitation, nucleation and metastable zones. The precipitation zone has an excess of macromolecules in solution which leads to formation of amphorphous aggregates. In the nucleation zone the excess of macromolecules aggregate in crystalline forms (crystal nuclei). Crystal formation is dependent upon the nuclei reaching a critical size, which is defined by the ratio between the surface area of nucleus and its volume. The optimal condition for obtaining well ordered crystals would be the formation of, a preferably single, nucleus in the nucleation zone just beyond the metastable zone. As the crystal nuclei grow, the solution returns to the metastable region and further nuclei cannot be formed. The growth of crystals should be slow so that a regular arrangement of molecules and ions leads to a well-formed crystal (Glusker & Trueblood. 1985).

There are various methods for growing crystals, such as vapour diffusion (hanging and sitting drop), dialysis (microdialysis, microdialysis and double dialysis), batch crystallisation, seeding (micro and macroseeding) and free interface diffusion method. The vapour diffusion method is one of the most common crystallisation techniques and has been used extensively for the work in this thesis (Figure 2.1.b).

In the vapour diffusion technique, the drop solution has less concentration of precipitants than the reservoir solution. Equilibration proceeds by diffusion of the volatile species (water or organic solvents such as MPD and PEG) until the vapour pressure in the droplet equals the one in the reservoir (Ducruix & Giege. 1992). This leads to a decrease in the volume of the droplet, therefore the concentration of all constituents in the drop will increase, leading to supersaturation in the drop, which may lead to formation of crystals or aggregates in the drop. The main parameters which determine the rate of equilibration are temperature, pH, initial drop volume (and initial surface to volume ratio of the drop and its dilution with the respective reservoir), water pressure of the reservoir and the chemical nature of the crystallising agent (Ducruix & Giege. 1992).

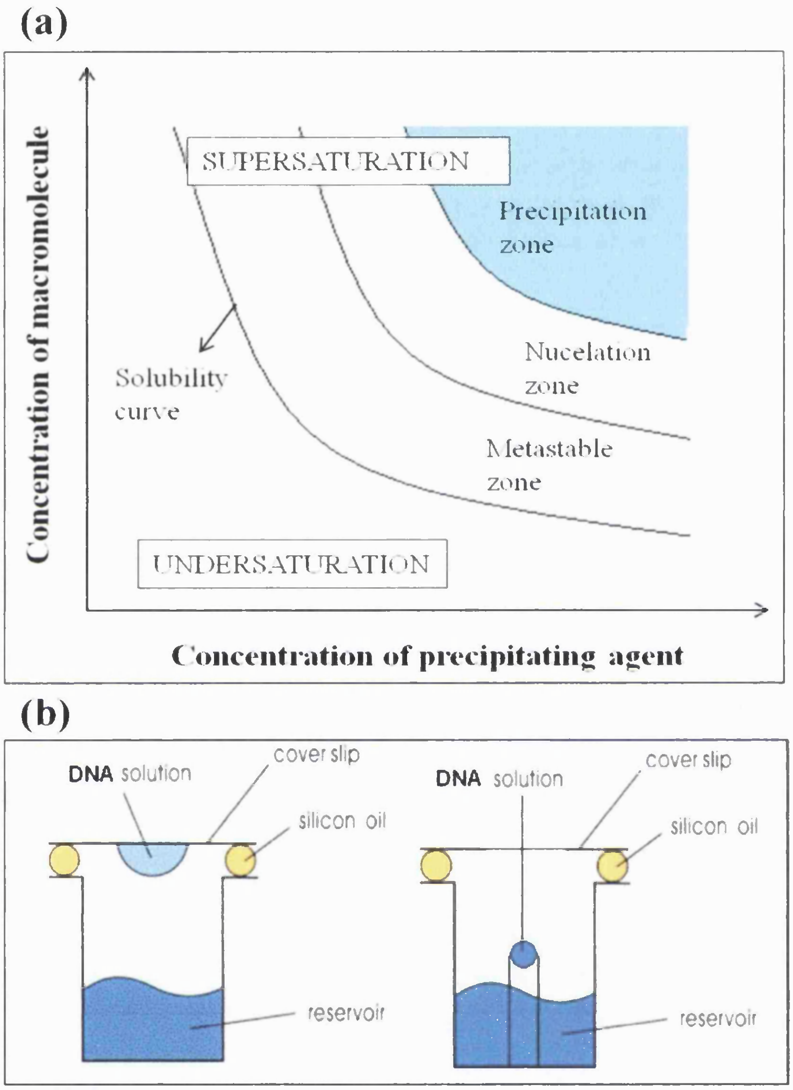


Figure 2.1. (a) Phase diagram for macromolecule concentration and precipitating agent concentration. (b) Schematic representation of the hanging and sitting drop vapour diffusion method of crystallisation.

Crystals obtained from crystallisation trials need to be removed from the drop and prepared for data collection. For data collection at room temperature crystal has to be mounted in a capillary and dried with paper wick. The capillary is sealed with wax and is ready for room temperature data collection. For data collection at low (typically

100K) temperatures, the crystal is mounted in a loop and flash-cooled in the presence of cryoprotecting agents, which prevent formation of ice and consequent deterioration of the crystal quality. Commonly used cryo-protectants include heavy oils, alcohol, polymers, sugars and sometimes high concentration of salts. However, the compatibility of crystal to cryo-protectants needs to be tested beforehand.

2.1.2. Crystals and lattices

Crystals are three-dimensional ordered structures which contain repeats of identical unit cells which are stacked together systematically in all three spatial directions. A single unit cell is a complete representative of the whole crystal and therefore if the exact content of the unit cell is known, the crystal structure can be described. The unit cell can be described by three axes (a, b, c) and three angles (α, β, γ), which leads to seven combinations of a, b, c, α, β and γ known as the seven crystal systems (Table 2.1). The crystal lattice is a regular three-dimensional array of points upon which the contents of unit cell may be considered to be arranged by infinite repetition to build up the crystal structure (Glusker & Trueblood. 1985).

There are 14 unique crystal lattices known as Bravais lattices. The concept of these crystal lattices combined with symmetry elements (screw axis/glide plane) gives rise to 230 distinctive arrangements known as space groups. They are listed and described in volume I of the *International Tables for X-ray Crystallography* (Hanh. 1987).

Crystal system	**Conditions imposed on cell geometry**	**Minimum symmetry requirement**
Triclinic	None	None
Monoclinic	$\alpha = \gamma = 90°$ (b is the unique axis; for macromolecules this is a 2-fold axis or screw axis) or: $\alpha = \beta = 90°$ (c is the unique axis)	One 2-fold axis
Orthorhombic	$\alpha = \beta = \gamma = 90°$	Three perpendicular 2-fold axes
Tetragonal	$a = b$; $\alpha = \beta = \gamma = 90°$	One 4-fold axis
Trigonal	$a = b$; $\alpha = \beta = 90°$; $\gamma = 120°$ (hexagonal axes) or: $a = b = c$; $\alpha = \beta = \gamma$ (rhombohedral axes)	One 3-fold axis
Hexagonal	$a = b$; $\alpha = \beta = 90°$; $\gamma = 120°$	One 6-fold axis
Cubic	$a = b = c$; $\alpha = \beta = \gamma = 90°$	Four 3-fold axes

Table 2.1. The seven crystal systems.

2.1.3. Theory of X-ray diffraction by crystal

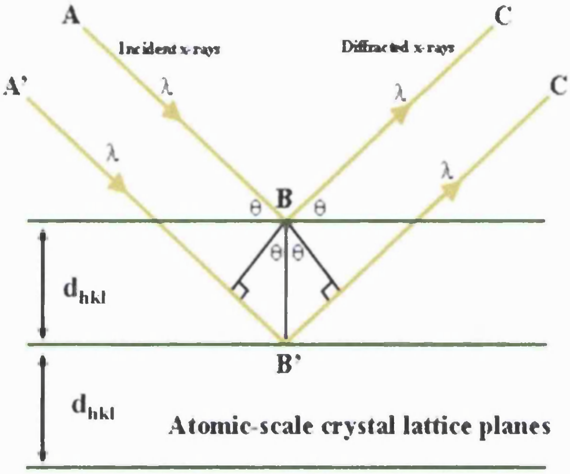
X-ray diffraction occurs by the interaction of photons with electrons within the atom, resulting in what can be thought of as oscillating electrons. When the electron returns to its ground state it releases energy in form of radiation at the same wavelength as the initial photon but in a different direction. The intensity of the scattered radiation is proportional to the number of electrons within the atom.

2.1.3.1. Bragg's law

If the weak scattering vector from a set of atoms is periodically repeated in the three dimensional order of a crystal, the X-rays scattered by these different atoms are in phase, yielding a wave of higher intensity. This is known as constructive interference which can be observed as the diffraction spots on the detector and is mathematically described by Bragg's law in the following equation (Bragg & Bragg. 1913) (Figure 2.2).

$$n\lambda = 2d_{hkl}\sin\theta$$

Bragg's law has to be satisfied to obtain a diffraction pattern. From the equation it can be observed that the distance between lattice planes is inversely related to the angle of incident beam. Thus by altering the angle of incidence by rotating the crystal, Bragg's law can be satisfied for a given wavelength.



Bragg's Law

$n\lambda = 2d_{hkl}\sin\theta$

where,

n = an integer,

λ = wavelength of incident beam,

d_{hkl} = distance between reflecting planes of atom and

θ = angle of incidence

Figure 2.2. Schematic description of Bragg's law in real space.

2.1.3.2. Reciprocal lattice and construction of Ewald sphere

To understand the crystal lattice from the experimentally accessible information, diffraction pattern; it is useful to consider diffraction pattern as lattice, which is inversely related to the real, crystal lattice. When the real, crystal lattice has axial dimensions a, b and c; the reciprocal lattice (diffraction lattice) is described as a*, b* and c*. The Ewald sphere illustrates Bragg's law in three dimensions. According to the Ewald construction, the radiation of wavelength λ is represented by a sphere of radius $1/\lambda$ (Figure 2.3). The crystal is represented by the reciprocal lattice, with its origin at a point on the Ewald sphere where the direct beam leaves the sphere. At instances when a point of the reciprocal lattice is in contact with the Ewald sphere those reciprocal lattice points meet the Bragg's law and diffraction occurs. To obtain diffraction from a number of reflections, the reciprocal lattice is moved to the surface of Ewald sphere, which is achieved by simple rotation of the crystal.

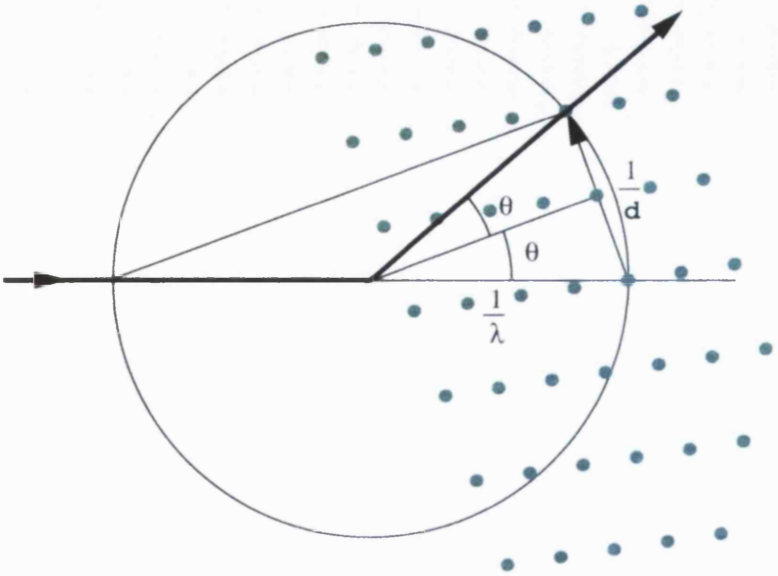


Figure 2.3. The construction of Ewald sphere (adapted from Dauter. 1999).

2.1.4. Data collection strategy

The Ewald sphere shows that diffraction data can be collected by simple rotation of crystals. However, all crystals are not perfectly ordered but have imperfect order, or mosaicity. As a result of this mosaicity, the diffracted X-ray beam is not a perfect linear beam but has a conical shape. This results in Bragg's condition being satisfied over a larger angle, resulting in the reflection forming ellipsoid shapes that are spread over a number of images in a standard rotation data collection of rotating a crystal in a succession of small angular ranges (typically 0.5-1.0 degrees and around 100 individual images) to collect the majority of available unique data (Rhodes. 2006).

The optimal strategy in terms of crystal rotation is determined by the spacegroup so as to obtain the unique part of data from the reciprocal lattice and therefore a complete data set.

2.1.5. Data processing

Diffraction data collected from exposing the crystal to X-rays (see section 2.5 below) are systematically processed where the collected diffraction spots (or intensities) are indexed, integrated and scaled together for further structure solution.

2.1.5.1. Indexing and integration

Several software packages such as MOSFLM (Leslie. 1999), XDS (Kabsch. 2010), HKL2000 (Otwinowski *et al.* 1997) are available for processing of collected reflections. The software can sort through the entire data set collected to detect the position of each reflection and the unit cell dimensions by taking into account factors such as X-ray wavelength, position of measured reflections and detector distance. Once the unit cell dimensions are determined the Bravais lattice can be assigned based on the highest fitting symmetry. With the determination of unit cell dimensions and Bravais lattice, the position of reflections in reciprocal space can be predicted (described as Miller indices *h, k, l*). The data set is then integrated by compiling all of the reflection data together, along with related Miller indices and the intensities (*I*) of the reflection. An error term is calculated from the integrated intensity associated with the background, termed as $I/\sigma I$.

2.1.5.2. Scaling of reflection

Scaling or data reduction is the process where a set of mathematically determined correction (scales) is applied and the data from each collected image are placed on a common scale. This is important as it improves consistency and maximises the number of sufficiently accurate measurements that can be used (Rhodes. 2006). Scaling is often carried out in 'bins', or groups of data, usually based on resolution and containing at least 100 – 200 reflections (McRee. 1999). Thus, scaling gives a suitable means for applying all data on arbitrary scale and allows poorly measured reflection to be rejected.

2.1.5.3. Structure factor and phase problem

The diffraction of X-rays is caused by the cloud of electrons in the molecule of the crystal and thus reveals the distribution of electron density in the molecules. This density map shows peaks from which, given sufficient resolution, the positions of the atoms can be deduced. The Fourier transform describes the mathematical relationship between an object (described by electron density map) and its diffraction pattern (described by the intensity data from its reflection). The intensity of a given reflection is proportional to the square of the structure factor amplitude which is mathematically described below:

$$F_{hkl} = \sum_{j=1}^{n} f_j e^{2\pi i(hx_j+ky_j+lz_j)}$$

where, f_j is atomic scattering factor of the electron cloud of atom j,

x,y,z are fractional coordinates of the electron cloud of atom j, and

$e^{[2\pi . i(hjx+kjy+ljz)]}$ is the periodic function describing the reflected wave.

The electron density at a position in real space $\rho(x,y,z)$, can be calculated by Fourier transform on all structure factor as is given by:

$$\rho(x,y,z) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l} F_{hkl}\, e^{-2\pi i(hx+ky+lz)+ i\alpha(hkl)}$$

where, V is the volume of unit cell,

F_{hkl} is the structure factor amplitude and

$\rho(x,y,z)$ is the electron density at position x, y, z.

Therefore, to calculate electron density at a given point within the unit cell, both amplitude (F_{hkl}) and phase (α(hkl)) needs to be known. The amplitude can be derived from intensity I(*hkl*) from the diffraction experiment. However, the phase information cannot be directly derived from the diffraction experiment and this is called the 'phase problem'.

The phase problem can be overcome by several direct and indirect techniques such as the Patterson function, molecular replacement, isomorphous replacement and anomalous dispersion. In this thesis molecular replacement is used to obtain phase information and hence this technique is described in detail.

Patterson function

A Patterson function is calculated by inverse Fourier function of the structure factor intensity with the phases set to zero. The Patterson map produces a three dimensional map with peaks corresponding to atoms in the real space. However, for large macromolecule systems which have thousands of atoms it is difficult to interpret the Patterson map. However, this technique is used to solve the substructure of heavy atoms and anomalous scatters in molecule. It also helps to determine the orientation of search model in molecular replacement.

Isomorphous replacement

This technique involves the introduction of strong scattering elements such as heavy atoms into the crystal. The position of these heavy atoms can be determined by a difference Patterson function and initial approximate phases can be determined. However, soaking of crystals with heavy atoms can cause change or even destroy the lattice structure. As a result the obtained data is no longer isomorphous with the native crystal (heavy atom free crystal) and the phase information thus obtained would be in error (Rhodes. 2006).

Anomalous dispersion

All elements scatter X-rays anomalously at wavelengths close to their absorption edge(s) this anomalous scattering can be used to deduce phases. Only atoms with a strong anomalous signal (typically metal atoms, but also lighter atoms such as sulphur) are used for this technique. The advantage of this technique is that the data is collected from a single crystal and hence the problem of non-isomorphous or destroying of the crystal lattice is avoided.

Molecular replacement

This is one of the most widely used techniques for determining phase information due to the increasing number of structures in PDB and/or NDB. This technique uses calculated phase information of the solved homologous structure to the structure factor of the unknown target macromolecule crystal. This method can only be applied when the level of sequence identity between two structures is typically greater than 25%, but is very case dependent. Molecular replacement is carried out by systematic rotation and translation function of the homologous model, so that it aligns with the unknown target structure in crystal. The known phases from the aligned search model are used with amplitudes from the diffraction of the target structure to calculate an initial electron density map. This technique involved six parameters consisting of three rotational functions and three translational functions. Molecular replacement programs AMORE and MOLREP (Collaborative Computational Project Number 4. 1994) uses the traditional rotation and translation functions based on Patterson methods. More recently, software such as PHASER (McCoy *et al.* 2007), make use of 'maximum likelihood' which takes into account the error in the model and improves the success rate of molecular replacement with poorer model and/or data.

2.1.6. Structure solution

Once the phase information is obtained, an initial electron density map can be calculated. A model is built to fit the electron density map. The model building primarily depends upon the quality and resolution of the electron density map. The initial model obtained needs to be refined by adjusting the model until the difference between the observed and calculated structure factor can be minimised. This is usually done by adjusting the model to better fit the electron density. Every atom is described by three positional (x, y, z) parameters and one temperature factor parameter. However, the ratio of measured reflections to the number of atoms of the macromolecular system (observation-to-parameter ratio) is generally poor. This is improved by including stereochemical data such as chirality, bond length, torsion angles and atomic contacts. This stereochemical data can be used in refinement in two ways:

2.1.6.1. Rigid body refinement

In this type of refinement subunit or blocks of the macromolecule subunits are treated as single entities with x, y, z and rotation angles varied greatly reducing the number of parameters refined. This method is only useful for early stages of refinement.

2.1.6.2. Simulated annealing

This method uses molecular dynamics for the refinement of macromolecular structures which is used in conjunction with other refinement techniques. The model is 'heated' in a molecular dynamic simulation, before being cooled back down (Brünger *et al.* 1987). This allows the model to adopt the most favourable low energy conformations rather than being trapped in a local minimum.

2.1.6.3. TLS refinement

TLS (translation/liberation/screw) refinement predicts the local positional displacement of atoms in a crystal structure, based on the assumption that all atoms are considered to be part of a rigid body. Standard anisotropic refinement requires high resolution and more parameters. TLS refinement helps to perform anisotropic temperature factor refinement on defined blocks of the macromolecule structure thereby using less parameters and can therefore be done at a medium resolution between 2.5 – 3.0Å.

2.1.6.4. Restrained refinement

This refinement employs geometric restraints on model such as bond length, bond angle, torsion angle, bond planarity and van der Waals contacts. The atomic positions are variable which ensures small parts of the structure can be easily moved. This method is applied throughout the process of refinement; however, it is particularly useful at the latter stages of refinement.

2.1.7. Structure validation

While defining the model by interpreting the electron density map, it is possible that the structure would contain errors. Therefore, to check the accuracy of the structural model, R-factor calculations are used (Rhodes. 2006). The R-factor compares the observed structure factors $|F_{obs}|$ with the structure factor calculated from the current model $|F_{calc}|$, over a group of reflections.

$$\text{R factor} = \frac{\sum_{hkl} |\text{Fobs} - \text{Fcalc}|}{\sum_{hkl} |\text{Fobs}|}$$

As the R-factor becomes close to zero the observed and calculated structure factors agree more closely. However, use of the R-factor alone can be biased towards the model and can lead to over-fitting of the model to the experimental data. A solution to this problem was suggested by Brünger to set aside a small fraction of reflection data (5-10%) as a so-called test set (Brünger. 1992). This test set is removed from all refinement procedures and is therefore unbiased. It is constantly monitored during refinement.

$$\text{R free} = \frac{\sum_{hkl\,(free)} |\text{Fobs} - \text{Fcalc}|}{\sum_{hkl\,(free)} |\text{Fobs}|}$$

Both R-factor and R-free are used in assessing the progress of refinement of a molecular model. Steady decrease in the R-factor whilst the R-free remains stable is an indication of over-fitting of the data. Thus R-free is an important unbiased indicator of refinement. The quality of the model can be assessed by programs such as WHATIF (Vriend. 1990), 3DNA (Lu & Olson. 2003) and PROCHECK (Laskowski *et al.* 1993). These programs assess the geometry and stereochemistry by comparing the bond length and bond angles to a library of ideal values. They also search for residues in close contact within the asymmetric unit or between symmetrically related molecules. Protein macromolecular structures can also be validated by using a *Ramachandran plot.*

2.2. Neutron crystallography

Neutron crystallography has been used as a powerful method to obtain structural information on hydrogen (or deuterium) atoms in macromolecules. Hydrogen atoms have important roles in biological molecules, such as those involved in reaction mechanisms or in their hydrogen-bonding and hydration structures. X-ray crystallography can also determine structural information on hydrogen atoms, but it requires structure to be determined at ultra-high resolutions (1Å and less). Obtaining crystal structures at such resolution can be extremely difficult or impossible for most macromolecular systems while neutron crystallography can provide information on hydrogen atoms at medium resolution ranges as low as 3.0-3.5Å. In addition neutron crystallography has no risk of radiation damage, which is an important factor to consider for X-ray crystallographic studies.

2.2.1. Neutron scattering

Both X-ray and neutron crystallography follows the same principles of Bragg's law, with their major difference being their nature of interaction with atoms. Whilst X-rays interact with the electrons, neutrons are scattered by the nuclei of the atom. When neutrons interact with the nucleus of an atom, it gives rise to coherent and incoherent scattering. Coherent scattering, is the termed used when the frequency of the incident radiation is same as the diffracted radiation. This is used for studying the structural biology of macromolecular systems. Incoherent scattering is caused by the presence of different isotopes in the same sample or the presence of spin resulting in the difference of frequency. This type of scattering is often termed as noise while studying molecular structure. Atoms with strong coherent scattering power can be easily distinguished using neutron crystallography.

2.2.2 Scattering lengths

As mentioned earlier X-ray interacts with electrons and therefore with increase in number of electrons in an atom the scattering length of the atom also increases. However, neutron which are scattered by nucleus of the atom has no such relation with the number of protons and neutrons of the nucleus (Chidambaram & Sikka. 2003). Due to this difference, neutron diffraction and scattering can be unique tools for locating hydrogen atoms and distinguishing isotopes such as deuterium.

Hydrogen comprises about half of the atoms in bio-molecules. The hydrogen atom has only one electron and has a low X-ray scattering length of 2.8 x 10^{13}cm and its isotope; deuterium also has same X-ray scattering length. However, the hydrogen nucleus has a large incoherent neutron scattering cross-section, which contributes to high backgrounds in neutron diffraction experiments. In addition, it has a negative coherent scattering length (-3.74 x 10^{13}cm) which is half the magnitude compared to the positive scattering lengths of C (6.65x10^{13}cm), N (9.40x10^{13}cm) and O (5.80x10^{13}cm) atoms. While deuterium, an isotope of hydrogen, has a strong positive coherent neutron scattering length (6.67 x 10^{13}cm), which is comparable to the scattering lengths of C and O.

2.2.3. Hydrogen and deuterium

Due to the negative coherent scattering length of hydrogen atoms, at a medium resolution of diffraction there is a partial cancellation of hydrogen density from its covalently-bonded C, N and O atoms (Chidambaram & Sikka. 2003). To avoid the high background from incoherent neutron scattering and cancellation of density due to negative coherent scattering of hydrogen atom, it can be replaced by deuterium atoms. This helps to reduce the signal to noise ratio and clear the density maps for better interpretation of the position of the atoms.

This hydrogen-deuterium exchange can be performed by co-crystallising and/or soaking macromolecular crystal systems in deuterated conditions. This allows replacement of water and labile hydrogen atoms such as those covalently bound to O and N atoms. It is also possible to prepare samples with 100% deuteration ("perdeuterated"). Perdeuterated samples of protein can be obtained by performing gene expression in a deuterated growth media. The ILL-EMBL Deuteration Laboratory (D-lab), is one such user facility where proteins, nucleic acids, enzymes and lipids are regularly prepared partially or fully deuterated for neutron studies.

The principal challenges of neutron crystallography are the low flux of neutron beam and the necessity to obtain crystals of sufficiently large volume, approximately 1 mm^3. With the recent developments in neutron crystallography instruments and the possibility to prepare perdeuterated crystals, it is now possible to study the crystal structures of macromolecular systems with a crystal volume of only 0.1-0.2mm^3 (Blakeley *et al.* 2010).

2.2.4. Structure determination and refinement

Determining phase information is also a problem with neutron crystallography. With this technique, due to the arbitrary relation between atom nucleus and their neutron scattering length, there are no 'heavy atoms' in neutron scattering. Thus phase determination using anomalous scattering is difficult. Further due the significant scattering lengths of hydrogen and deuterium atoms, there would be more atoms to be located at the phase determination stage making the Patterson map crowded. A neutron shake-and-bake algorithm has been used to solve the structure of 199 atom cyclosporine by only using experimental neutron diffraction data (Hauptman & Langs. 2003). A study was performed where the difference in neutron scattering lengths of isotopes was used to determine phase information similar to the isomorphous replacement technique (Chidambaram & Sikka. 2003; Johnson. 1967). Some nuclei such as ^{113}Cd, ^{149}Sm,

^{155,157}Gd, etc have high resonant absorption for thermal neutron and were used in late 1970's for phase determination by anomalous scattering technique. However, this technique is not widely used due to the large time required for data collection due to higher absorption of neutrons in crystals containing anomalous scatterers (Chidambaram & Sikka. 2003).

Phase information for neutron data is generally obtained by performing joint X-ray and neutron data collections. The positions of non-hydrogen atoms in the crystal are determined from the X-ray structure. The phases calculated from the X-ray structure is used as a starting set for neutron phase determination.

There have been developments in the software packages available for the joint X-ray and neutron structure determination and refinements. *nCNS* a structure solution program has been developed which combines global X-ray, neutron and energy refinement with maximum-likelihood simulated annealing refinement (Brünger *et al.* 1998). *PHENIX* can be used for automated determination of macromolecular structure using joint X-ray and neutron data (Adams *et al.* 2010).

In spite of the challenges involved with neutron crystallography, it is an important and growing technique to determine information regarding hydrogen atoms. With the recent development in neutron data collection instruments (discussed later), the availability of perdeuterated samples and with advances in structure refinement methods it is now possible to obtain neutron structures in less time and with smaller crystals.

2.3. Small angle X-ray scattering (SAXS)

SAXS can be used to study structural information (mainly shape and size) of biological systems in solutions at low resolution (10-20Å). It can be used to study formation of protein-protein or protein-DNA complexes in real time without need to trap any intermediate conformations.

The scattering vector (Q) is the difference in the direction of the incident wave and scattered wave and it is given by:

$$Q = 4\pi\ (\sin\theta)\ /\ \lambda$$

Plotting the measured scattering intensity as a function of Q provides the SAXS scattering curve of the particle. The low Q region of the scattering curve can be defined by the Guinier approximation and the interpretation of this region gives the experimental parameters $I(0)$ (intensity at zero angle) and the R_g^2 (radius of gyration of the particle). The Guinier approximation can be describes as

$$I(Q) = I(0)\ \exp(-R_g^2 Q^2/3)$$

The radius of gyration can be defined as the root mean squared distances of all the atoms to the centre of the particle, thus it describes the particle size. $I(0)$ can be used to describes the molecular weight of the particle at the given concentration of the particle, provided the particles in solution are identical and non-interacting (Jacrot & Zaccai. 1981). However, if interparticle interaction is observed then these interparticle correlations have to be taken into considerations in the analysis of the scattering curve. The intensity as a function of the magnitude of the scattering vector is given as:

$$I(Q) = F(Q)\ S(Q)$$

where, $F(Q)$ is the interparticle form factor and $S(Q)$ is the interparticle structure factor. For determining the shape of the particle, the distance distribution function $p(R)$ which describes the average distance and their distribution within the particle is real space is calculated by Fourier transform of $I(Q)$. The shape of this $p(R)$ function also provides important information regarding the shape of the particle.

The analysis of the SAXS scattering curves is done by using different programs which are part of the ATSAS data analysis suite developed by Svergun and group at EMBL, Hamburg. The programmes that have been used in this study are described.

a) PRIMUS and GNOM

PRIMUS (Konarev *et al.* 2003) is used for simple treatment of scattering curves by background subtraction, averaging of curves. The Guinier region can be then determined to calculate R_g and $I(0)$ for the particles. The Guinier region determination is used as a guidance to evaluate the quality of the data for calculation of the $p(\mathrm{R})$ function in GNOM (Semenyuk & Svergun. 1991). The $p(\mathrm{R})$ function can be improved by manually changing the $\mathrm{D_{max}}$. GNOM analysis also calculates R_g and $I(0)$ which can be compared with the PRIMUS Guinier approximation and therefore provides additional criteria to judge the $p(\mathrm{R})$ function.

b) Ab initio modelling

Following the satisfactory $p(\mathrm{R})$ function calculation, the three-dimensional model of the particle can be calculated. Several modelling programs are available such as the DAMMIN, DAMMIF, GASBOR, MONSA. In this study fast mode DAMMIF method was used. This technique uses the simulated annealing to construct a compact interconnected model yielding a scattering pattern that fits the experimental curve (Franke & Svergun. 2009).

c) Rigid body fitting

The simulation of solution scattering from macromolecules with known atomic structure and fitting it to the experimental scattering curves from SAXS can be performed using CRYSOL. This program uses multipole expansion of the scattering amplitude to calculate the spherically averaged scattering pattern and takes into account the hydration shell (Svergun *et al.* 1995). The solution scattering curve of the PDB structure is calculated and is fitted to the SAXS experimental data.

d) Multi-component analysis

OLIGOMER fits an experimental scattering curve of a multicomponent mixture to find the volume fraction of individual components in the mixture. The experimental scattering intensity $I(s)$ from a mixture of different components is written as

$$I(s) = \sum w_i I_i(s))$$

where, w_i and $I_i(s)$ are the volume fraction and scattering intensity for the i number of components. Given the intensities (form factor) of individual components, OLIGOMER finds the volume fraction by solving linear equations using algorithm of nonnegative or unconstrained least-square to minimise the discrepancy between the experimental and calculated scattering curves (Svergun *et al.* 1994-2004).

2.4. Surface Plasmon Resonance (SPR)

This technique can measure biomolecular interactions in real time. One of the interacting molecules is immobilised on a sensor chip, referred to as ligand (DNA in our experiment) and the other is free in solution and passed over the chip surface, referred as the analyte (protein and/or protein-ligand solution in our experiment). SPR studies can be used to determine whether there is any binding between the interacting molecules and how strong the interaction is by measuring binding affinity and it can also be used to measure the actual association and dissociation rates.

SPR is a physical process that occurs when plane-polarised light hits a metal film (generally gold) under total internal reflection conditions. Surface plasmons are surface electromagnetic waves that propagate in a direction parallel to the metal/dielectric interface. Plasmon velocity (and therefore the momentum) changes when the composition of the medium changes. The change in momentum causes change in angle of incident light at which resonance occurs. This change in resonant angle is accurately measured and used for SPR studies. The SPR angle mainly depends on the properties of metal film, wavelength of the incident beam, temperature and the refractive index of the media on either side of the metal chip (BIACORE. 1998). The SPR signal is directly dependent on the change in refractive index of the medium on the surface side of the metal chip as all the other components are kept constant. The binding of biomolecules result in change of the refractive index on the sensor surface which is measured as a change in resonance angle.

The association is the first phase of biomolecular interaction which shows the binding of analyte (protein and/or protein-ligand) and ligand (DNA). On binding the ligand and analyte remain bound for a random amount of time, following which dissociation occurs. The reaction between the immobilised ligand (L) and flowing analyte (A) follows a pseudo first order kinetics (Jonsson *et al.* 1991). During the association phase, the complex [LA] increases as a function of time and can be given as:

$$\frac{d[LA]}{dt} = ka.[L].[A] - kd.[LA]$$

where, **association rate constant (k_a)** describes the rate of complex formation, i.e. the number of [LA] complexes formed per second in 1M solution of [L] and [A], while the **dissociation rate constant (k_d)** describes the stability of the complex, i.e. the fraction of the complexes that decay per second.

The analyte is continuously flowed over the immobilised ligand and therefore, the binding reaches a steady state, where the net binding is zero. The equilibrium constants are determined by measuring the concentration of free interactants and the complex at equilibrium:

$$\frac{ka}{kd} = \frac{[LA]}{[L][A]} = Ka \ (or) \ \frac{kd}{ka} = \frac{[L][A]}{[LA]} = Kd$$

where, **K_a** is the **association constant** and **K_d** is the **dissociation constant**. The best parameter to compare the strength of binding is the dissociation rate constant (k_d) because this parameter gives the time an interaction exists.

The SPR data collected is analysed by fitting the curves obtained. Software is available to find the best fit values of the variables in the model (rate constants, affinities, etc.) which is then interpreted and a model must be chosen. The best fit of the data provides the best model. Several kinetic models exist such as one-to-one, mass transfer, decaying model, etc. For this study the 1:1 Langmuir model was chosen, which assumes the reaction between the [L] and [A] follows the pseudo first order kinetics (Jonsson *et al.* 1991). The quality of the fit and the model chosen can be validated by looking at the residual plot, Chi^2 value and the SE (standard error) value. The width of the residual plot at the given time indicated the noise level and the shape reveals systematic difference between the fitted curve and experimental data (BIACORE. 1997). SE indicated the sensitivity of the fit to variations in the parameters. A high SE would indicate that the variation in the parameter value will not have much effect on the closeness of the fit (BIACORE. 1997).

2.5. Instruments

This section describes the source of X-ray and neutron used for the crystallography and SAXS experiments and also outlines the SPR instrument used.

2.5.1. X-ray source

X-rays are electromagnetic radiation of wavelength 0.1-100Å. X-rays for crystallographic studies can be generated by transition of the electron from a higher energy shell to a lower energy resulting in emission of X-ray photon. Rotating anodes and synchrotrons are commonly used for X-ray crystallographic studies, while Bio-SAXS studies are often performed at high intensity synchrotron sources.

At synchrotrons, electrons are circulated in a storage ring at velocities near the speed of light. When the path of electrons is changed using strong magnets, X-rays are emitted. The X-rays thus generated are channelled and conditioned through experimental facilities ("beamlines") using devices such as slits, attenuators, monochromators and mirrors.

The X-ray crystallographic and SAXS studies in this thesis were performed using beamlines at ESRF, Grenoble. The macromolecular crystallography beamlines (MX beamlines) include high energy tunable beamline (ID 14-4, ID 23-1 and ID 29), fixed wavelength beamlines (ID 14-1 and ID 14-2), microfocus beamline (ID 23-2). The Bio-SAXS beamline is ID 14-3.

In this study the tunable beamline ID 23-1 was used to obtain anomalous information on crystals (described in Chapter 4 and 5) while other crystallographic data collection were performed on ID 23-2, ID 14-4 and ID 29 (see Chapters 4 and 5). The Bio-SAXS experiments (Chapter 6) were performed at the ID 14-3 beamline dedicated for solution scattering studies.

2.5.2. Neutron source

Neutrons can be generated by fission chain reaction and by spallation. The spallation technique uses high energy protons to bombard metallic targets and thermodynamically it is more efficient than fission at generating neutrons. However, spallation requires high energy to accelerate the protons. Neutrons are also directed using mirrors and monochromators like X-rays. However, as neutrons are electrically neutral their direction cannot be easily maintained and hence the neutron sources have poor flux compared to X-ray sources (both rotating anode and synchrotron). Due to the low flux of neutrons, using monochromatic beam would make data collection time longer and hence the white beam or Laue diffraction is used for neutron data collection. The ILL, Grenoble is one of the strongest nuclear fission reactors and the neutron experiments for this study were performed here using the LADI-III instrument (Chapter 3).

LADI-III uses a large cylindrical detector composed of neutron sensitive image plate which completely surrounds the sample and allows large number of Bragg's reflection to be recorded simultaneously. Data collection uses the quasi-Laue method to provide rapid survey of reciprocal space while reducing background scattering and reflection overlap compared to use of the full white beam (Blakeley *et al.* 2008). An improved reading system is also employed which is located within the cylindrical detector and provides a two to threefold gain in neutron detection.

2.5.3. SPR Instrument

The BIACORE X instrument was used for the SPR experiments (Chapter 6). The components of this instrument are the liquid delivery pump which maintains the flow of liquid, an integrated fluid cartridge (IFC) where the samples can be injected and two flow cells formed by the IFC pressing against the sensor chip. A sensor chip is required to immobilise a ligand to study the binding affinities. Several sensor chips are available, for this study the sensor chip SA with streptavidin layer which binds biotinylated ligands (in our study, biotinylated DNA was used).

Chapter 3

Crystallisation studies to obtain large volume crystals

Neutron crystallography is a well adapted technique for the determination of the location of hydrogen (or deuterium) atoms and the protonation states in a macromolecule. Historically, the principal challenge for this technique is to obtain crystals of sufficiently large volume, approximately 1mm^3. With the recent developments in neutron crystallography instruments and the possibility to prepare samples with 100% deuteration ("perdeuterated"), it is now possible to study the crystal structures of macromolecular systems with a crystal volume of only 0.1-0.2mm^3 (Blakeley *et al.* 2010). However, is difficult to obtain or synthesise perdeuterated DNA sequences. Nevertheless a neutron structure of A-form DNA grown in buffer prepared with deuterated (heavy) water has been successfully studied with crystal volume of 0.06mm^3 (Leal *et al.* 2009).

This chapter describes the steps involved in obtaining crystals of DNA and MGBL complexes of sufficient quality for neutron diffraction studies. As part of the process of optimising conditions to obtain large volume crystals for neutron studies, several high quality smaller crystals were obtained which were used for the high-resolution X-ray crystal structures of DNA ligand complexes described in Chapter 4 and Chapter 5. The systematic screening, crystallisation condition optimisation and different crystallisation techniques used for obtaining neutron size crystals of DNA ligand complex system are described. The chapter also details the preliminary neutron test results obtained.

3.1. Ligand selection and preparation

The MGBLs belonging to the diamidine group have been exhaustively studied by crystallography and biophysical techniques. However, these studies have never directly revealed the hydrogen atom locations and therefore neither the detailed hydrogen-bonding nor the hydration networks important for the drug binding. Several MGBLs synthesised by the groups of David Boykin and David Wilson at Georgia State University were kindly provided for research purposes. For the X-ray and neutron structural studies, the selection of the ligands was based upon their structural features and potential hydrogen-bonding patterns with DNA (see schematics in Figure 3.1). Several of the ligands, such as berenil, Hoechst 33258, DB75 (Furamidine) and DB921, were selected as prime candidates for neutron diffraction studies. The reasons for the selection are based on the fact that the existing X-ray crystal structure of DB921 in complex with DNA showed that water molecules play an important role in the ligand-DNA interaction (Miao *et al.* 2005) and that previous crystallisation studies have shown that berenil, Hoechst 33258 and DB75 provide large volume crystals. The ligand solutions were prepared using 30mM of sodium cacodylate buffer at pH 6.5 (or pD 6.9). The stock solutions of the ligands were prepared at 10mM concentration.

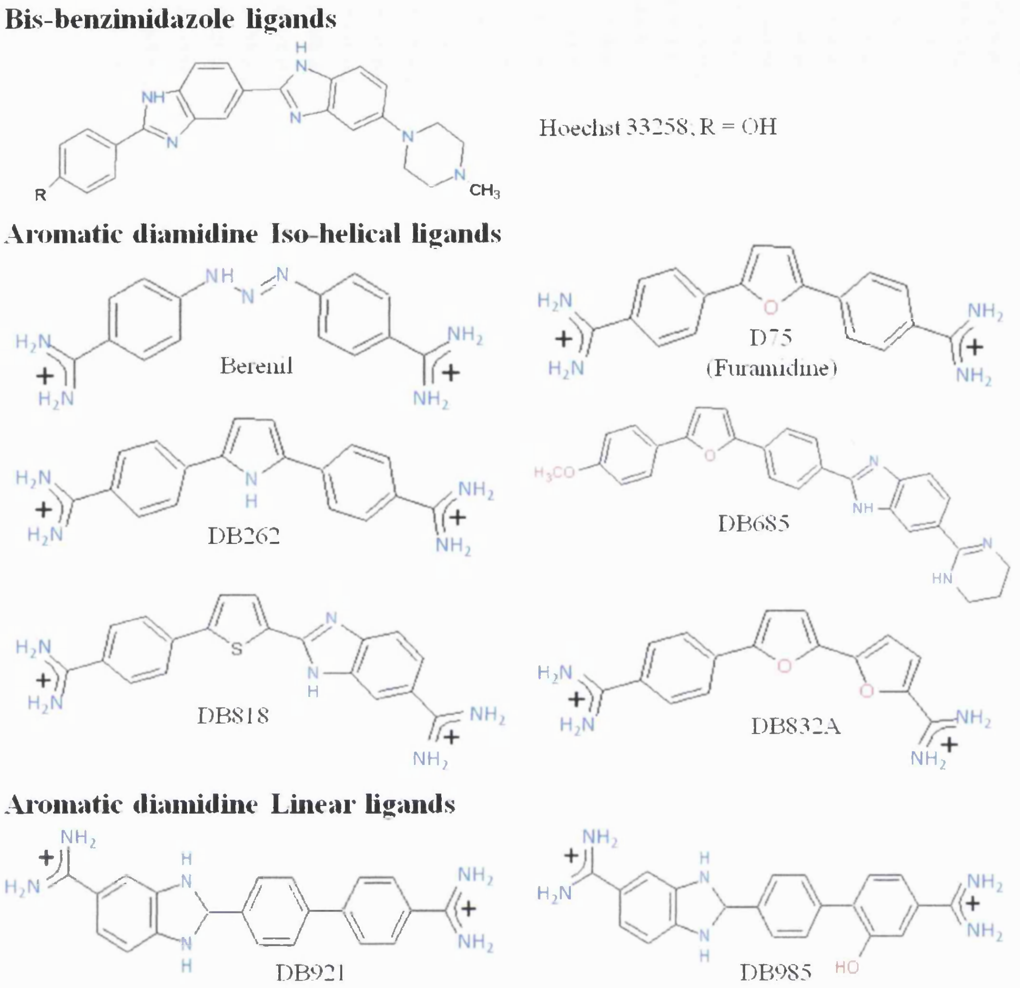


Figure 3.1. Chemical schematics of bis-benzimidazole and aromatic diamidine ligands selected for X-ray and neutron studies.

3.2. DNA selection and preparation

In a quantitative DNase I footprinting study on a representative of each class of minor groove binding ligands (MGBLs) done by K.R. Fox and his group (Abu-Daya *et al.* 1995), it was observed that MGBLs have a stronger preference towards the AATT and AAAA binding sites than TTAA and TATA. Based on these observations, the following A-T rich DNA sequences were selected to carry out the crystallographic studies:

- CGC GAA TTC GCG (A_2T_2; 12monomeric sequence)
- CGC AAA TTT GCG (A_3T_3; 12monomeric sequence)

The cytosine-guanine caps are essential to ensure stacking of the bases and to maintain the integrity of the DNA. The HPLC purified single stranded DNA sequences were ordered from Eurogentec Ltd and IBA. The DNA sequences were further purified by anion exchange FPLC to obtain pure samples for crystallisation.

3.2.1. DNA purification

Chemically synthesised oligonucleotides usually contain imperfect synthesis products which can impede the crystallisation experiment. Although HPLC purified oligonucleotides were ordered commercially, an additional step of purification using anion exchange chromatography was performed to ensure the purity and quality of the DNA.

The mono Q HR 10/10 anion exchange column (GE Healthcare) was used on an ÄKTA FPLC system (GE Healthcare). The anion exchange chromatography was performed at 4°C in the cold room. The column was equilibrated with 90% buffer A containing 10mM NaOH and 10% buffer B containing 10mM NaOH and 1.0M NaCl at 4ml/min. The DNA sample (1mM concentration) was prepared with 800µl of 90% buffer A and 10% buffer B. The sample was centrifuged 10000rpm on a benchtop Eppendorf centrifuge to remove any large particles. The DNA sample was injected onto the column with an isocratic flow of 10% buffer B. The DNA was then eluted over an increasing salt concentration gradient flow of 40-75% of buffer B at 4ml/min (Figure 3.2). The fractions of 1ml were eluted into fraction collection tubes containing 100µl of neutralising solution (1M HEPES, pH 7.5) to neutralise NaOH upon elution.

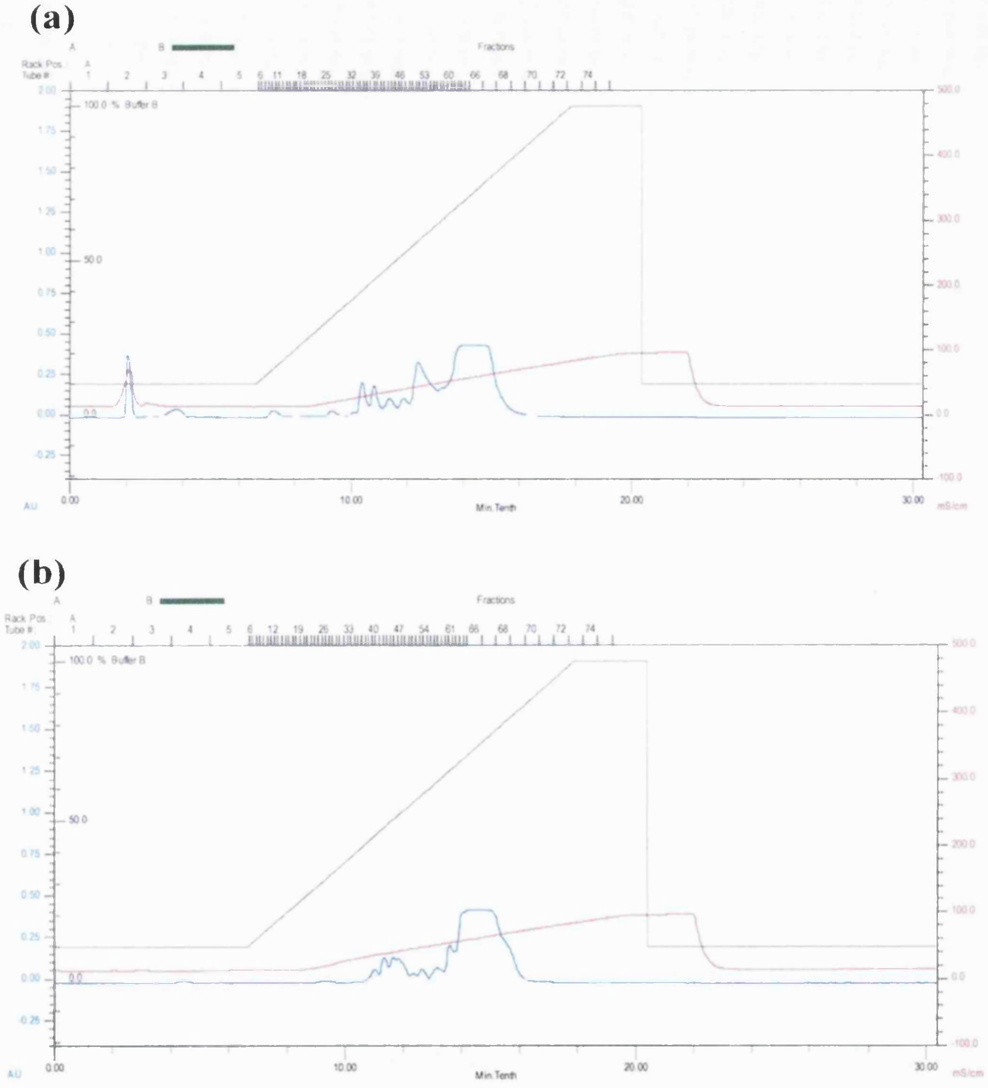


Figure 3.2. UV chromatogram profile of (a) A_2T_2 and (b) A_3T_3 DNA after anion exchange chromatography. The UV absorption spectrum measured at A_{260nm} for nucleic acid is shown in blue, the conductivity in red, and the percent of buffer B in black.

The fractions containing the desired peaks were pooled together and dialysed overnight in the cold room against 4 litres of water (changing water at least every 3-4 hours) using a Slide-A-Lyzer dialysis cassette (Thermo Scientific) with 3500Da molecular weight cut-off. The sample was then concentrated and further desalted using the 5ml D-Salt^{TM}

Polyacrylamide desalting column from Thermo Scientific. The column was first washed with 5 column volumes of distilled water. Following this 500µl of DNA was introduced onto the column. Water was then added onto the column and 500µl fractions were collected in Eppendorf tubes until the DNA sample was completely eluted from the column. Small aliquots of these fractions were measured at A_{260nm} using a UV spectrophotometer to detect the presence of DNA (Figure 3.3). The fractions containing DNA were collected and lyophilised to dryness. The DNA sample was now ready for the further annealing to be used in the crystallisation experiments.

The purification of DNA helped to improve the crystal morphology by reducing the showers of small crystals which otherwise appeared in the crystallisation drops. Moreover, with some ligands it was found then possible to obtain single crystals instead of multiple crystals growing together (Figure 3.4).

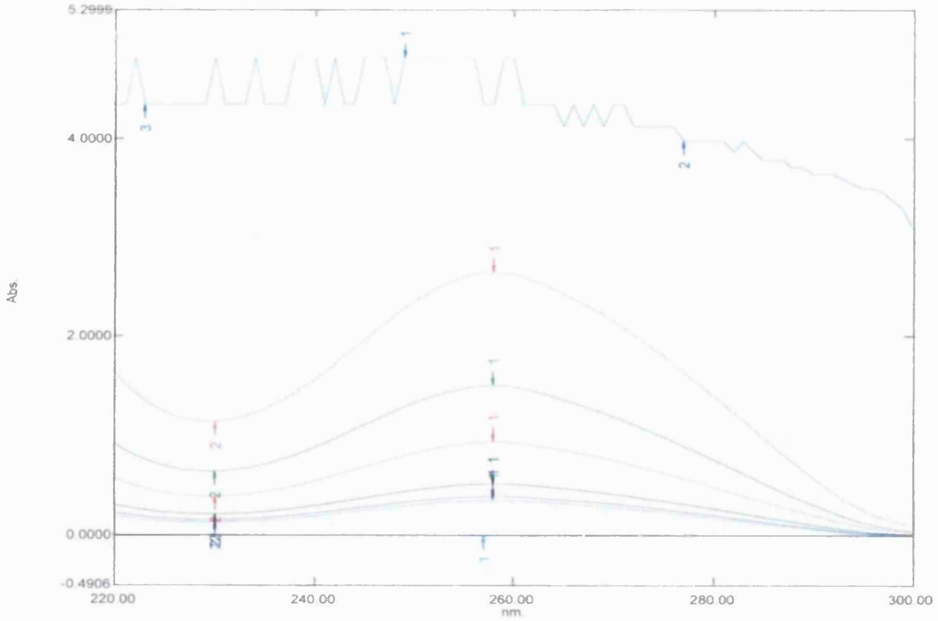


Figure 3.3. UV absorption spectrum at A_{260nm} indicating presence of DNA in different fractions collected after the desalting of DNA.

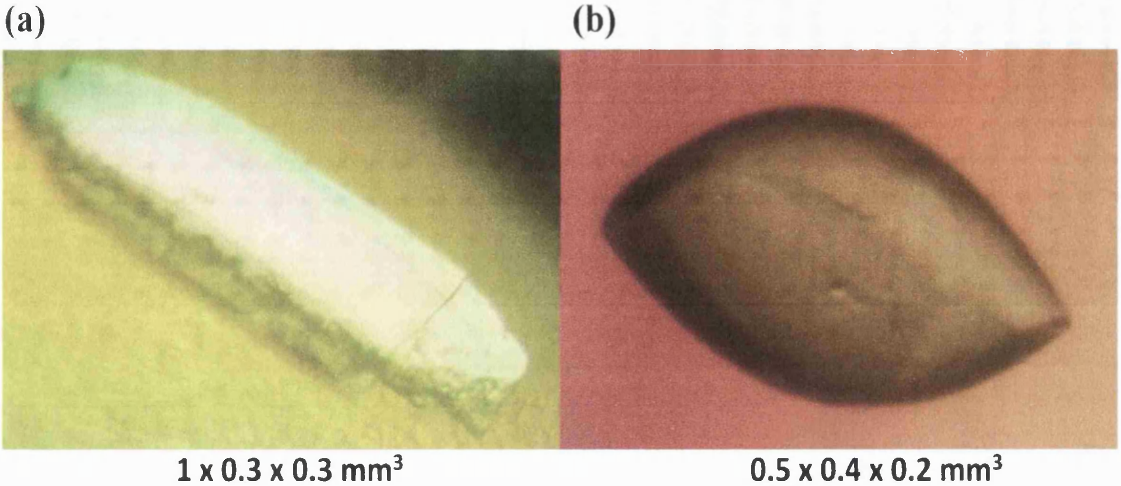


Figure 3.4. Crystals of DB985 in complex with A_3T_3 DNA obtained (a) before DNA purification (b) after purification of DNA, this crystal has better morphology (Note that this picture was taken whilst the crystal was still in growth.)

3.2.2. DNA annealing

The lyophilised DNA was dissolved in heavy water (D_2O) and a UV spectrophotometer (A_{260nm}) was used to determine the quantity of DNA obtained. Approximately 52mg and 42mg of A_2T_2 and A_3T_3 DNA were obtained after the purification step from the 125mg of each DNA ordered. The DNA samples were split into small aliquots and frozen at -80°C to be annealed when required.

Each aliquot of the DNA was then diluted by adding approximately 150µl of D_2O and the absorbance was measured to determine the exact concentration of DNA present in the aliquot. Since the concentration of DNA in each aliquot was different, appropriate volumes of buffer solution, (30mM sodium cacodylate pD 6.9 prepared in D_2O) were added to prepare the DNA for annealing. The DNA sample was annealed by heating on a water bath at 85°C for 15-20 minutes and allowed to cool to room temperature overnight. The annealed sample was again measured by UV spectrophotometer at A_{260nm} to determine the concentration of annealed DNA. This concentration was used to prepare further dilution required for crystallisation conditions.

3.3. Crystallisation

3.3.1. Initial crystallisation screening

Extensive crystallisation studies on the A_2T_2 and A_3T_3 DNA sequences have been carried out in the presence of MGBLs (see references in table 3.1). This prior work was used to provide the starting points for the crystallisation conditions to obtain initial crystals for the MGBLs listed in Figure 3.1. The crystallisation conditions along with the in-drop concentrations are listed in Table 3.1. All of the conditions extracted from literature included varying concentration of magnesium chloride ($MgCl_2$), spermine tetrahydrochloride, 2-methyl-2,4-pentadiol (MPD), sodium cacodylate (pH 6.5), DNA and ligand.

The required concentrations of DNA and ligands were prepared by stepwise dilution using 30mM sodium cacodylate (pH 6.5). MPD was used as the precipitating agent. The crystallisation solutions were first prepared in hydrogenated solutions. After initial hits were obtained, the conditions were adapted for deuterated solutions. This was done by preparing the additives used for crystallisation in D_2O. While using D_2O solutions in buffer there is a change in the acidity calculation, where essentially pD is equivalent to (pH + 0.4) (Krężel & Bal. 2004). Therefore the buffer solution with sodium cacodylate prepared in D_2O solution was made with a pD of 6.9 instead of 6.5. Both the hanging drop and sitting drop vapour diffusion techniques were used. Each crystallisation plate was duplicated at two temperatures: 12°C and 20°C. The conditions 1 to 13 listed in Table 3.1 were set for both A_2T_2 and A_3T_3 DNA sequences with all of the MGBLs listed in Figure 3.1.

It was observed that conditions 1, 4, 5, 6, 7, 9, 12 and 13 were more favourable to obtain crystals whilst the other conditions 2, 3, 8, 10 and 11 did not give any crystals. Hence these conditions were not repeated with deuterated solutions.

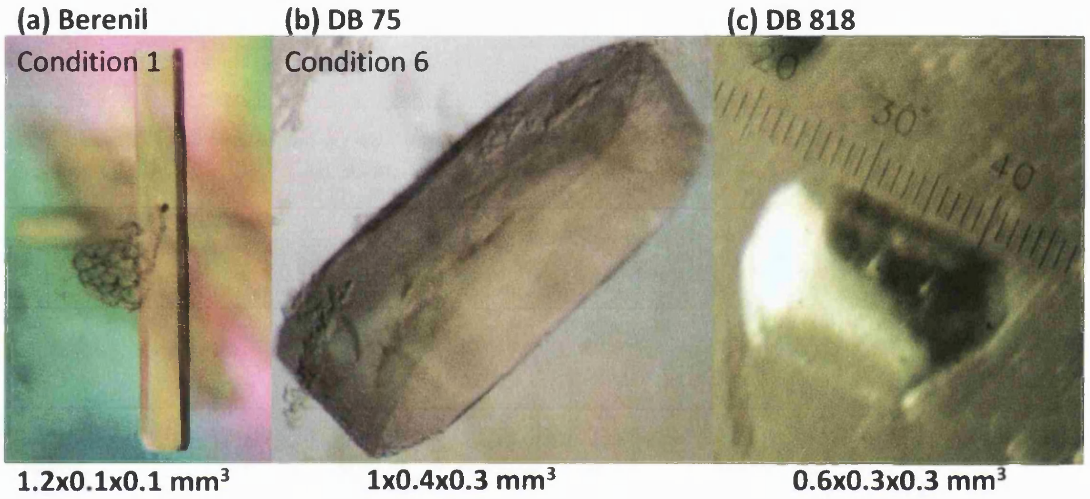
Crystals of iso-helical ligands (berenil, DB75, DB262, DB818 and DB832A) and linear ligands (DB921 and DB985) complexed with A_2T_2 and A_3T_3 DNA sequences were obtained from conditions 1, 4, 5, 6, 7, 9, 12 and 13. Conditions 4, 5 and 12 mainly resulted in showers of small crystals of approximately (0.1 x 0.2 x 0.1) mm^3, which were suitable for X-ray data collection. Conditions 1, 6, 7, 9 and 13 resulted in large crystals of:

- berenil complexed with A_3T_3 DNA obtained from conditions 1 (1.2 x 0.1 x 0.1) mm^3 and condition 6 (0.6 x 0.3 x 0.25) mm^3
- DB75 complexed with A_3T_3 DNA obtained from condition 6 (1 x 0.4 x 0.3) mm^3; condition 7 (0.9 x 0.05x 0.05) mm^3 and condition 13 (0.9 x 0.1 x 0.2) mm^3
- DB818 complexed with A_3T_3 obtained from condition 9 (0.6 x 0.3 x 0.3) mm^3
- DB921 complexed with A_3T_3 were obtained from condition 9 (0.9 x 0.3 x 0.2) mm^3
- DB985 complexed with A_3T_3 obtained from condition 1 (0.6 x 0.2 x 0.1) mm^3; condition 9 (1 x 0.3 x 0.3) mm^3 and condition 13 (1.3 x 0.3 x 0.1) mm^3.

Photographs of several of these crystals are shown in Figure 3.5.

These crystals were first observed in hydrogenated conditions. Following this the conditions were repeated in deuterated conditions (with changes as described above) and similar results were observed at 12°C. It was also observed that at 12°C the crystallisation conditions of Table 3.1 set in hydrogenated and then deuterated conditions provided more reproducible results than 20°C crystallisation trials.

Some of the most promising conditions were further optimised as explained later in section 3.3.2.



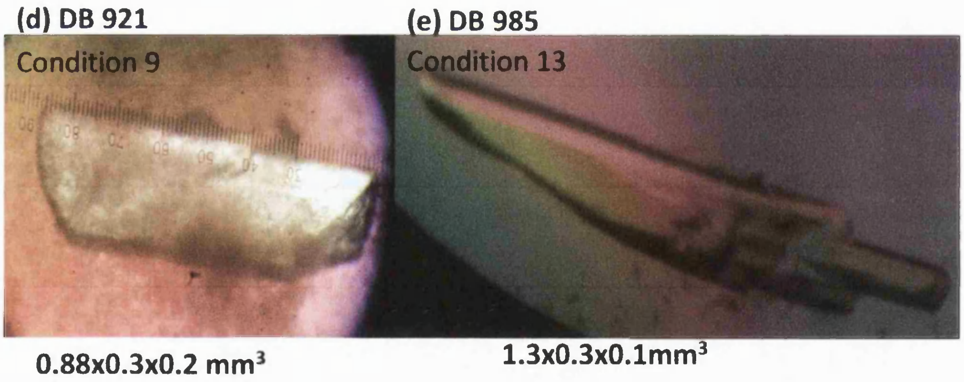


Figure 3.5. Crystals obtained by reproducing literature conditions in Table 3.1: (a) crystal of Berenil with A_3T_3 DNA from deuterated condition 1; (b) crystal of DB75 with A_3T_3 DNA from hydrogenated condition 6; (c) crystals of DB818 with A_3T_3 DNA from hydrogenated condition 9; (d) crystals of DB921 with A_3T_3 from hydrogenated condition 9; (e) crystal of DB985 with A_2T_2 DNA from deuterated condition 13.

	Condition												
	1	2	3	4	5	6	7	8	9	10	11	12	13
Crystallisation method	**(sitting)**	**(hanging)**	**(hanging)**	**(hanging)**	**(hanging)**	**(sitting)**	**(sitting)**	**(hanging)**	**(sitting)**	**(hanging)**	**(sitting)**	**(hanging)**	**(sitting)**
Additive													
$MgCl_2$	31.24mM	9.375mM	4.5mM	10mM	10mM	28.57mM	18mM	3mM	42.87mM	5.83mM	0	54.54mM	166.7mM
MPD	1.0417%	3.125%	5%	6%	5%	11.43%	7.5%	5%	10%	5%	19.5%	9.54%	11.25%
Sodium cacodylate	5mM	20mM	0mM	0mM	2.5mM	0mM	0mM	0mM	0mM	0mM	1mM	0mM	0mM
Spermine	0mM	0mM	0.25mM	0.5mM	1mM	0.257mM	1mM	0.5mM	0.343mM	1mM	0.7mM	0.2272mM	0mM
DNA	0.125mM	0.375mM	0.6mM	0.4mM	0.5mM	0.857mM	0.4mM	0.6mM	0.428mM	0.5mM	1mM	0.909mM	0.825mM
MGBLs	0.167mM	0.5mM	1mM	1mM	0.75mM	0.89mM	0.6mM	1mM	1.4mM	0.75mM	1mM	0.909mM	1.667mM
Drop volume	24µl	10µl	10µl	15µl	6µl	14µl	10µl	10µl	14µl	6µl	6µl	11µl	12µl
Reservoir	50% MPD 500µl	50% MPD 500µl	55% MPD 1000µl	50% MPD 1000µl	50% MPD 700µl	40% MPD 2000µl	40% MPD 1000µl	40% MPD 1000µl	35% MPD 2000µl	50% MPD 700µl	50% MPD 500µl	40% MPD 1000µl	45%MPD 1000µl
Reference	(Munde *et al.* 2007)	(Miao *et al.* 2005)	(Trent *et al.* 1996)	(Trent *et al.* 1996)	(Mallena *et al.* 2004)	(Laughton *et al.* 1996)	(Mazur *et al.* 2000)	-	(Wood *et al.* 1995)	-	(Vega *et al.* 1994)	(Clark *et al.* 1996)	(Nunn *et al.* 1997)

Table 3.1. The different crystallisation conditions for MGBLs as extracted from the literature and used as the starting crystallisation screen for this work. The final concentrations of the additives in the drop are shown.

Two MGBLs, DB685 and DB832A, resulted in micro crystals from these early crystallisation trials (Table 3.1). Hence to extend the screening conditions and in an attempt to obtain better quality crystals a commercially available crystallisation screen, Nucleic Acid Mini screen (NAM, Hampton Research) was tested for these ligands (Appendix I). The NAM screens were prepared by making hanging drop plates with drops containing 1µl each of ligand (2mM) and annealed DNA (1mM) and 2µl of screen solution against 500µl of 35% MPD in the well. Positive crystallisation conditions were obtained for both ligands. Crystals of DB685 were obtained from condition 5 (10% MPD, 40mM Na cacodylate pH 6.0, 12mM spermine tetra-HCl, 80mM KCl and 20mM $MgCl_2$) and condition 20 (10% MPD, 40mM sodium cacodylate pH 7.0, 12mM spermine tetra-HCl, 80mM NaCl and 20mM $BaCl_2$). These conditions were further optimised to obtain improved crystals of DB685 and A_2T_2 DNA. Several synchrotron X-ray data collections were performed with the crystals diffracting up to 1.5Å resolution. However, after processing and phasing of the data none showed a convincing difference electron density corresponding to the ligand DB685 bound in the minor groove of the DNA (Figure 3.6.b). For ligand DB832A, crystals of good morphology were obtained from several NAM screen conditions (1, 2, 4, 12, 18, 21 and 23). Many of these DB832A crystals were tested using synchrotron X-rays and it was consistently observed that these crystals gave only diffraction up to 3.0 - 3.5Å resolution at best (Figure 3.6.d). Further optimisation did not result in a better diffraction quality and hence no further work was carried out on these crystals.

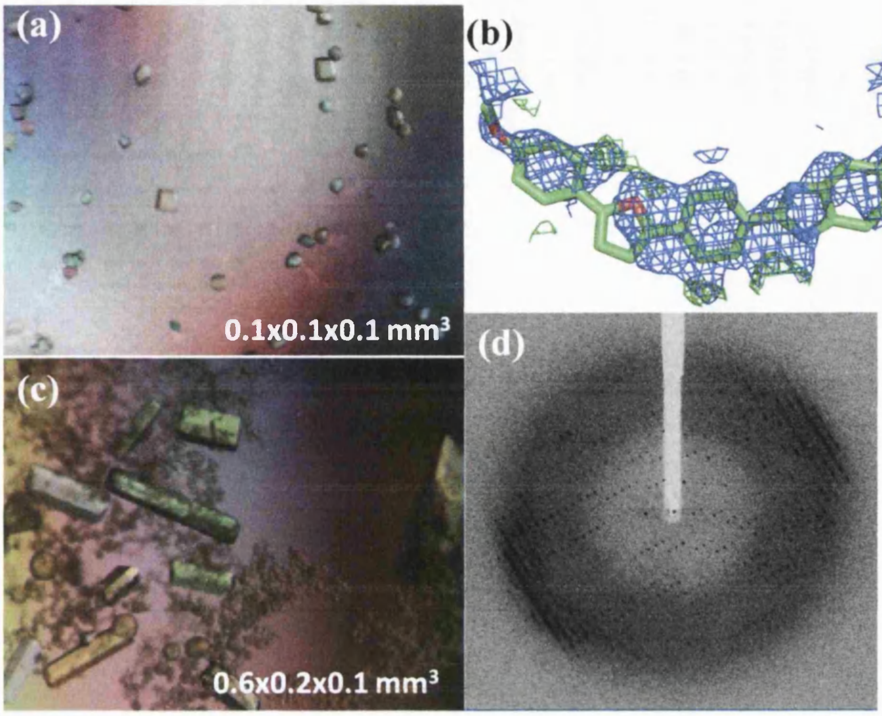


Figure 3.6. (a) Crystals of DB685 complexed with A_2T_2 DNA obtained from condition 5. (b) The $2F_o$-F_c electron density map observed in minor groove of DNA at 1.0 sigma level (c) Crystal of DB832A and A_3T_3 DNA obtained from condition 23. (d) X-ray diffraction pattern of DB832A crystal extending to 3.0Å resolution.

Another commercially available crystallisation screen, NATRIX (Hampton Research), was used for the DB75 and A_3T_3 complex (Appendix II). Crystals were obtained from condition 31 of the NATRIX screen which contained 5mM magnesium chloride hexahydrate, 50mM HEPES sodium pH 7.0 and 25% PEG monomethyl ether 550. The crystals from the NATRIX screen were only obtained after approximately 1 year, while the literature survey conditions in Table 3.1 resulted in crystals within 2-3 weeks.

3.3.2. Optimisation of initial conditions

The large crystals of ligand-DNA complexes obtained in hydrogenated conditions were hydrogen-deuterium exchanged by changing the hydrogenated MPD well solutions with

MPD well solution made in D_2O. Most of these crystals were stable with change in the well solution (Figure 3.5.b, c, d., images taken after the well solution was changed).

The exceptions to this were the crystals of DB262 with A_3T_3 DNA. These crystals were obtained from hydrogenated conditions 1 and 9 with approximately (0.6 x 0.2 x 0.1) mm^3 and (0.9 x 0.2 x 0.1) mm^3 volume respectively. These crystals were grown in sitting drop conditions. The well solutions for condition 1 and 9 were 50% and 35% MPD respectively which were then replaced with 50% and 35% MPD made in D_2O solution. The DB262 crystals developed cracks approximately 1 week after the hydrogen-deuterium exchange.

In an attempt to obtain larger crystals suitable for neutron studies, the most promising conditions found from the initial screens (Section 3.3.1) were optimised. The optimisation of the conditions was performed only at 12°C as the equipment available for maintaining the plates at this temperature was more stable. Another reason for working at 12°C was, as mentioned above, that the crystallisation conditions of Table 3.1 set in hydrogenated and then deuterated conditions provided more reproducible results than at 20°C.

Optimisation involved firstly the variation of the concentrations of precipitating agent (MPD) and DNA-ligand in a regular fashion from the original screen condition. The next step of optimisation involved varying the concentration of each additive whilst the concentration of precipitating agent in the well was kept stable.

Optimisation of berenil crystals

Crystallisation conditions 1 and 6 (Table 3.1) were optimised for berenil and A_3T_3 DNA crystals. The original condition 1 contained drop concentrations of 31.24mM $MgCl_2$, 1.04% MPD, 0.12mM annealed DNA, 0.16mM berenil and 5mM sodium cacodylate precipitated against a well solution containing 50% MPD, all in deuterated conditions. This condition provided single rod shaped crystals and also contained multiple crystals growing as clustered groups (Figure 3.7.a and 3.7.b). This condition was optimised by simultaneously varying the precipitating agent and DNA-ligand concentration. The precipitating agent concentration was varied from 25% to 55% and the DNA-ligand concentration ratio was varied from 0.1:0.2 up to 0.3:0.4. Crystals were obtained from this plate at different time intervals. Well defined crystals with good morphology were obtained within 2 weeks at a low concentration of DNA-ligand (0.12 and 0.16mM respectively) and 55% of MPD (Figure 3.7.c) whilst at the same concentration of precipitating agent MPD and higher concentration of DNA-berenil, showers of crystals were observed within 1 week (Figure 3.7.d). Within 3-5 weeks of setting the plate better crystals were observed in conditions ranging from 45% to 50% of precipitating agent concentration (Figure 3.7.e). The low concentration of MPD and DNA-ligand crystallisation conditions also produced crystals in approximately 8-10 weeks and these crystals were larger than the crystals obtained from the original condition (Figure 3.7.f).

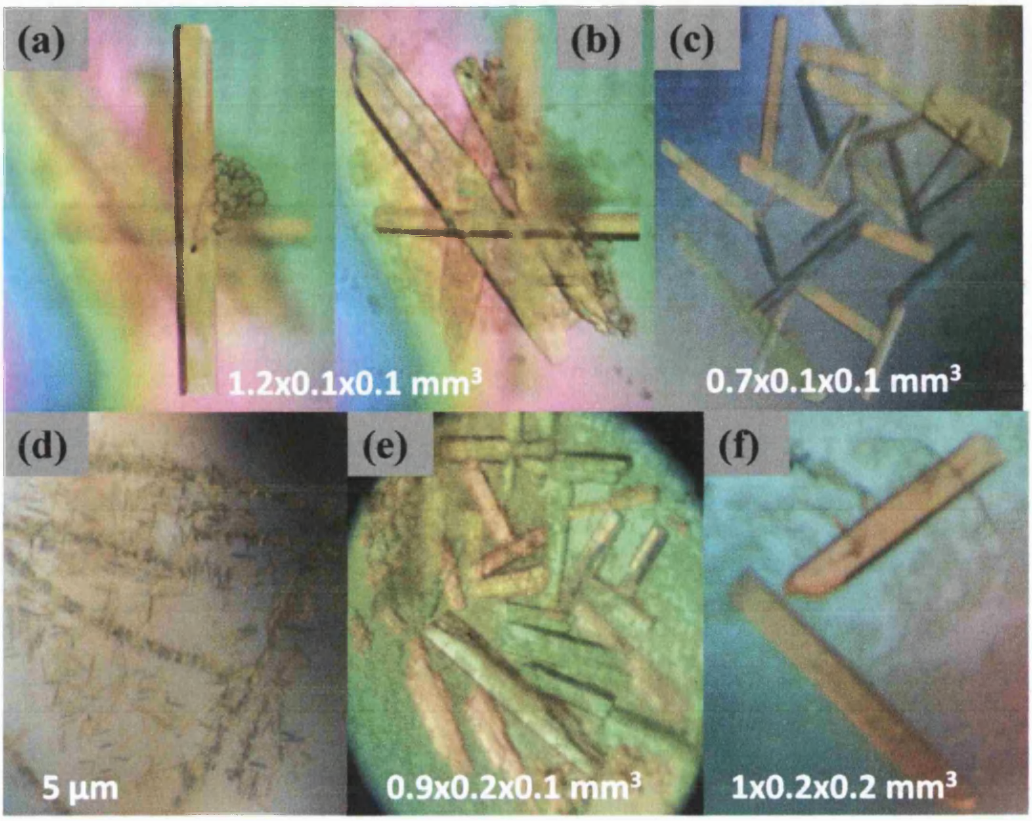


Figure 3.7. Crystals of berenil and A_3T_3 DNA obtained after optimisation of condition 1 by changing the reservoir and DNA-ligand concentration.

To further improve the crystallisation condition and to obtain good quality crystals within a good time frame, further optimisation was performed where the concentration of each additive was varied whilst keeping the concentration of the precipitating agent constant. Crystals of good quality were obtained with this optimisation; although there was little improvement in the volume of the crystals (Figure 3.8.a). The concentration of precipitating agent and the pD of the solution was also varied. The MPD concentration was varied from 40% to 55% and the pD of the solution was varied from 4.5 to 7.0. It was observed this change in pD also helped in obtaining better quality crystals with good morphology but it did not improve the volume of the crystal (Figure 3.8.b). With the change in pD good quality crystals were observed in the range of 5.5 and 7.0.

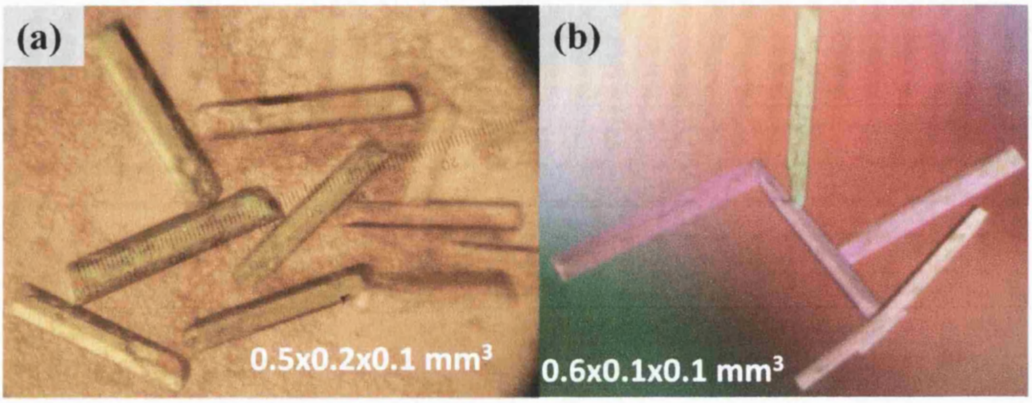


Figure 3.8. Crystals of berenil and A_3T_3 DNA obtained from (a) changing concentration of each additive and (b) changing pD and concentration of precipitating agent.

Both micro and macro seeding techniques were also tested to obtain large crystals of berenil for neutron studies. The crystals yielded by the optimisations trial by changing MPD and DNA-ligand concentration were used as the seed crystals. Pre-equilibration of the drop is an important factor to be considered for a seeding experiment. Since the exact time required for the drop equilibration was not known, the seeding experiment was performed over a period of time. The first set of seeding experiments (macro and micro seeding) was done 1 hour after the drop was set while the last seeding was done 5 days after the drop was set. For the macroseeding experiment, the crystal was first washed with cryo solution. The cryo solution contained twice the concentration of the additives present in the condition in which the seed crystal was obtained along with glycerol. This was done to remove any micro-crystalline particles present on the surface of the crystal and therefore to minimise nucleation sites and maximise the chance of a single crystal growing in the drop. The crystal was then seeded into the equilibrated drop. For the microseeding method, the crystal was taken from the early optimisation techniques and washed with the cryo solution and then the crystal was crushed using an acupuncture needle and diluted to 500X, 1000X and 2000X using 30mM sodium cacodylate pD 6.5 solution. These solutions were used as seed solutions for seeding equilibrated drops using cat whiskers.

Macroseeding after 3 days of equilibration of the drop gave better results at obtaining crystals of berenil and A_3T_3 DNA complex. However, the crystals obtained were not of sufficient volume for neutron diffraction study (Figure 3.9.a). Microseeding of the drop with seed solution diluted at 2000X provided promising result with crystals of approximately (1.1 x 0.2 x 0.1) mm^3 volume (Figure 3.9.b).

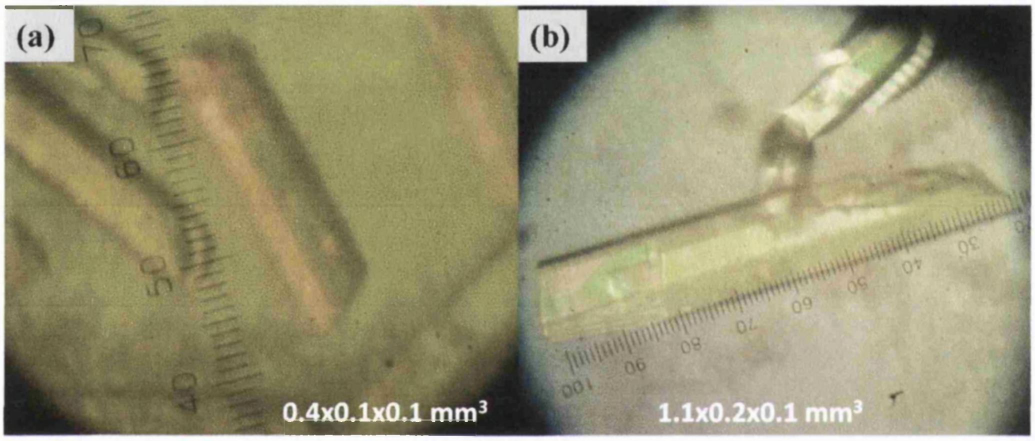


Figure 3.9. Crystal of berenil and A_3T_3 DNA obtained (a) after macroseeding and (b) after microseeding of the equilibrated drop.

Crystals of berenil complexed with A_2T_2 DNA were also obtained from condition 13 (Table 3.1), 12μl sitting drop solution containing 166.7mM $MgCl_2$, 11.25% MPD, 0.825mM annealed DNA and 1.667mM ligand equilibrated against 45% MPD as reservoir solution all in deuterated solutions. The crystal volume was approximately (0.4 x 0.3 x 0.2) mm^3 but the crystal morphology was not good and appeared as multiple crystals (Figure 3.10.a). Hence this condition was further optimised by changing the precipitating agent and DNA-ligand concentration. A better quality and slightly bigger crystal was obtained within 5 weeks from drop containing 166.7mM $MgCl_2$, 12.5%

MPD, 0.5mM DNA and 1mM berenil equilibrated against 40% MPD. The crystal volume was approximately (0.6 x 0.2 x 0.2) mm^3 and had well defined edges and good morphology (Figure 3.10.b).

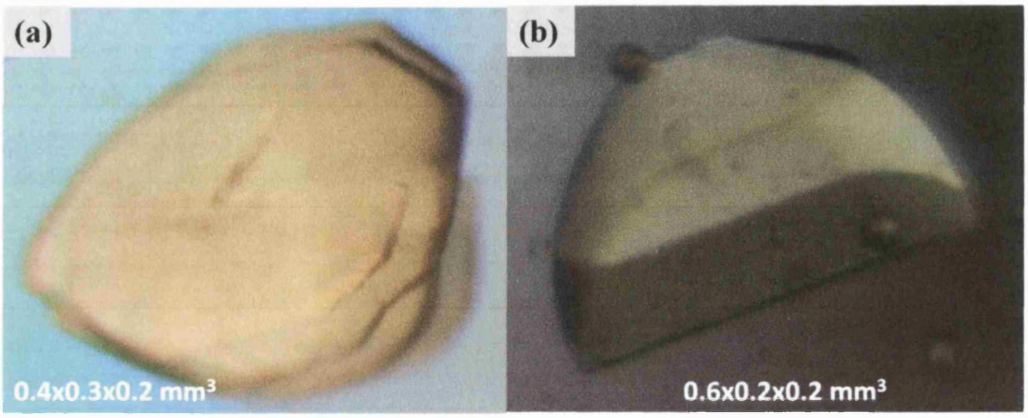


Figure 3.10. Crystal of berenil and A_2T_2 DNA obtained (a) from screening study, condition 13 and (b) by optimising condition 13.

Optimisation of DB75 crystals

The most promising crystal for neutron study obtained was of the DB75 and A_3T_3 DNA from hydrogenated condition 6 with the drop containing 28.57mM $MgCl_2$, 11.43% MPD, 0.257mM spermine, 0.857mM double stranded DNA and 0.89mM ligand equilibrated against 2ml of 40% MPD reservoir solution. This condition was reproduced in deuterated solution and this also provided crystal of good volume crystal although the crystal morphology was not good (Figure 3.11.a). To improve the quality of the crystal the optimisation of condition 6 was performed by varying the precipitating agent and DNA-ligand concentration. The optimisation resulted in several good quality crystals with volumes suitable for neutron studies (Figure 3.11.b). It was observed that changing reservoir solution from 20% to 40% MPD resulted in large crystals in 4-5 weeks. Higher concentration of MPD (50% and 55%) in the well resulted in large crystals growing as clusters growing in 1-2 weeks (Figure 3.11.c). Further optimisation of condition 6 by changing the concentrations of additives in the drop also resulted in several good crystals. The most promising crystal was obtained with drop solution containing 0% MPD while the concentrations of other additives were same as in condition 6. This drop was equilibrated against 40% MPD in the well (Figure 3.11.d). Several of the best DB75 and A_3T_3 crystals from this optimised conditions were tested for neutron diffraction (see Section 3.4).

All the condition 6 optimisation drops had a crystalline spherulite-like object present in the drop, either attached to the crystal or close to it (Figure 3.11.b and 3.11.d). Several of these objects were tested using X-ray diffraction and showed a diffraction pattern of a DNA-ligand crystal but the diffraction was poor and showed multiple crystal lattices were present.

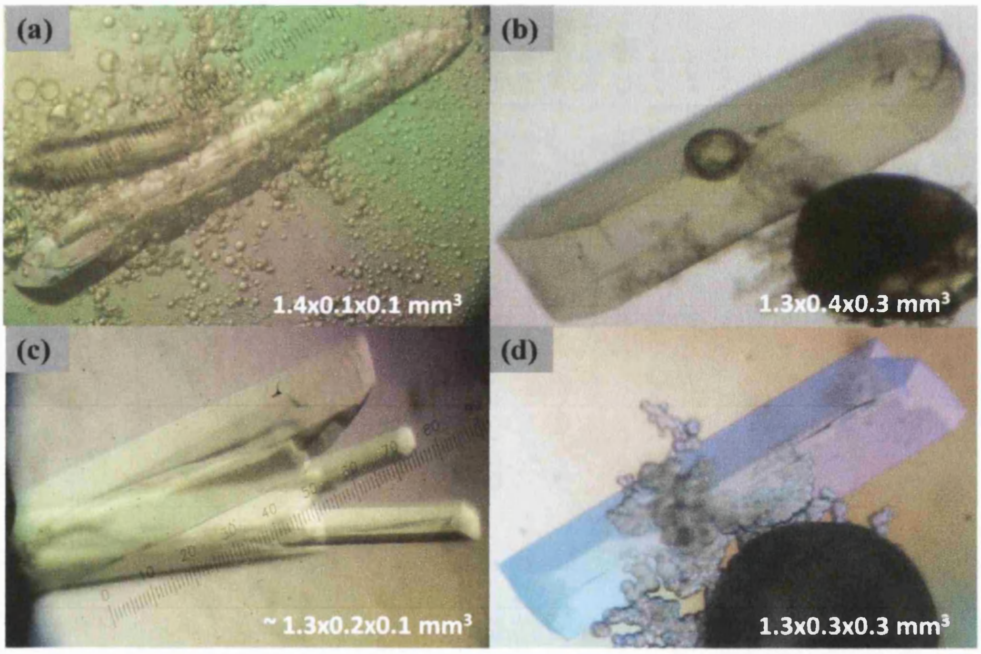


Figure 3.11. Crystals of DB75 and A_3T_3 obtained in deuterated conditions (a) from condition 6 of volume (1.4 x 0.1 x 0.1) mm^3 (b) by optimising condition 6 with low MPD concentration in well, (1.3 x 0.4 x0.3) mm^3 (c) by optimising condition 6 with high MPD concentration in well and (d) by optimising condition 6 with 0% MPD in drop solution, (1.3 x 0.3 x 0.3) mm^3.

A feeding technique was also tested with this condition 6 where the crystals in the drop were supplied with fresh drop solution containing higher concentration of DNA and ligand. It was observed that the crystals obtained earlier gradually dissolved into solution and did not re-crystallise.

Another technique attempted was where the reservoir solution was gradually changed. The crystallisation drop was first equilibrated against higher concentration of MPD, in case of condition 6, 40% MPD was first tested. As nucleation was observed the reservoir solution was changed gradually to lower concentration of MPD. This change in reservoir solution was done over several steps to reduce the rate of crystallisation with the aim to obtain large volume crystals growing over a period of time. A crystal of approximately (1.1 x 0.2 x0.1) mm^3 volume was obtained from this technique. However, the crystal volume was smaller than crystals obtained from the early optimisation trial of changing the concentration of precipitating agent and DNA-ligand (Figure 3.11).

Crystals of the DB75 and A_3T_3 DNA complex were also obtained from condition 7 (Table 3.1). The crystals were obtained from both hydrogenated and deuterated conditions and were of similar volume of approximately (0.9 x 0.05x 0.05) mm^3 (Figure 3.12.a). The drop solution for this condition contained 18mM $MgCl_2$, 7.5% MPD, 1mM spermine, 0.4mM double stranded DNA and 0.6mM DB75 equilibrated against 40% MPD reservoir solution. As a starting optimisation trial the precipitating agent MPD was replaced by PEG. Different PEGs were used for this trial: PEG400, PEG1000, PEG2000, PEG3350, PEG6000 and PEG8000. It was observed that replacing MPD with PEG did result in crystals but of different morphology and much smaller at approximately (20 x 30 x 20) μm^3 (Figure 3.12.b). It was also observed that the low molecular weight PEGs (PEG400, PEG1000 and PEG2000) resulted in better (larger) crystals compared to the high molecular weight PEGs. The crystals obtained using the high molecular weight PEGs were essentially microcrystals with no defined crystal morphology or shape. Some of the crystals obtained from low molecular weight PEGs were tested by synchrotron X-ray diffraction. It was observed that these crystals diffracted only up to 2.0Å resolution while the crystals obtained using MPD diffracted to higher resolution (typically 1.2 to 1.7Å).

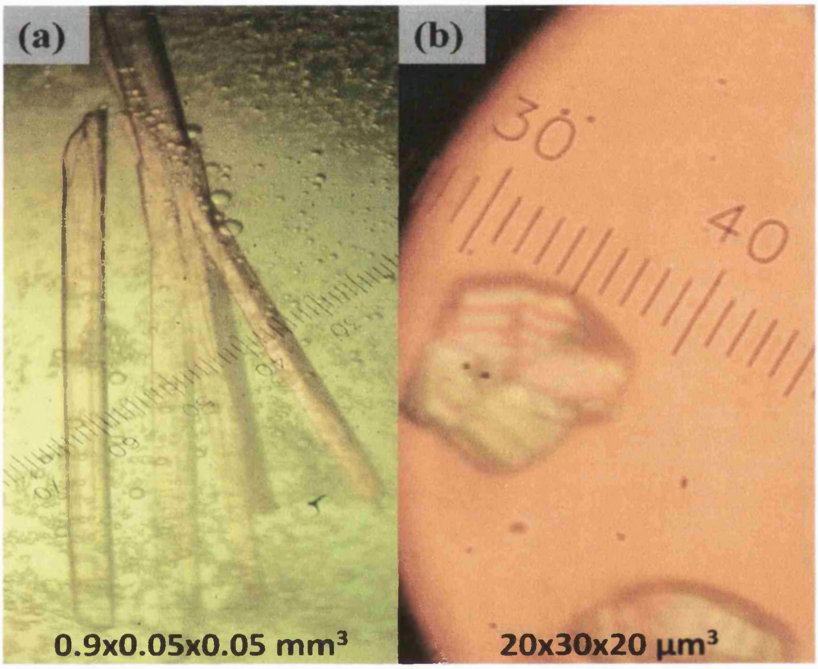


Figure 3.12. Crystals of DB75 and A_3T_3 DNA obtained from condition 7 using (a) MPD and (b) PEG400 as precipitating agent.

Optimisation of DB985

The first large crystal of DB985 was, surprisingly, obtained from un-purified DNA sample. The crystal volumes were approximately (1.0 x 0.3 x 0.3) mm^3 obtained from condition 9 and (1.4 x 0.2 x 0.3) mm^3 obtained from condition 13. These conditions were reproduced in deuterated conditions and further optimised by changing the precipitating agent and DNA-ligand concentration, although these crystallisation trials were performed using the purified DNA samples. There was a significant difference in the morphology of the crystals obtained before and after DNA purification (Figure 3.13).

The crystals obtained by using the purified DNA samples were tested by X-ray diffraction. The crystals diffracted X-rays to approximately 1.2Å resolution and a complete data set was collected. The structure obtained from the refinement of this crystal was judged to be highly similar to the X-ray structure obtained from the crystal of DB985 A_3T_3 with unpurified DNA which is described in Chapter 4. The two structures were superposed using COOT (Emsley & Cowtan. 2004). The structures have an RMS deviation of 0.3Å and from the electron density maps it is evident that the water network and ligand position are the same as the completely refined DB985-A_3T_3 (not purified DNA sample) crystal structure. This indicated that both crystals of DB985 obtained from using purified and not purified DNA samples were similar with insignificant structural changes, although the crystal morphology was very different.

The optimisation of condition 13 resulted in large oval-shaped crystals. The largest crystal obtained was of approximately (0.9 x 0.4 x 0.4) mm^3 volume which was tested for neutron diffraction (see Section 3.4 below).

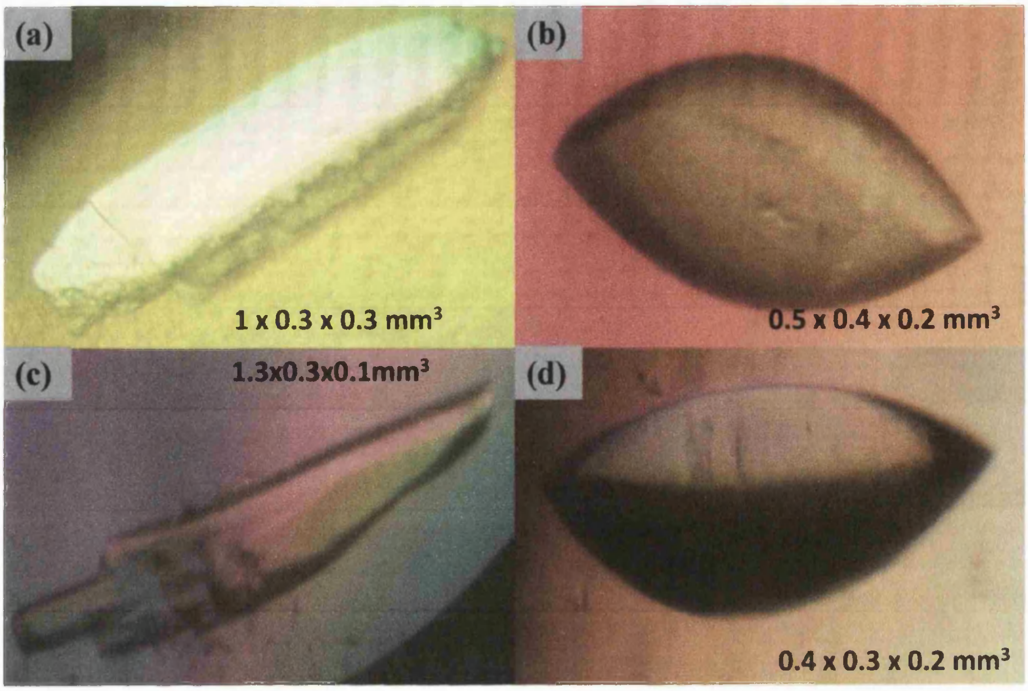


Figure 3.13. Crystals of DB985 and A_3T_3 DNA obtained from (a) condition 9 using DNA which was not purified (b) optimised condition 9 using purified DNA (c) condition 13 using DNA before purification and (d) optimised condition 13 using purified DNA sample.

3.3.3. Other crystallisation methods

The vapour diffusion technique of crystallisation was the most widely used method in the screening and optimisation as mentioned in the earlier sections. However, in an attempt to obtain large crystals suitable for neutron diffraction, other crystallisation techniques were also attempted including crystallisation under oil, batch crystallisation and crystallisation in capillaries. These attempts are summarised below.

Slowing down equilibration: heavy oils

A crystallisation trial for the berenil and A_3T_3 DNA complex was performed under oil. In this technique heavy oil was layered over the reservoir solution. This was done to reduce the equilibration rate of the reservoir and drop solution. In these tests, different concentrations of heavy oils, (a) 100% paraffin oil, (b) 100% silicone oil and (c) mixture of 50% paraffin and 50% silicone oil, were tested. It was observed that in the presence of these heavy oils over the reservoir, the drop solution dried out completely with heavy precipitation within 24 hours (Figure 3.14.a). One of the crystallisation conditions with 100% paraffin oil over the reservoir solution was microseeded. This drop solution resulted in crystals although they had poor quality (Figure 3.14.b).

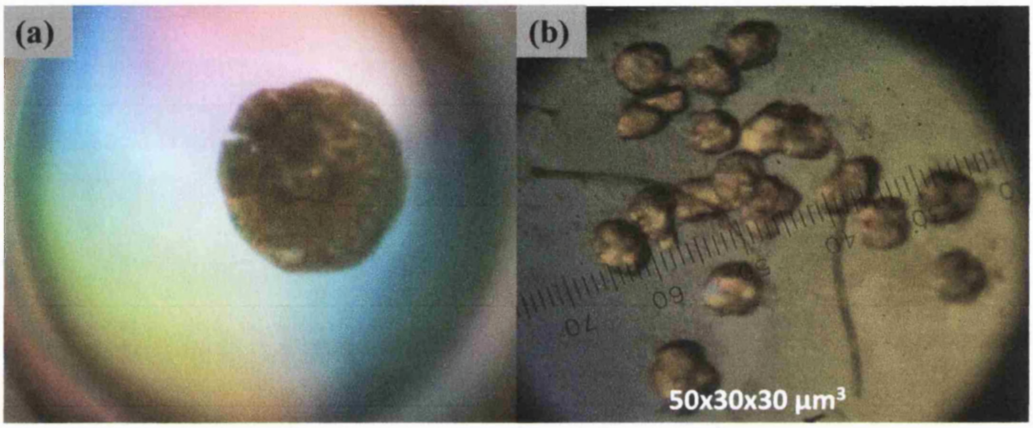


Figure 3.14. (a) Crystallisation drop after equilibrations against 55% MPD reservoir solution layered with 60% paraffin oil. (b) Crystals obtained from microseeding a crystallisation drop with reservoir solution containing MPD solution layered with 100% paraffin oil.

Growing larger crystals: batch crystallisation

The crystallisation conditions which provided crystals of suitable volume for neutron diffraction studies (Section 3.3.2) were adapted for batch crystallisation to obtain bigger crystals. The optimised crystallisation condition for DB75 and A_3T_3 DNA were used for batch crystallisation trials. For the batch trial the quantity of DNA in the drop was increased to approximately 0.5mg compared to the original condition containing only approximately 0.15mg of DNA. All the other additives and ligand concentrations were adjusted accordingly to compensate for the increased DNA concentration. The crystallisation drop was prepared such that the concentration in the drop would be similar to a pre-equilibrated drop. Four trial batch conditions were set in an attempt to be in the range of equilibrated drop. The batch drops were set with no reservoir solution. However, for 3 of the batch trial conditions it was observed that on addition of ligand to the batch solution containing high concentrations of DNA led to immediate precipitation. The clear batch solution was set for crystallisation and it was observed that this solution precipitated after couple of days. The batch crystallisation trials on DB75 and A_3T_3 DNA did not provide any promising results and, given the large amounts of material required, were not further pursued.

Growing larger crystals: scale-up of vapour diffusion

Since the batch crystallisation did not provide any promising results, the normal vapour diffusion technique which resulted in neutron size crystals of DB75 and A_3T_3 DNA was scaled up.

Condition 6 (Table 3.1) was used for the scale-up vapour diffusion trial. From the earlier optimisation study of this condition (Section 3.3.2) it was evident that better crystals were obtained with 20% to 40% of MPD in the reservoir. However, the conditions with lowest concentrations of reservoir solution required 4-5 weeks to crystallise. Given the large volumes of drop solution used for the scale-up vapour diffusion trial, the crystallisation time required would be long. Therefore, the scale-up vapour diffusion trials were set up by varying the reservoir solution only between 40% and 45% MPD.

Due to the volumes involved, the scale-up trial could not be performed in standard vapour diffusion trays but was performed by using a cylindrical plastic container with lid and a watch glass mounted in this container. This set up resembled the normal sitting drop vapour diffusion technique. Twelve scale-up vapour diffusion containers were set-up, 6 of which had 40% MPD in the reservoir and the remainder had 45% MPD. The drop solution contained varying concentrations of MPD. With this experiment it was observed that 5 drops precipitated (these drops contained lower concentration of MPD in drop) while the remaining drop solutions dried out over a period of 1 month without any crystal being obtained. This could have been due to the insufficient sealing of the plastic container used for this trial. The scale-up vapour diffusion trial did not provide any promising results.

Growing larger crystals: in-capillary growth

The DB75 and A_3T_3 DNA complex was also attempted to be crystallised in quartz capillaries. The positive crystallisation condition 6 was used for this trial. The drop solution used for scale-up vapour diffusion was loaded into a 1mm diameter capillary. The capillary was sealed with wax and a hole of made using an acupuncture needle (0.16mm diameter) in the seal to allow slow evaporation.

Some of the crystallisation solution which had higher concentration of MPD dried out within a few weeks of the capillary being set - possibly because the hole made in the seal was too large. This trial did not provide any promising crystals. However, one of the conditions which included 0% MPD in the crystallisation solution resulted in a small needle shaped crystal of approximately (50 x 10 x 10) μm^3 volume. This small crystal was only obtained after approximately 3 months of setting up the condition and hence was not further optimised.

3.4. Neutron diffraction test

The large crystals obtained were tested on LADI-III instrument at Institut Laue Langevin (ILL), Grenoble. LADI-III is a Laue-based neutron diffraction beamline and the neutron wavelengths used for data collection were from 3.0 – 4.2Å.

Crystal 1

The first crystal tested on LADI-III was of the DB75 and A_3T_3 DNA complex. This crystal was obtained by optimising condition 6. The DNA and ligand concentration in the drop was 1.5mM and the drop was equilibrated against 45% MPD in the well. The crystal was mounted in a 1.5mm wide capillary with drop solution present on both sides of the crystal (Figure 3.15.a). The crystal was dried completely using paper wick and is shown in Figure 3.15.a.

A single exposure of 18 hours was made (Figure 3.15.b). The crystal diffracted to beyond 3Å resolution and was indexed in orthorhombic $P2_12_12_1$ space group with unit cell dimensions of a = 23.97Å, b = 39.45Å, c = 62.87Å and α, β, γ = 90°. The diffraction pattern, however, suggested that the crystal was not single and hence was not further collected.

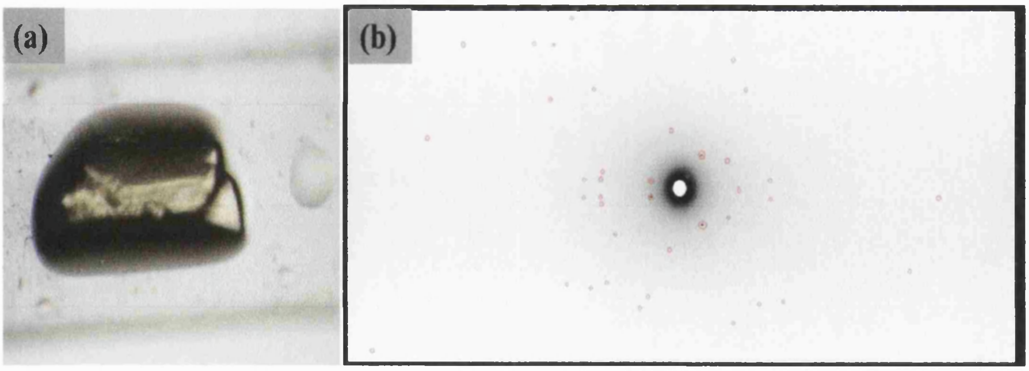


Figure 3.15. (a) Crystal of DB75 and A_3T_3 DNA mounted in capillary for neutron data collection (b) Diffraction pattern of this crystal obtained from LADI-III.

Crystals 2, 3 and 4

Following the first neutron test the crystallisation conditions were further optimised and developed to obtain more crystals for neutron data collection (Section 3.3.2). Three different crystals were tested on LADI-III to determine the best diffracting crystal to be collected. The first crystal tested was of the DB75 and A_3T_3 DNA complex. This crystal was also obtained from optimised condition 6 where the drop contained 0% MPD with all other additives at the same concentration as condition 6 and the drop was equilibrated against 40% MPD. The crystal volume was approximately (1.3 x 0.3 x 0.3) mm^3 (Figure 3.16.a). The crystal was mounted in a capillary and two images were collected by exposing the crystal for 14 hours and 16 hours. This crystal diffracted to 4Å resolution (Figure 3.16.a).

The next crystal tested was of the DB985 and A_3T_3 DNA complex. The crystal was obtained by optimising condition 13 with 166.7mM $MgCl_2$, 11.25% MPD, 1.2mM annealed DNA and 1.44mM DB985 in the drop equilibrated against 50% MPD. The crystal volume was approximately (0.9 x 0.4 x 0.4) mm^3. Two images were collected by exposing the crystal to neutrons for 3 hours and 9 hours. This crystal had poor diffraction extending to 4.0Å resolution. The final crystal tested was a crystal of the DB75 and A_3T_3 DNA. This crystal was obtained from a drop solution containing 28.57mM $MgCl_2$, 11.43% MPD, 0.257mM spermine, 1.2mM DNA and 1.2mM DB75. The drop was equilibrated against 25% MPD solution in the well (Figure 3.16.b). The crystal was exposed for 2 hours and 14 hours. This crystal diffracted beyond 3.5Å resolution and more spots could be observed compared to the other crystals (Figure 3.16.b).

This last crystal was kept aside after the test and would have been used for neutron data collection. However, the neutron beamline became unavailable unexpectedly which precluded the full data collection within the time scale of this thesis work.

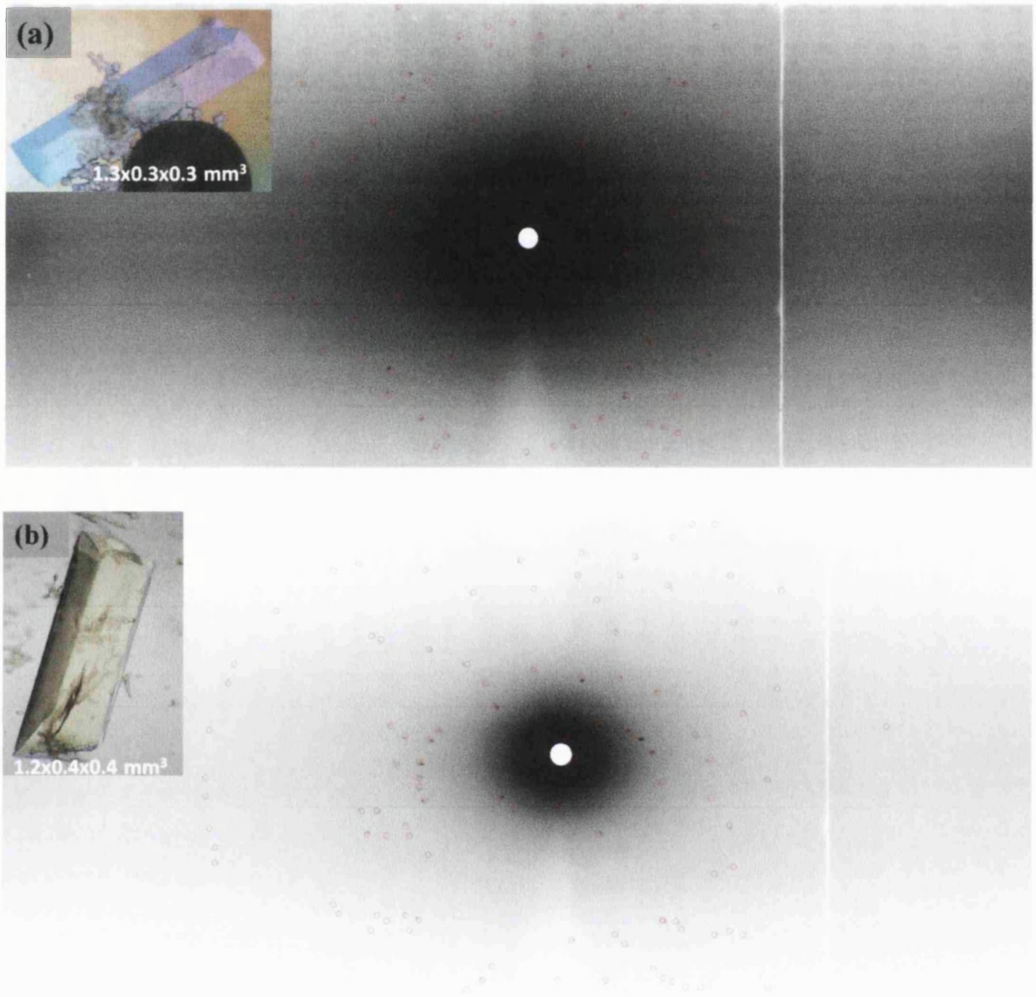


Figure 3.16. Crystals of DB75 and A_3T_3 DNA test on LADI-III and their neutron diffraction images.

3.5. Discussion

Several different MGBLs were crystallised in this study with two DNA sequences, d(CGCGAATTCGCG)$_2$ and d(CGCAAATTTGCG)$_2$. The crystallisation trials performed resulted in high-resolution X-ray crystal structures of MGBLs and DNA complexes which are discussed in detail in following chapters. Crystals of some MGBLs DB685 and DB832A were obtained but their X-ray crystal structures could not be determined due to poor electron density in minor groove for DB685 crystal and weak diffraction of DB832A crystals. The crystallisation conditions of these ligands needs to be further optimised to obtain their crystal structure.

The first trials of crystallisation were performed in hydrogenated conditions. For neutron studies these conditions were reproduced and further optimised in deuterated solutions. A wide range of crystallisation experiments were attempted, but it was observed that the sitting drop vapour diffusion technique provided the most suitable crystals for neutron study. Crystallisation conditions 1, 6, 7, 9 and 13 (Table 3.1) resulted in the best crystals for neutron study. The MGBLs berenil, DB75, DB921 and DB985 also resulted in better crystals for neutron studies.

Crystals of DB75 and DB985 in complex with A_3T_3 DNA produced under deuterated conditions were tested for neutron diffraction. These crystals provided good neutron diffraction result but could not be collected due to unexpected beamline unavailability. The crystals of MGBLs and DNA complex are available for future neutron study.

Chapter 4

X-ray crystallographic studies of linear ligands

This chapter describes detailed crystallographic studies of the interaction of two linear ligands, DB985 and DB921, with two dodecamer sequences d(CGCGAATTCGCG)$_2$ (A_2T_2) and d(CGCAAATTTGCG)$_2$ (A_3T_3). The linear ligands have a benzimidazole and two phenyl rings with charged amidine groups at the termini and belong to the *out-of-shape* class of DNA minor groove binding ligands (Miao *et al.* 2005), lacking the classical iso-helical crescent shape that matches the curvature of the minor groove of DNA.

The focus of the study is to determine the structures at sufficiently high resolution that would enable sufficiently reliable structures to be obtained, thus providing detail of the DNA-ligand-water interactions. For the crystallisation study, HPLC purified oligonucleotides were ordered from Eurogentec Ltd. or IBA BioTAGnology and were further purified (as described in Chapter 3) by anion-exchange chromatography using a Pharmacia monoQ HR 10/10 column on a BioRad FPLC workstation followed by dialysis and desalting on a bench top D-salt^{TM} polyacrylamide desalting column.

4.1. X-ray crystal structure of DB985 in complex with DNA

The crystal structure of DB985 in complex with A_2T_2 and A_3T_3 DNA sequences was obtained. The corresponding DNA single strands were annealed as a 5mM stock solution of 30mM sodium cacodylate buffer at pH 6.5 by heating on a water bath at

85°C for 15-20 minutes and then left to cool gradually overnight to room temperature. Further dilutions of the DNA and ligand were made using 30mM sodium cacodylate buffer at pH 6.5.

4.1.1. Materials and methods

4.1.1.1. Crystallisation

The crystals of DB985-A_2T_2 complex were obtained from condition 1 shown in Table 3.1. A 24µl sitting drop was set up containing 15µl of 50mM $MgCl_2$, 2µl of 1.5mM double-stranded DNA, 4µl of 1mM DB985 compound, 1µl 25% v/v ((±)-2-methyl-2,4-pentanediol (MPD) and 2µl of 60mM sodium cacodylate buffer at pH 6.5. The drop was equilibrated against a reservoir of 500µl of 50% v/v MPD solution at 285K (Figure 4.1a).

The DB985-A_3T_3 crystals were obtained from condition 6 (Table 3.1). The sitting drop solution of 14µl contained 2µl of 200mM $MgCl_2$, 4µl of 3mM double-stranded DNA, 2.5µl of 5mM DB985 compound, 4µl 40% v/v MPD and 1.5µl of 2.4mM spermine tetrahydrochloride. The drop was equilibrated against a reservoir of 2000µl of 40% v/v MPD solution at 285K (Figure 4.1b).

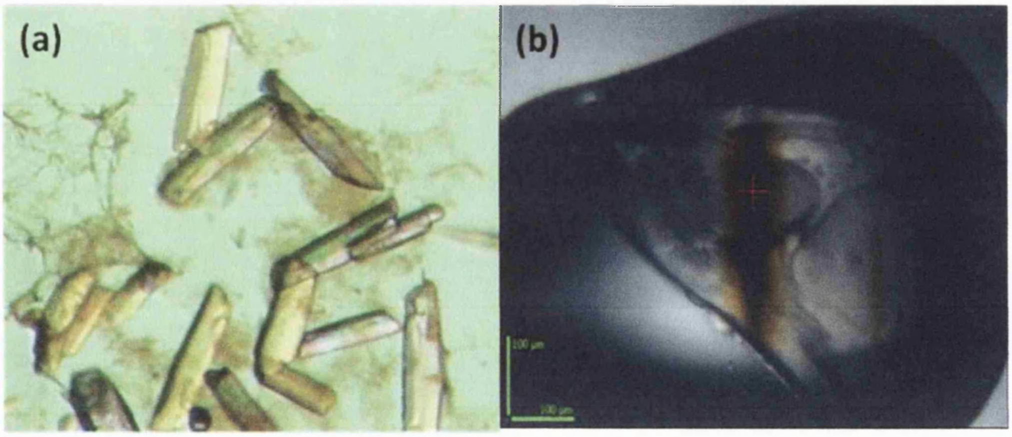


Figure 4.1. Crystals of DB985 with (a) A_2T_2, crystal size (0.4 x 0.1 x 0.1) mm^3 and (b) A_3T_3 crystal size (0.6 x 0.4 x 0.3) mm^3.

4.1.1.2. Data collection

A DB985-A_2T_2 crystal was flash-cooled in liquid nitrogen and a complete X-ray diffraction data set were collected to 1.25Å resolution on the ID14-4 beamline (McCarthy *et al.* 2009) using a Q315r ADSC X-ray detector at the European Synchrotron Radiation Facility (ESRF, Grenoble, France) using X-rays of wavelength 0.98Å. The crystal-to-detector distance was set at 140.5mm. 180 images were collected with an oscillation range of 1.0° and an exposure time of 0.2 seconds per image (Figure 4.2.a). Reflections were integrated with the program MOSFLM (Leslie. 1999) and scaled, merged and converted to structure factors using SCALA (Evans. 2006) and CTRUNCATE (Stein & Ballard. 2009) from the CCP4 suite (Collaborative Computational Project Number 4. 1994).

The DB985-A_3T_3 crystal was cryo-protected using 100% paraffin oil and cryo-cooled in liquid nitrogen before exposing it to the X-ray beam. Diffraction data were collected on the micro-focus beamline ID23-2 at ESRF using X-rays of 0.87Å wavelength and a MAR225 detector (Flot *et al.* 2010). Two sets of data were collected on the same sample to avoid the consequences of detector saturation at low resolution. 180 images were collected for the low resolution pass (1.8Å resolution) and the detector distance was set at 210.3mm with oscillation range of 1.0° and an exposure time of 0.3 seconds per image (Figure 4.2.b). A further 180 images were collected for a high resolution pass (1.2Å resolution) with the detector distance set at 121.8mm and an oscillation range of 1.0° and exposure time of 1.0 second per image (Figure 4.2.c). The two resulting sets of data were combined using POINTLESS (Evans. 2006) before scaling using SCALA (Evans. 2006) and converting to structure factors using CTRUNCATE (Stein & Ballard. 2009). Data collection statistics for both structures are shown in Table 4.1.

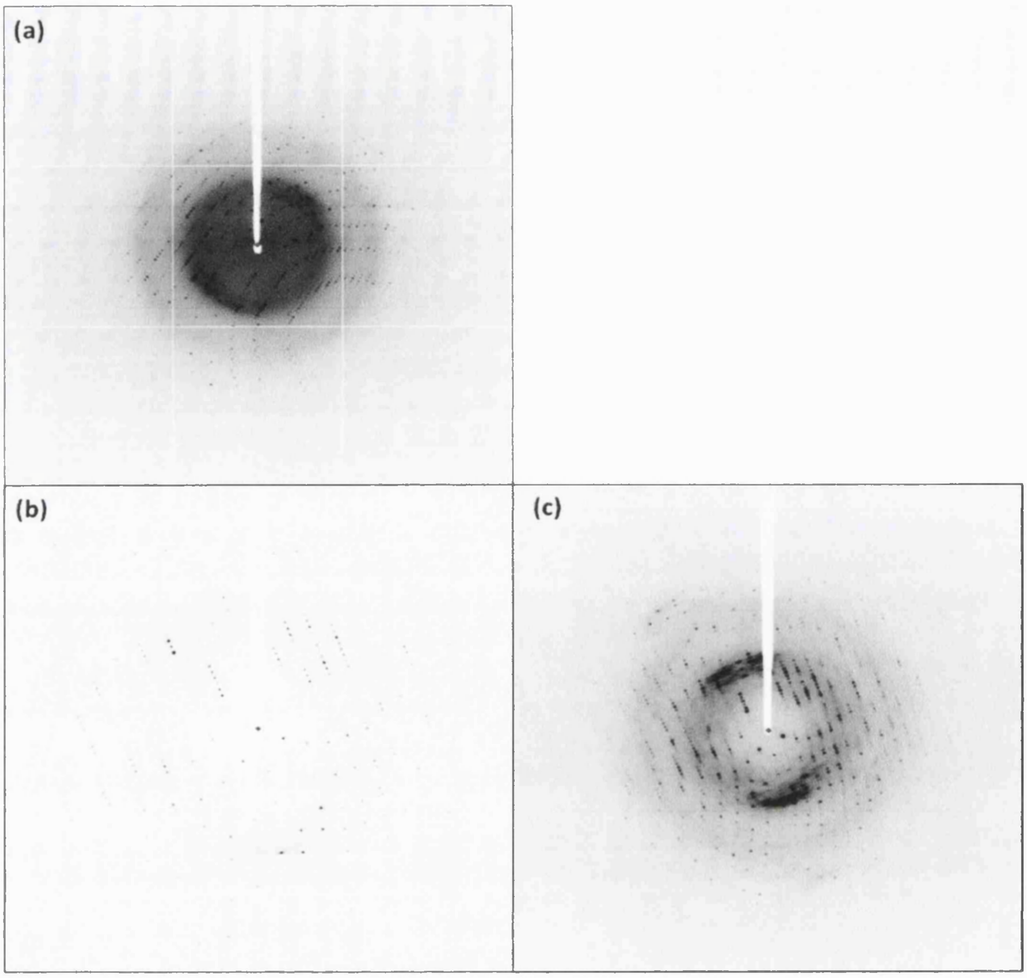


Figure 4.2. (a) Diffraction pattern of DB985-A_2T_2 crystal as collected on the ID 14-4 beamline. (b) Low resolution pass (resolution extending up to 1.8Å) and (c) high resolution pass (extending beyond 1.0Å resolution) diffraction images of DB985-A_3T_3 crystal collected on the ID 23-2 beamline.

4.1.1.3. Structure solution and refinement

Phasing information for both structures were obtained by the molecular replacement method using the PHASER program from the CCP4 suite (McCoy *et al.* 2007). The native DNA sequences of A_2T_2 [PDB code 1FQ2] (Sines *et al.* 2000) and A_3T_3 [PDB code 1S2R] (Woods *et al.* 2004) were used as molecular models for structure solution.

The models were stripped of all water molecules and ions prior to molecular replacement.

The refinement was started with a cycle of rigid body optimisation followed by restrained refinement using REFMAC5 (Murshudov *et al.* 1997). After the initial rigid-body refinement the R-factor and R-free for DB985-A_2T_2 were 50.9% and 51.7% and for DB985-A_3T_3 were 49.9% and 49.4%, respectively. The resulting sigma weighted $2F_o$-F_c and F_o-F_c electron density maps COOT (Emsley & Cowtan. 2004) were used to visualise and model ligand binding in the DNA minor groove. Continuous electron density corresponding to the ligand was observed in the minor groove of both DNA sequences (Figure 4.3.a). Water molecules were added using the automated COOT water placement procedure and also placed manually by visual inspection of the $2F_o$-F_c map and hydrogen bonding distances with neighbouring atoms. Where water molecules were less than 2.0Å apart, each water molecule was assigned a partial occupancy of 0.5. Hydrated magnesium ions were placed manually on the basis of their octahedral geometry. The R-factor and R-free at this stage for the structures were 20.9% and 24.5% for DB985-A_2T_2 and 23.6% and 26.3% for DB985-A_3T_3, respectively.

Six parameters per atom are required to model atomic anisotropic displacement parameters for a macromolecular structure (Winn *et al.* 2001). Due to an inadequate data-to-parameter ratio, anisotropic B-factors were introduced only for the DNA atoms during refinement. This resulted in a further decrease in R-factor and R-free to 17.2% and 22.8% for DB985-A_2T_2 and to 18.8% and 22.4% for DB985-A_3T_3, respectively. Idealised model coordinates of DB985 were calculated using the Dundee PRODRG server (van Aalten *et al.* 1996). At this stage of refinement, the complete ligand was built into the $2F_o$-F_c and F_o-F_c electron density (Figure 4.3.b).

The final R-factor and R-free values for the DB985-A_2T_2 structure are 15.5% and 20.7% and for the DB985-A_3T_3 structure are 15.3% and 20.7% respectively. Final coordinates were validated using the WHATIF web interface (Vriend. 1990) and analysed using w3DNA (Lu & Olson. 2003). The crystallographic refinement statistics for these structures are listed in Table 4.1.

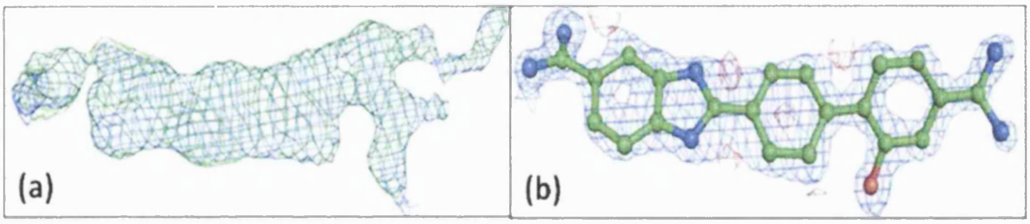


Figure 4.3. (a) The minor groove continuous $2F_o$-F_c and F_o-F_c electron density corresponding to DB985, after a first cycle of restrained refinement, shown in blue and green respectively. (b) The final $2F_o$-F_c and F_o-F_c electron density map calculated for the DB985 bound in the minor groove of the A_2T_2 DNA shown in blue and red respectively. The $2F_o$-F_c and F_o-F_c maps are contoured at 1.0 σ level and 3.0 σ level, respectively.

DNA sequence and Ligand	d(CGCGAATTCGCG)$_2$ and DB985	d(CGCAAATTTGCG)$_2$ and DB985
Data Collection		
Space group	$P2_12_12_1$	$P2_12_12_1$
Unit cell dimensions (Å,°)	a=24.03, b=39.91, c=65.78, α=β=γ=90.00	a=24.29, b=40.29, c=64.68, α=β=γ=90.00
Resolution (Å)	39.90-1.25 (1.32-1.25)	40.29-1.20 (1.26-1.20)
Wavelength (Å)	0.98	0.87
No. measured reflections	117062 (17186)	200871 (20567)
No. unique reflections	18006 (2600)	20606 (2967)
Completeness (%)	99.00 (100.00)	100.00 (100.00)
I/σ(I) for the data set	17.30 (3.10)	16.70 (3.70)
Data redundancy	6.50 (6.60)	9.70 (6.90)
Rmerge (%)	5.30 (54.70)	7.2 (41.90)
Refinement		
Resolution range	14.76-1.25 (1.28-1.25)	14.40-1.20 (1.23-1.20)
No. reflections	17027 (1230)	19484 (1413)
Rfactor (%)	15.48 (21.20)	15.29 (20.80)
Rfree (%)	20.70 (27.40)	20.31 (25.80)
RMS deviation		
Bond length (Å)	0.020	0.025
Bond angle deviation (Å)	2.57	2.72
Number of atoms		
DNA	486	486
DB985 ligand	28	28
Magnesium	3	3
Water (full/half occupancy)	172 / 19	170 / 11
Mean temperature factor of atoms (Å^2)		
DNA	15.78	17.29
DB985 ligand	21.04	22.89
Magnesium	14.20	15.92
Waters	30.05	31.69

Table 4.1. X-ray data collection and refinement statistics are summarised for DB985 structures with A_2T_2 and A_3T_3 DNA. The high resolution shell statistics are shown in parentheses.

4.1.2. Results and discussion

The asymmetric unit of both crystal structures contains a duplex DNA (chain A and chain B), one molecule of DB985 and three magnesium ions. In the DB985-A_2T_2 structure, 172 full-occupancy and 19 half-occupancy water molecules were identified and included in the final model. The DB985-A_3T_3 structure contains 170 full-occupancy and 11 half-occupancy water molecules. For both structures, one molecule of DB985 ligand is bound within the central AATT or AAATTT bases pair region in the minor groove. In the A_2T_2 structure one of the magnesium ions has full occupancy while the other two have been modelled with 0.25 occupancy. In the A_3T_3 structure, one of the magnesium ions was modelled with full occupancy, the second at 0.50 occupancy and the third at 0.25 occupancy.

4.1.2.1. DB985-groove contacts in d(CGCGAATTCGCG)$_2$

Direct drug-DNA contacts: The benzimidazole amidine group nitrogen (N1) (Figure 4.4.a and 4.4.b) forms a direct hydrogen bond with the O2 of thymine 8 (3.1Å) and O4' of cytosine 9 (3.3Å) of chain A. The inner facing nitrogen atom (N5 in Figure 4.4.a, 4.4.b) of the benzimidazole ring forms a pair of bifurcated hydrogen bonds with the O2 of thymine 7 (3.1Å) on chain A and with the O2 of thymine 19 (2.9Å) of chain B.

Water-mediated interactions: Within the minor groove of the DNA there is a complex water network mediating contact between ligand and DNA. The N1 on the amidine group of the drug also forms a hydrogen bond with water 88 (2.7Å long) which is responsible for a further hydrogen bond with O2 of cytosine 9 (2.9Å) on chain A. This water 88 (Figure 4.4.b) is part of an important network in the minor groove giving rise to further water mediated hydrogen bonds with O2 of cytosine 9, O4' of guanine 10 on chain A, and N3 of adenine 17 and O4' of adenine 18 on chain B. The positively charged N2 atom of the amidine group of DB985 also forms a hydrogen bond with a water molecule (2.0Å) which also contacts with OP1 of thymine 20 of chain B via another water molecule. The hydroxyl moiety of DB985 is oriented away from the minor groove and forms a hydrogen bond with OP1 of the thymine 7 phosphate via two-water bridges (Figure 4.4.c). The nitrogen atoms N5 and N6 of the amidine group attached to the phenyl ring at the other end of the molecule do not form any direct H-bond with the DNA. However, they participate in a dense water network which further hydrogen bonds with the bases of both chains (Figure 4.4.b, 4.4.c and 4.11.a). Water 22 is a conserved water molecule in the minor groove of DNA, observed in DB985-A_3T_3 and native A_2T_2 and DB921-A_2T_2 and DB921-A_3T_3 crystal structures (Figure 4.11). The ligand binding site extends over four base pairs and forms five direct hydrogen bonds with the DNA bases along with several water-mediated hydrogen bonds.

***Outer interactions*:** There is a ribbon of water molecules that connects and arches over the two strands of the DNA. Water 98 is close to phosphate oxygen OP1 of cytosine 9 of chain A, which further bridges with two waters; water 171 and water 175 and finally contacts phosphate oxygen OP1 of thymine 20 in chain B.

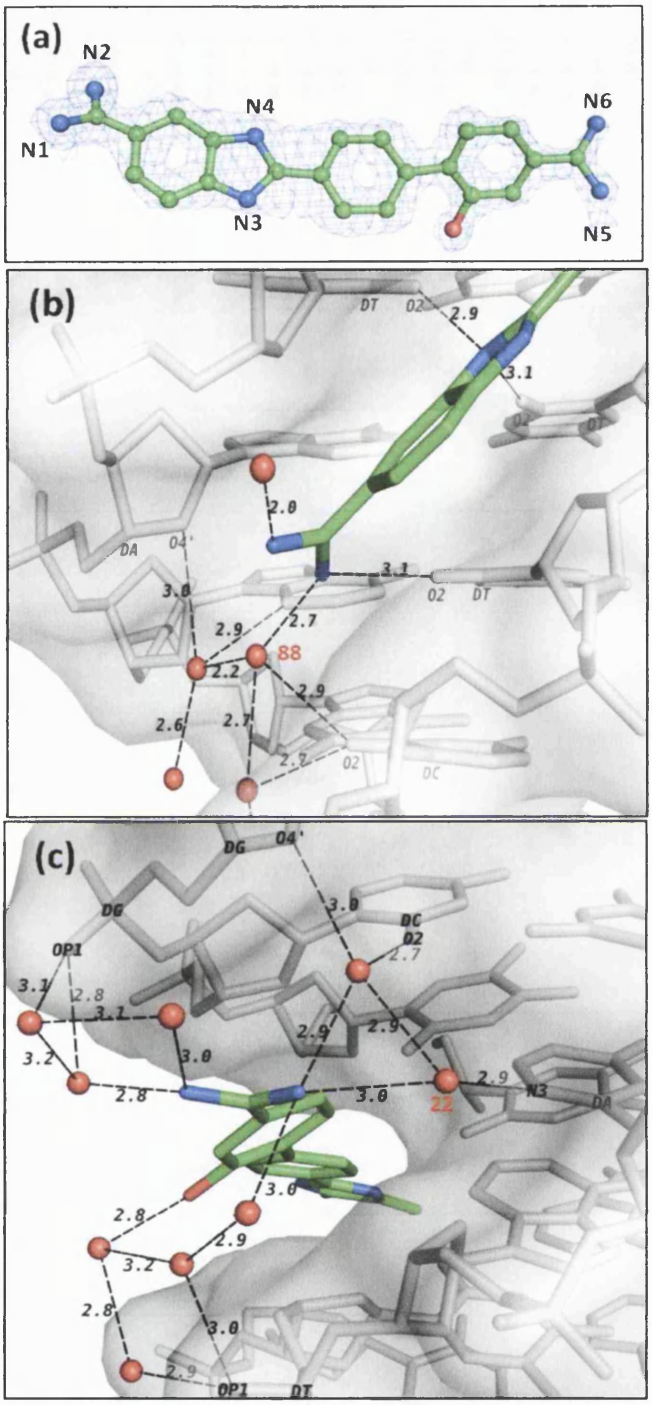


Figure 4.4. (a) The final $2F_o$-F_c electron density map calculated for the DB985 bound in the minor groove of the A_2T_2 DNA at 1.0σ level contour. (b) A side view of the DNA minor groove showing the benzimidazole-amidine end of the ligand. (c) A top view, looking into the minor groove at the phenyl-amidine end of the ligand. DB985 is shown as green stick model. Water molecules in the groove are shown as red spheres and DNA is shown as grey surface.

4.1.2.2. DB985-groove contacts in d(CGCAAATTTGCG)$_2$

Direct drug-DNA contact: The inner-facing N1 of the amidine group attached to the benzimidazole ring of the ligand forms a direct hydrogen bond to O2 of thymine 8 (3.2Å) and O4' of thymine 9 (3.2Å) of chain A. The inner facing N3 of the benzimidazole ring forms a bifurcated hydrogen bond with O2 of thymine 7 (2.9Å) of chain A and O2 of thymine 19 (2.9Å) of chain B. These interactions are similar to the ligand contacts as seen in DB985-A_2T_2 structure.

Water-mediated interactions: The N1 of the ligand forms hydrogen bond with a 0.5 occupancy water molecule 59A, which further forms hydrogen bond with O2 of thymine 9 (2.9Å) similar to the DB985 interaction with AATT. Water 59A also forms hydrogen bonds with O4' of adenine 18 (3.2Å) and N3 of adenine 17 (2.8Å). The N2 of the amidine group also forms a strong hydrogen bond with a water molecule (2.1Å). The hydroxyl moiety in this structure is facing away from the minor groove and it forms a hydrogen bond with phosphate oxygen OP1 of thymine 7 via two water molecules. The amidine group on the other end of the ligand attached to the phenyl ring does not form any direct hydrogen bonds with the DNA sequences, contrasting with the amidine group on the other end. The N5 and N6 of the amidine group are involved in a dense water network and forms hydrogen bonds with both strands of DNA (Figure 4.5.c and 4.11.b).

Outer interactions: The OP1 of thymine 8 of chain A is connected to OP1 of thymine 21 of chain B, via two water molecules, 170 and 30.

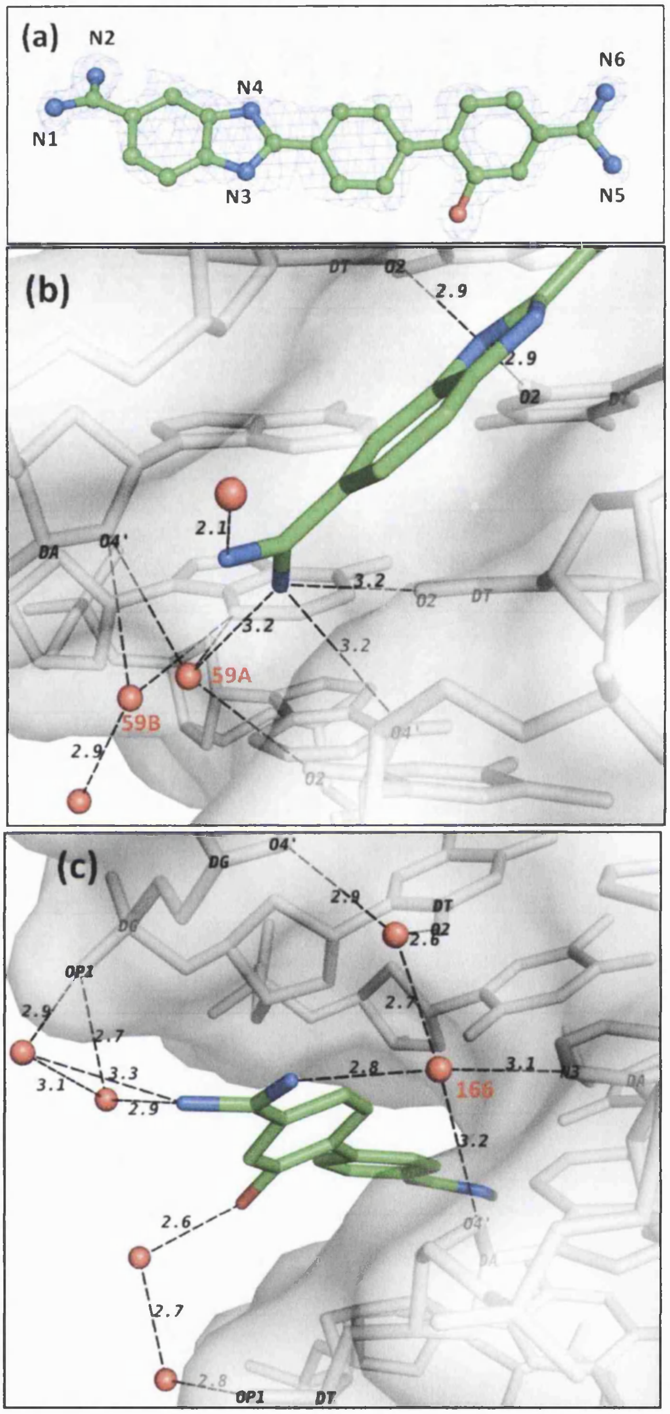


Figure 4.5. (a) The final $2F_o$-F_c electron density map calculated for the DB985 bound in the minor groove of A_3T_3 at 1.0σ level contour. (b) A side view of the DNA minor groove showing the benzimidazole-amidine ends of the ligands. (c) A top view, looking into the minor groove at the phenyl-amidine end of the ligand. DB985 is shown as green stick model. Water molecules in the groove are shown as red spheres and DNA is shown as grey surface.

4.2. X-ray crystal structure of DB921 and d(CGCAAATTTGCG)$_2$

The crystal structure of DB921 reported here is in complex with A_3T_3 DNA to obtain higher resolution structural information and determine structural difference, if any with other linear ligand-DNA complex structures. The crystal structure of DB921 in complex with A_2T_2 has previously been determined PDB code 2B0K (Miao *et al.* 2005). The annealed oligonucleotide solution and other dilutions were prepared as described earlier in section 4.1.

4.2.1. Materials and methods

4.2.1.1. Crystallisation

The DB921-A_3T_3 crystals were grown with the same drop condition as for the DB985-A_3T_3 crystal (Section 4.1.1.1.). The drop was equilibrated against a reservoir of 2000μl of 50% v/v MPD solution at 285K (Figure 4.6).

Figure 4.6. Crystal of DB921 with A_3T_3 used for data collection had approximate dimensions of (0.2 x 0.05 x 0.05) mm^3.

4.2.1.2. Data collection

The DB921-A_3T_3 crystal was flash-frozen and a complete data set was collected on the micro-focus beamline ID23-2 at ESRF using X-rays at 0.87Å wavelength and a

MAR225 detector (Flot *et al.* 2010). The crystal detector distance was set at 152.90mm. 180 images were collected with oscillation range of 1.0° and an exposure time of 1.0 second per image. Reflections were integrated with the programme MOSFLM (Leslie. 1999) and scaled, merged and converted to structure factors using SCALA (Evans. 2006) and CTRUNCATE (Stein & Ballard. 2009) from the CCP4 suite (Collaborative Computational Project Number 4. 1994).

Following the refinement of the structure, ambiguous electron density was observed near a phosphate group of DNA. To further assess this, a low energy data set was collected on the same crystal to obtain anomalous information. This was done on the tunable wavelength beamline ID 23-1 at ESRF using X-rays of 1.9Å wavelength (http://www.esrf.eu/UsersAndScience/Experiments/MX/About_our_beamlines/ID23-1) and a Q315r ADSC X-ray detector. The crystal detector distance was set at 123.68mm and 720 images were collected with oscillation range of 1.0° and an exposure time of 0.5 seconds per image. The data was collected up to 1.9Å. The data was processed as described above. Data collection statistics for both structures are shown in Table 4.2.

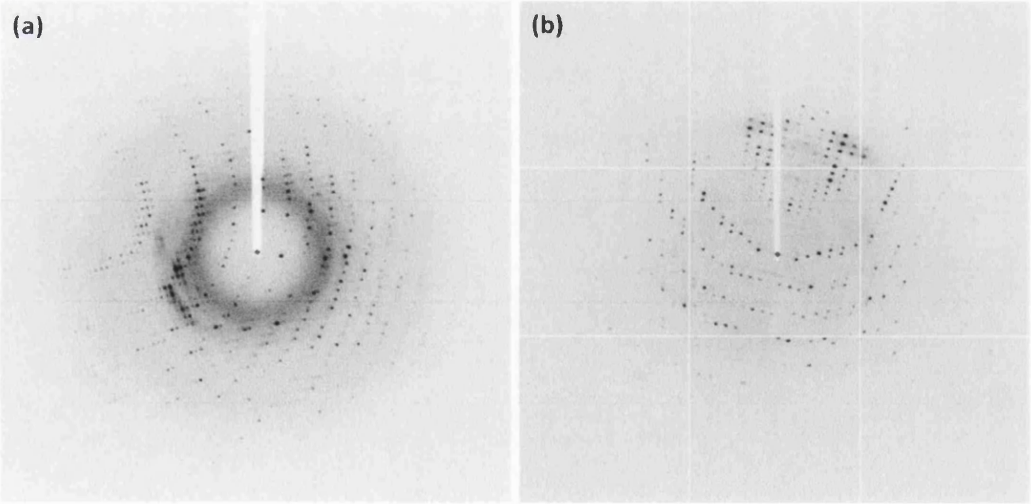


Figure 4.7. Diffraction pattern of the DB921-A_3T_3 crystal as collected on (a) ID 23-2 and (b) ID 23-1 beamlines at ESRF.

4.2.1.3. Structure solution and refinement

The initial structure solution and refinement was done as described in section 4.1.1.3 using the native DNA sequence of A_3T_3 as a model, PDB code 1S2R (Woods *et al.* 2004).

After the initial rigid body refinement the R-free and R-factor were 50.5% and 51.3% respectively. Further to this water molecules were added in a similar way to that described in section 4.1.1.3. The hydrated magnesium ions were identified and placed manually based on their octahedral geometry. The R-factor and R-free at this stage were 23.5% and 26.9% respectively.

The idealised model coordinates of DB921 were calculated using the Dundee PRODRG server (van Aalten *et al.* 1996). Although a continuous electron density corresponding to the ligand was observed in the minor groove, the electron density of the drug at this stage did not account for the whole drug molecule (Figure 4.8.a). Therefore, only the clearly visible benzimidazole and the central benzene rings were built into the electron density (Figure 4.8.b). Further refinement then resulted in clear $2F_o$-F_c and F_o-F_c electron density maps for the second benzene ring and this was then added to the model (Figure 4.8.c). The R-factor and R-free at this stage were 18.8% and 24.7% respectively.

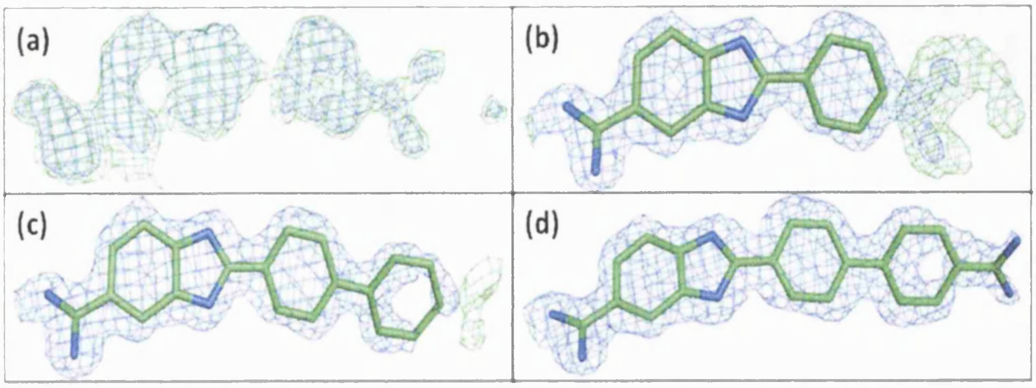


Figure 4.8. The electron density for DB921 in the minor groove of A_3T_3 DNA. The $2F_o$-F_c map is shown in blue at 1.0 σ level contour and the F_o-F_c map is shown in green at 3.0 σ level. The DB921 ligand is shown as a stick model. (a) Initial continuous electron density as seen in minor groove of DNA. (b) Following the refinement with model including only the benzimidazole and central phenyl ring, a prominent region was observed in the F_o-F_c map corresponding to the other benzene ring. (c) The last benzene ring was added to the model and refined which further showed the amidine group in the F_o-F_c map at 2.5 σ level corresponding to. (d) Final $2F_o$-F_c electron density map for the ligand at 1.0σ level.

At this stage of refinement strong positive F_o-F_c density was observed near the phosphate backbone of guanine residue number 12 on the Chain A (Figure 4.9.a), indicating an inadequately represented feature in the model. This ambiguous density close to the phosphate backbone could have been due to alternate conformation of the sugar-phosphate backbone (Figure 4.9.b). Hence, a low energy dataset was collected on the same crystal to obtain a peak in a phased anomalous difference map to unambiguously determine the position(s) of the phosphorus. Structure factors obtained from this low energy data set was further processed using the CAD and FFT programs from the CCP4 program suite to obtain a phased anomalous difference map. This map indicates a single anomalous peak for the phosphorus atom (Figure 4.9.c) which confirms the absence of alternated conformation for this residue. The model was further refined with a water molecule as shown in Figure 4.9.d.

The final R-factor and R-free values for the DB921-A_3T_3 structure are 18.4% and 25.0% respectively. Final coordinates were validated using the WHATIF web interface (Vriend. 1990) and analysed using w3DNA (Lu & Olson. 2003). The crystallographic refinement statistics for these structures are listed in Table 4.2.

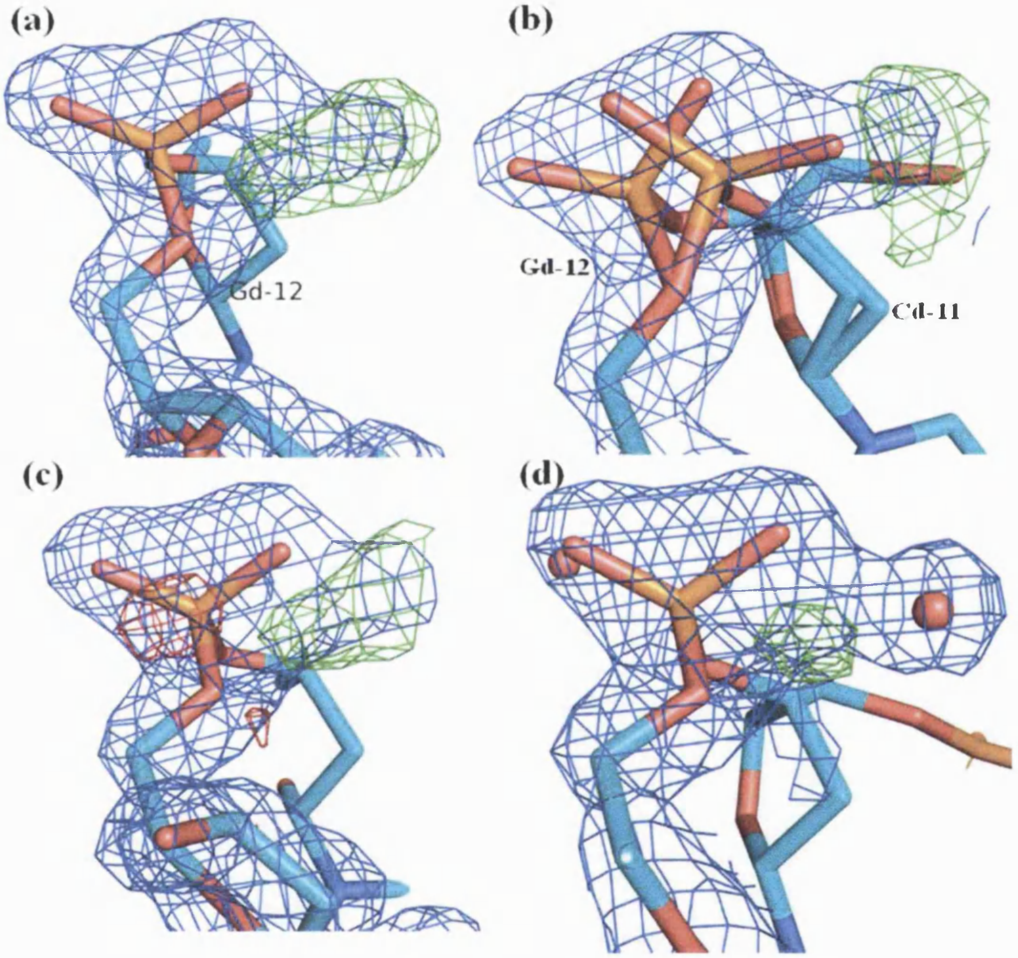


Figure 4.9. This figure illustrates the electron density maps near the phosphate backbone of guanine 12. DNA is represented as stick model. The $2F_o$-F_c electron density map is shown in blue at 1.0 σ level and the F_o-F_c map is shown in green at 3.0 σ level. (a) The strong positive F_o-F_c electron density map is shown in green (b) The model shown with alternate phosphate conformation of guanine 12 and sugar moiety of cytosine 11. (c) The phased anomalous difference map shown is red at 3.0 σ level, confirming the position of the phosphorus atom and single conformation of the phosphate backbone. (d) The final refined model with water molecules.

DNA sequence and Ligand	d(CGCAAATTTGCG)$_2$ and DB921 at ID 23-2	d(CGCAAATTTGCG)$_2$ and DB921 at ID 23-1
Data Collection		
Space group	$P2_12_12_1$	$P2_12_12_1$
Unit cell dimensions (Å,°)	a=23.62, b=38.88, c=62.66, $\alpha=\beta=\gamma=90.00$	a=23.62, b=38.71, c=62.64, $\alpha=\beta=\gamma=90.00$
Resolution (Å)	38.88-1.42 (1.50-1.42)	62.64-2.15 (2.27-2.15)
Wavelength (Å)	0.873	1.90
No. measured reflections	48637 (6981)	77949 (10752)
No. unique reflections	11003 (1597)	3416 (470)
Completeness (%)	96.60 (98.20)	99.80 (99.80)
$I/\sigma(I)$ for the data set	13.80 (2.80)	29.70 (8.50)
Data redundancy	4.40 (4.40)	22.80 (22.90)
Rmerge (%)	4.90 (47.00)	6.90 (38.90)
Refinement		
Resolution range	33.04-1.42 (1.48-1.42)	
No. reflections	10451 (784)	
Rfactor (%)	18.41 (27.20)	
Rfree (%)	25.04 (35.90)	
RMS deviation		
Bond length (Å)	0.025	
Bond angle deviation (Å)	3.41	
Number of atoms		
DNA	486	
DB985 ligand	27	
Magnesium	1	
Water (full/half occupancy)	103 / 6	
Mean temperature factor of atoms (Å^2)		
DNA	19.51	
DB985 ligand	27.42	
Magnesium	16.48	
Waters	31.67	

Table 4.2. X-ray data collection statistics is summarised for the two data sets collected for the DB921-A$_3$T$_3$ crystal on ID 23-2 and ID 23-1, along with the final refinement statistics. The high resolution shell statistics are shown in parentheses.

4.2.2. Results and discussion

The asymmetric unit of the crystal contains a duplex DNA (chain A and B), one molecule of DB921 bound within the central AAATTT tract of the minor groove, a fully hydrated magnesium ion and 103 full-occupancy and 6 half-occupancy water molecules. The ligand forms few direct bonds with the DNA but many water mediated contacts.

Direct drug-DNA contact: The N1 atom of the amidine group at the benzimidazole end of the ligand DB921 forms a direct hydrogen bond with the O2 of thymine 8 (3.0Å) of chain A. The inner facing nitrogen atom (N5) of the benzimidazole ring forms a pair of bifurcated hydrogen bonds with the O2 of thymine 7 (3.1Å) on chain A and O2 of thymine 19 (3.0Å) of chain B.

Water-mediated interactions: The N1 of the ligand forms a hydrogen bond with the 0.5 occupancy water molecule 99B (2.9Å) which further forms hydrogen bonds with O2 of thymine 9 (2.7Å) of chain A and O4' of the sugar moiety of adenine 18 (3.2Å) and N3 of adenine 17 (2.8Å) of chain B. The 0.5 occupancy water 99A further participates in the water network present in the minor groove of the DNA as shown in Figure 4.10.b. While the N2 of the ligand forms hydrogen bonds with water (2.9Å) which further participates in the water network in the minor groove similar to water 99A (Figure 4.10.b). The water network at the phenyl end of the ligand is very sparse in comparison to both DB985 structures. The N5 of the amidine group of DB921 forms a hydrogen bond with symmetry-related water 92 (3.0Å) which further hydrogen bonds with O1P of thymine 7 (2.7Å) and symmetry related O2P of adenine 6 (2.5Å) of chain A. The N6 of the amidine group forms a H-bond with water 55 (3.5Å). This water molecule interacts with another water molecule (3.2Å) which further interacts with OP1 of cytosine 23 of chain B. Water 79 in this structure is also observed in the other structures of DB985-

A_2T_2; DB985-A_3T_3 and DB921-A_2T_2 (Figure 4.11). However, water 79 does not form H-bonds with DB921 as seen in other structures because the amidine group is twisted out of plane, leaving the N6 atom 4.0Å away from water 79 (Table 4.3).

Outer interactions: A water bridge is formed which arches over the minor groove and connects the phosphate backbone of the two strands of DNA. The OP1 atom of guanine 12, chain A hydrogen bonds with water 40 (3.1Å). This further interacts with water 16 (2.6Å) which hydrogen bonds with OP1 of adenine 18 (3.0Å) and O3' of adenine 17 of chain B.

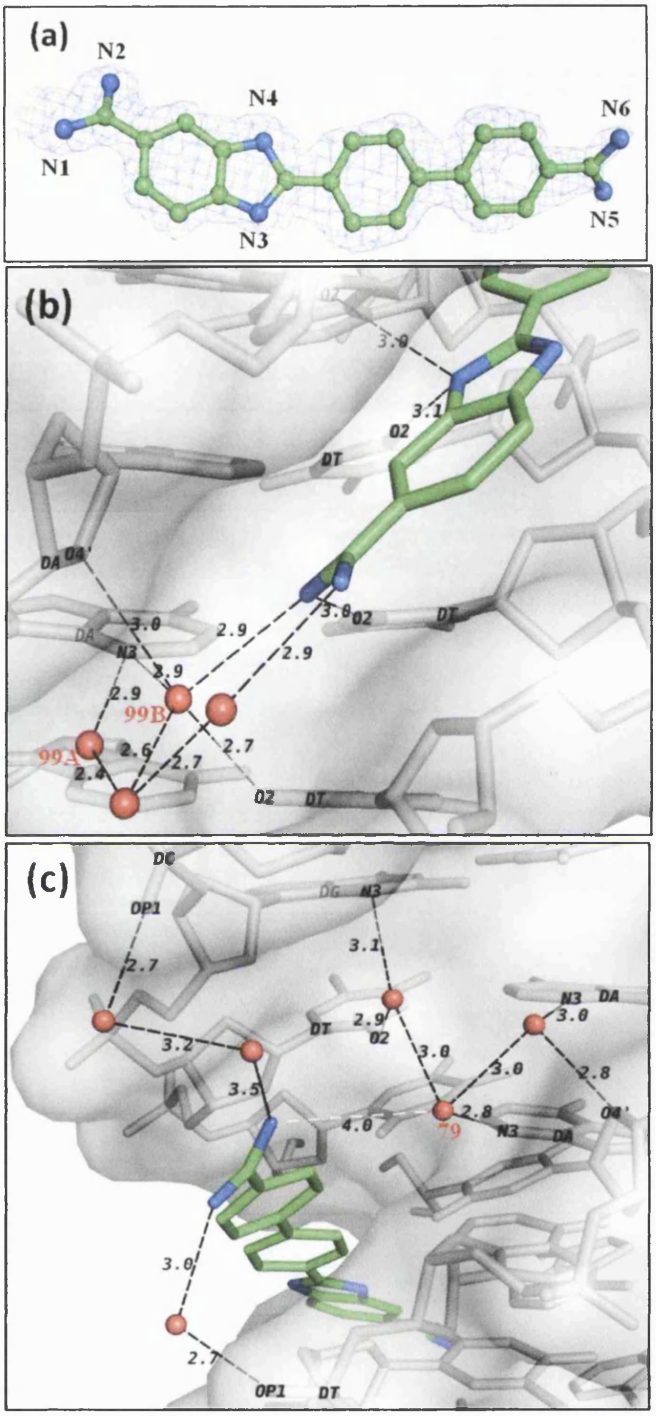


Figure 4.10. (a) The final $2F_o$-F_c electron density map calculated for the DB921 bound in the minor groove of A_3T_3 DNA at 1.0σ level. (b) A side view of the DNA minor groove showing the benzimidazole-amidine ends of the ligand, DB921. (c) A top view, looking into the minor groove at the phenyl-amidine end of the ligand. DB921 is shown as green stick model. Water molecules in the groove are shown as red spheres and DNA is shown as grey surface.

4.3. Comparison of the linear ligand structures

DB985 and DB921 are structurally very similar and both belong to the so called *out-of-shape* class of diamidine derivatives. DB921 lacks the hydroxyl moiety attached to the benzene ring as seen in DB985. The structures reported here are compared with DB921-A_2T_2 structure reported earlier, PDB code: 2B0K (Miao *et al.* 2005). In the DB985 crystal structures reported here, three magnesium ions were identified based on the octahedral geometry of the ion and was modelled at different occupancies, whilst both DB921 structures contains a single full occupancy magnesium ion.

The ligands in the reported structures are placed in the central A-T rich region of the minor groove of the DNA. The positions of the ligands in all four structures, DB985 and DB921 with A_2T_2 and A_3T_3, are similar. However, the geometry of the ligands in individual crystal structures is slightly different. The torsion angles of the ligands in the reported structures are in Table 4.3. While the benzimidazole and adjacent phenyl ring is planar in the DB921 complex with A_2T_2, it is twisted out of plane for other structures. The amidine group at the phenyl end of the ligands is twisted out of plane in all structures. In the DB921-A_3T_3 structure the group is twisted in an anti-clockwise direction (Table 4.3) which explains the different hydrogen bonding pattern in the structure (Figure 4.11.d).

The DNA chain of the reported structures were LSQ superposed using Coot (Emsley & Cowtan. 2004), to compare their structural similarity with the DB921-A_2T_2 complex structure (PDB code 2B0K), which is listed in Table 4.4.

Torsion angle / Structure	Benzimidazole-phenyl ring	Phenyl-Phenyl ring	Phenyl-amidine group	Benzimidazole-amidine group
DB985-A_2T_2	7°	33°	-34°	39°
DB985-A_3T_3	13°	36°	-29°	32°
DB921-A_2T_2 (2B0K)	1.0°	28°	-32°	45°
DB921-A_3T_3	18°	25°	28°	16°

Table 4.3. This table shows the torsion angles of the ligands (DB985 and DB921) in the different structures.

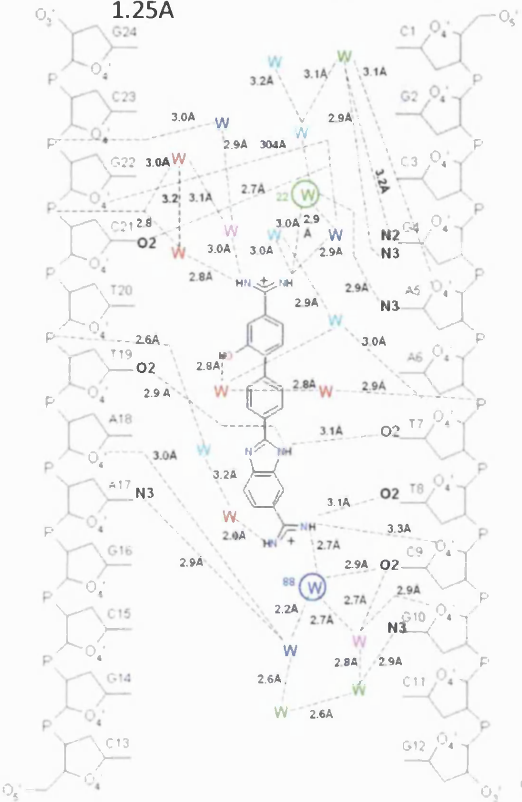
Structure	Number of atoms matched	Mean deviation (Å)	RMS deviation (Å)
DB985-A_2T_2	485	0.169	0.202
DB985-A_3T_3	518	0.54	1.54
DB921-A_3T_3	262	0.73	0.91

Table 4.4. This table shows the structural similarity of the reported structures in comparison to DB921-AATT structure in PDB (code 2B0K).

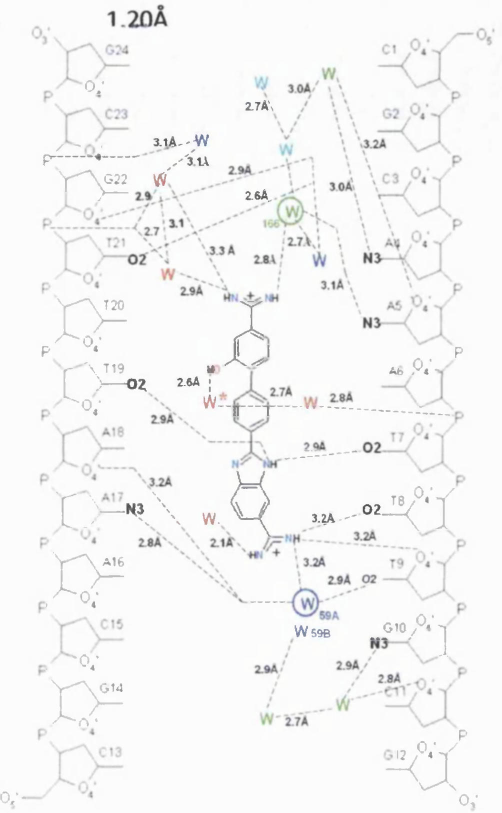
The DB985-A_2T_2 and DB985-A_3T_3 structures are closely related with very similar water network arrangements in the minor groove of the both the DNA sequences. Almost all water molecules in the minor groove are conserved in both DB985 structures (DB985-A_2T_2 and DB985-A_3T_3) and form similar hydrogen bonding pattern with the ligand and DNA (Figure 4.11.a and 4.11.b). In the DB921-A_2T_2 structure, the arrangement of the water molecules is sparse as compared to the DB985 structures and DB921-A_3T_3 structure at the benzimidazole end of the ligand (Figure 4.11). Both the ligands form similar direct hydrogen bonds with the bases of the DNA (Figure 4.11). The detailed network of water molecules in the minor groove of DB985-A_2T_2; DB985-A_3T_3; DB921-A_2T_2 and DB921-A_3T_3 can be seen in Figure 4.11.

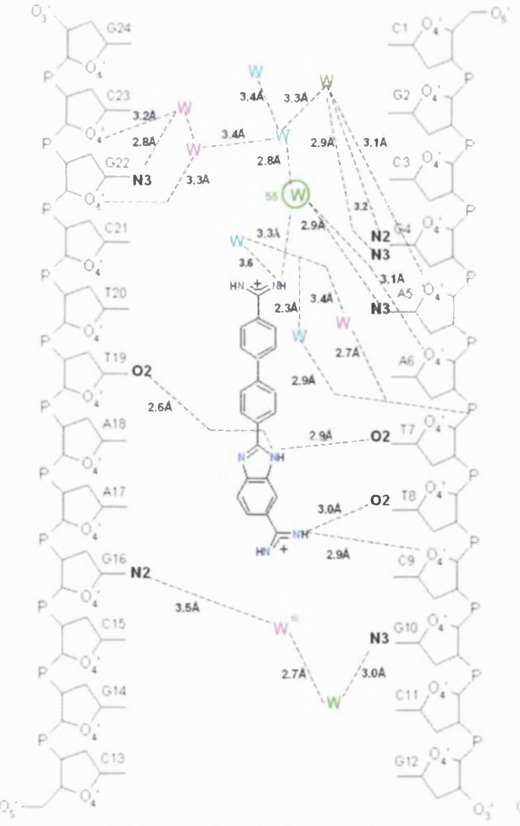
Figure 4.11. Schematic representation showing the hydrogen bonding of the linear ligands (a) DB985-A_2T_2, (b) DB985-A_3T_3, (c) DB921-A_2T_2 and (d) DB985-A_3T_3. Water molecules (shown in green) are present in all the above structure. Water molecules shown in blue are observed in DB985-A_2T_2; DB985-A_3T_3 and DB921-A_3T_3 structures. Cyan water molecules are present in DB985-A_2T_2; DB985-A_3T_3 and DB921-A_2T_2 structures. Red waters indicate presence in only DB985 structures, while magenta indicates their presence only in the individual structures.

(a) DB985 and d(CGCG<u>AATT</u>CGCG)$_2$ at 1.25Å

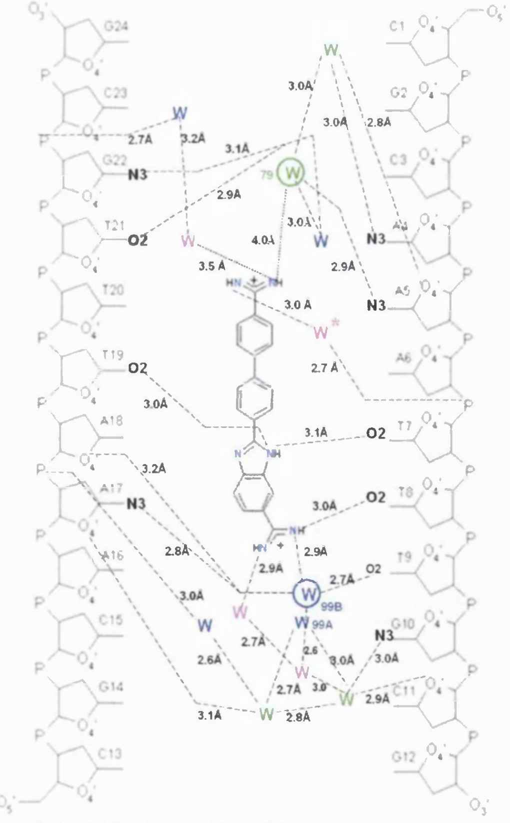


(b) DB985 and d(CGC<u>AAATTT</u>CGCG)$_2$ at 1.20Å





(c) DB921 and d(CGCG<u>AATT</u>CGCG)$_2$ at 1.64Å



(d) DB921 and d(CGC<u>AAATTT</u>GCG)$_2$ at 1.42Å

4.4. Comparison with the native B-DNA oligonucleotides

The ligand-DNA interactions described in this chapter are all associated with minor groove interactions that involve the displacement of ordered water from the minor groove of DNA. It is instructive therefore to consider these interactions in the context of the water structures in the native oligonucleotides.

The native A_2T_2 (PDB code: 1FQ2 at 1.2Å resolution) (Sines *et al.* 2000) and native A_3T_3 (PDB code: 1S2R, 1.5Å resolution) (Woods *et al.* 2004) structures were used as molecular model for structure solution of all structures reported here. The structures reported here are hence compared against these models to observe the change in DNA structure in presence of the ligands. The crystallisation conditions for the native DNA were slightly different to that of the ligand bound structures. The w3DNA web interface was used to compare the change in the minor groove width variation and local base-pair step parameters for the ligand bound and native structure (Lu & Olson. 2003).

The minor groove width in the DB985-A_2T_2 structure increases by a minimum of 0.8Å to a maximum of 2.0Å in the central AATT region when compared against the native A_2T_2 structure (PDB code: 1FQ2). For the DB985-A_3T_3 and DB921-A_3T_3 structures the minor groove width increases by a minimum of 0.5Å to a maximum of 1.3Å in comparison to the native A_3T_3 structure (PDB code 1S2R). It can also be seen in that there is a substantial increase in the tilt and roll parameters for the ligand bound structures in comparison to the native sequence for both A_2T_2 and A_3T_3 sequence (Table 4.5). While the shift, slide and rise parameters are similar in both native and ligand bound DNA sequences, with changes observed also in the twist parameter. However, the small structural differences observed could be due to different refinement programs and parameters of DNA used for the native structures.

When the native structures are compared to the ligand bound structure it is observed that the ligands (DB985 and DB921) displace the primary and secondary shells of hydration in the A-T tract of the DNA where the ligand binds. In the central A-T tract of both A_2T_2 and A_3T_3 native structure, there are five water molecules belonging to the primary shell of hydration which form direct bond with DNA and eight secondary layer of water molecules which are displaced by the ligand DB985. While the ligand DB921 displaces six water molecules belonging to the primary shell of hydration and eight water molecules of the secondary shell of hydration from the central A-T tract of A_3T_3 native structure (PDB code 1S2R).

Native A_2T_2 structure (PDB code: 1FQ2)							Native A_3T_3 structure (PDB code: 1S2R)						
Step	Shift	Slide	Rise	Tilt	Roll	Twist	Step	Shift	Slide	Rise	Tilt	Roll	Twist
GA/TC	-0.2	-0.1	3.4	-0.4	2.6	37.8	AA/TT	0.3	-0.0	3.4	0.3	0.0	37.7
AA/TT	0.1	-0.4	3.4	-0.3	-0.0	36.8	AA/TT	0.1	-0.4	3.3	-0.9	-0.6	37.8
AT/AT	-0.0	-0.6	3.3	-0.2	-1.0	32.4	AT/AT	0.1	-0.6	3.3	0.8	-0.4	30.7
TT/AA	-0.1	-0.4	3.3	1.7	-2.1	33.7	TT/AA	-0.3	0.3	3.3	1.6	-1.4	37.8
TC/GA	-0.2	-0.3	3.5	0.4	-1.3	42.5	TT/AA	-0.0	0.1	3.4	1.7	-3.4	41.0
DB985-A_2T_2							DB985-A_3T_3						
Step	Shift	Slide	Rise	Tilt	Roll	Twist	Step	Shift	Slide	Rise	Tilt	Roll	Twist
GA/TC	0.0	0.1	3.4	0.3	4.1	38.6	AA/TT	0.0	0.2	3.3	-1.2	2.1	37.8
AA/TT	0.5	-0.2	3.3	-0.5	3.8	34.9	AA/TT	0.5	-0.2	3.3	-1.8	4.0	36.2
AT/AT	-0.0	-0.6	3.3	-0.3	0.7	31.1	AT/AT	-0.2	-0.6	3.3	0.1	0.8	30.7
TT/AA	-0.4	-0.3	3.3	1.4	1.1	35.3	TT/AA	-0.4	-0.2	3.3	0.8	2.9	35.8
TC/GA	0.1	0.3	3.4	1.7	-0.4	41.1	TT/AA	-0.2	0.2	3.3	2.3	-1.4	39.9
DB921-A_2T_2							DB921-A_3T_3						
Step	Shift	Slide	Rise	Tilt	Roll	Twist	Step	Shift	Slide	Rise	Tilt	Roll	Twist
GA/TC	0.0	0.0	3.4	-0.3	-1.0	39.2	AA/TT	0.2	0.3	3.3	-0.5	0.3	38.0
AA/TT	0.5	-0.3	3.4	1.0	4.6	36.3	AA/TT	0.5	-0.1	3.3	-2.3	3.9	36.1
AT/AT	0.1	-0.7	3.3	-0.0	-1.0	30.3	AT/AT	0.0	-0.5	3.3	0.3	1.7	36.1
TT/AA	-0.4	-0.3	3.4	2.1	2.0	35.3	TT/AA	-0.4	-0.2	3.3	0.8	5.0	36.1
TC/GA	-0.1	0.3	3.3	2.8	-1.2	41.0	TT/AA	-0.4	0.1	3.3	1.5	-3.5	36.1

Table 4.5. The local base-pair step parameters in the central A-T tract of the native DNA and in the ligand bound structures.

4.5. Discussion

In this chapter the detailed crystal structures of two ligands DB985 and DB921 belonging to the *out-of-shape* class of ligands with two A-T rich sequences A_2T_2 and A_3T_3 are reported. It was observed that the ligands DB985 and DB921 both bind to the central A-T rich region of the DNA. They form direct hydrogen bonds with the bases of the DNA at the benzimidazole end of the ligand. However, due to the linear shape of the ligands they are unable to make direct contact with the DNA base at the phenyl end of the ligand. This is overcome by the stable water molecules present at the phenyl end of the ligand in the minor groove of the DNA. The water molecules participate in indirect contact of the ligand with bases of the DNA, hence providing a curved mode of binding which matches the classical iso-helical ligand binding mode. The water network in the minor groove of the DNA is very stable and plays a very important role by improving the ligand-DNA contact.

Biophysical studies have been done on both DB985 and DB921 which show that both these ligands have stronger binding to DNA in comparison to other iso-helical liagnds binding to the AATT site (Miao *et al.* 2005). This further shows that this class of ligand is promising and needs to be further studied and optimised to develop better drug with more binding efficacy and biological significance.

Chapter 5

X-ray crystallographic studies of iso-helical ligands

X-ray crystal structures of iso-helical diamidine ligands such as distamycin, berenil, Hoechst 33258, DB75, etc. have been studied in the past. This chapter describes the X-ray crystal structure of two iso-helical ligands DB818 and DB262 both with the double stranded d(CGCAAATTTGCG)$_2$ (A_3T_3) DNA sequence. High resolution X-ray structures are reported and these structures are compared to the linear ligand structures presented in the earlier chapter.

5.1. X-ray crystal structure of DB818 and d(CGCAAATTTGCG)$_2$

The crystal structure of DB818-A_2T_2 complex has been studied at 1.77Å resolution (Mallena *et al.* 2004). This study describes the crystal structure of DB818 with a longer A-T tract DNA, d(CGCAAATTTGCG)$_2$ (A_3T_3) at higher resolution. The structures were compared to observe the ligands interaction with two DNA sequences. This ligand was chosen due to its structural similarity to the linear ligands studied in Chapter 4.

5.1.1. Materials and methods

5.1.1.1. Crystallisation

The oligonucleotides used for the crystallisation was purified and annealed as described earlier (Chapter 3). The DNA and ligand dilutions were prepared with 30mM sodium cacodylate solution pH 6.5. Crystals of DB818-A_3T_3 DNA were obtained from condition 8 shown in Table 3.1. The hanging drop solution of 10µl contained 2µl of 15mM $MgCl_2$, 2µl of 25% MPD, 2µl of spermine tetra HCl, 2µl of 3mM duplex A_3T_3 DNA and 2µl of 5mM DB818 ligand. This drop solution was equilibrated against 40% MPD in the reservoir (Figure 5.1).

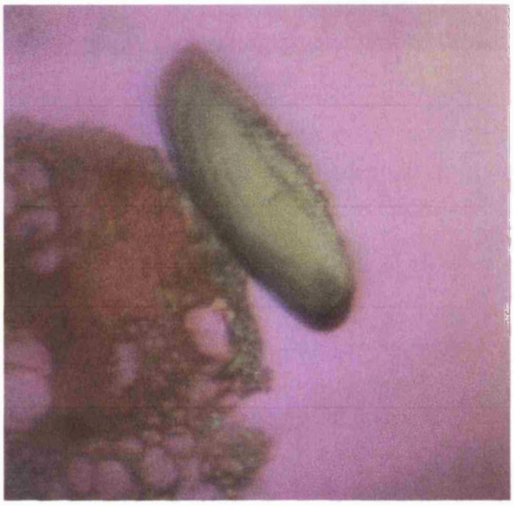


Figure 5.1. Crystal of DB818-A_3T_3 complex of crystal size (0.3 x 0.1 x 0.1) mm^3.

5.1.1.2. Data collection

The crystal of DB818-A_3T_3 was cryo-protected by sweeping it rapidly through 100% paraffin oil and then flash-cooled in liquid nitrogen before exposing to X-rays. Diffraction data set was collected on the ID14-4 beamline (McCarthy *et al.* 2009) using a Q315r ADSC X-ray detector at the European Synchrotron Radiation Facility (ESRF, Grenoble, France) using X-rays of wavelength 0.98Å. The program EDNA was used to generate data collection strategy. According to the strategy suggested by EDNA, two passes of data collection were carried out at high and low resolution limits to avoid the consequences of detector saturation at low resolution. The low resolution pass data was collected first. 139 images were collected to 2.5Å resolution with an oscillation range of

2.5° and an exposure time of 0.2 seconds per image. 256 images of the high resolution pass were collected to 1.4Å resolution with an oscillation range of 0.5° and exposure time of 1.33 seconds per image. The reflections from the two data sets were integrated with the programme MOSFLM (Leslie. 1999) and then combined using POINTLESS (Evans. 2006). This was then scaled and converted to structure factors using SCALA (Evans. 2006) and CTRUNCATE (Stein & Ballard. 2009) from the CCP4 suite (Collaborative Computational Project Number 4. 1994). Following the refinement of this structure, some ambiguity was observed in the electron density map for the ligand.

In an attempt to remove this ambiguity through making use of the presence of a sulphur atom in the ligand, a low energy data set was obtained on another crystal from the same crystallisation drop. This data was collected on the tunable beamline ID 23-1 at ESRF (http://www.esrf.eu/UsersAndScience/Experiments/MX/About_our_beamlines/ID23-1) using X-ray radiation at 1.77Å wavelength and a Q315r ADSC X-ray detector. 720 images were collected with oscillation range of 1.0° and an exposure time of 0.5 seconds per image. The data was collected to 2.0Å resolution. The data was processed as described above. Data collection statistics for both structures are shown in Table 5.1.

5.1.1.3. Structure solution and refinement

Phasing information was obtained by the molecular replacement method using PHASER from the CCP4 suite (McCoy *et al.* 2007). The native DNA sequences of d(CGCAAATTTGCG)$_2$ [PDB code 1S2R] (Woods *et al.* 2004) was used as molecular models for structure solution. The model was stripped off all water molecules and ions prior to molecular replacement.

The refinement was started with a cycle of rigid body optimisation followed by restrained refinement using REFMAC5 (Murshudov *et al.* 1997). After the initial rigid body refinement the R-factor and R-free was 49.9% and 54.2%, respectively. The $2F_o$-F_c and F_o-F_c electron density maps obtained from this refinement were used to analyse ligand binding in the DNA minor groove. The electron density maps were viewed using COOT (Emsley & Cowtan. 2004). Continuous electron density was observed in the minor groove of both DNA sequences (Figure 5.2.a). Water molecules were added using the automated COOT water placement procedure and also placed manually by visual inspection of the $2F_o$-F_c map and hydrogen bonding distance with neighbouring atoms. Where water molecules were less than 2.0Å apart, they were each assigned a partial occupancy of 0.5. Hydrated magnesium ions were placed manually on the basis of octahedral geometry. The R-factor and R-free at this stage for the structure were 24.6% and 29.5%.

The idealised model coordinates of DB818 were calculated using the Dundee PRODRG server (van Aalten *et al.* 1996). Although a continuous electron density was observed in the minor groove, the electron density of the drug at this stage did not clearly indicate the position in which the ligand was bound. From the electron density maps it appeared that the ligand could be bound in two positions (Figure 5.2.b and 5.2.c) – rotating 180 degrees around the central point of the ligand.

The DB818 molecule contains a thiophene ring, hence to determine the position of the ligand a low energy data set was collected. The structure factors obtained from this low energy data set were processed with CAD and FFT programmes from the CCP4 programme suite to obtain a phased anomalous difference map using calculated phases from the partially refined structure. This map indicates a single anomalous peak for the sulphur atom (Figure 5.2.d) which confirmed the position of the ligand. However, when the ligand was refined in this position (position 2, Figure 5.2.c) it was observed from the electron density map that a cyclic ring could be modelled into the structure attached to the benzimidazole ring. To confirm that ligand used for crystallisation study was the expected molecule, mass spectroscopy was performed. Approximately 10 μM of ligand sample was sent to *Laboratoire Lésions des Acides Nucléiques* at CEA Grenoble to determine the mass of the sample. The expected molecular weight for DB818 ($C_{19}H_{16}N_6S$) was 360 Da. The electron spray ionisation (ESI) mass spectroscopy was performed in both positive and negative mode. The ESI mass spectroscopy confirmed that the ligand solution was pure and had a molecular weight of 359 Da and 361 Da from negative and positive mode (Figure 5.3). This confirmed that the ligand used for crystallisation experiment was DB818 and it was then refined in the position confirmed by the anomalous peak with water molecules for the additional electron density (Figure 5.2.e).

The final R-factor and R-free for this structure are 19.8% and 26.4% respectively. Final coordinates were validated using WHATIF web interface (Vriend. 1990) and analysed using w3DNA (Lu & Olson. 2003). The crystallographic refinement statistics for the structure is listed in Table 5.1.

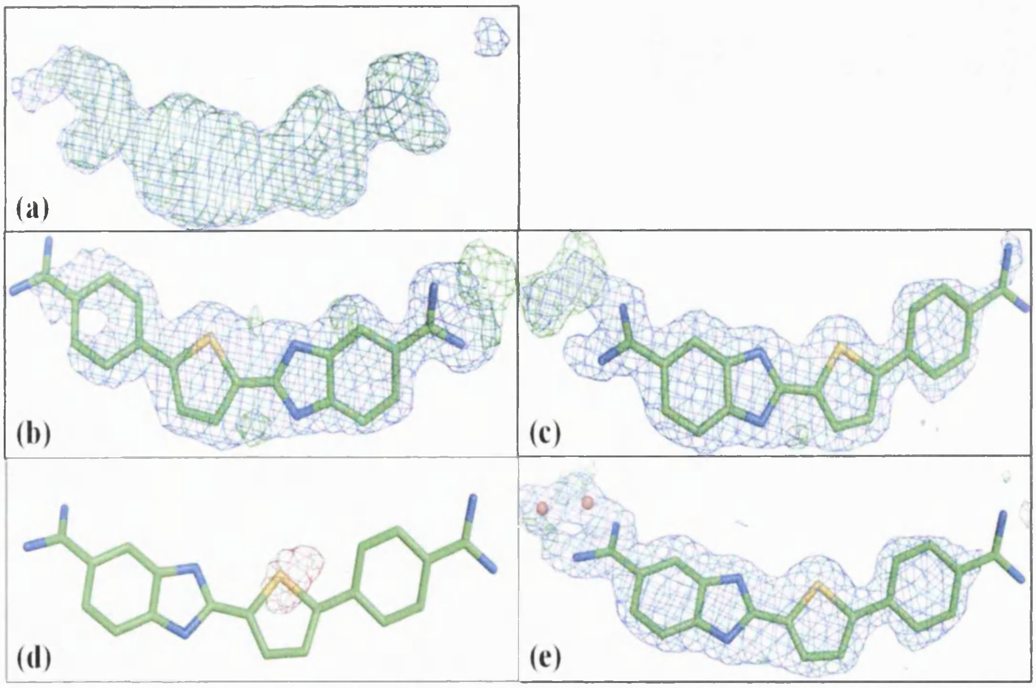


Figure 5.2. (a) The continuous $2F_o$-F_c and F_o-F_c electron density in the minor groove of the DNA after first cycle of restrained refinement. The $2F_o$-F_c and F_o-F_c map are shown at 1.0 σ level in blue and at 3.0 σ level in green, respectively. (b) DB818 in position 1 and (c) DB818 in position 2 in the minor groove of the DNA. (d) The anomalous peak shown in red corresponding to the sulphur atom of thiophene ring which confirms the binding position of ligand in position 2. (e) The final electron density map calculated for the DB818 bound in position with two 0.5 occupancy water molecules.

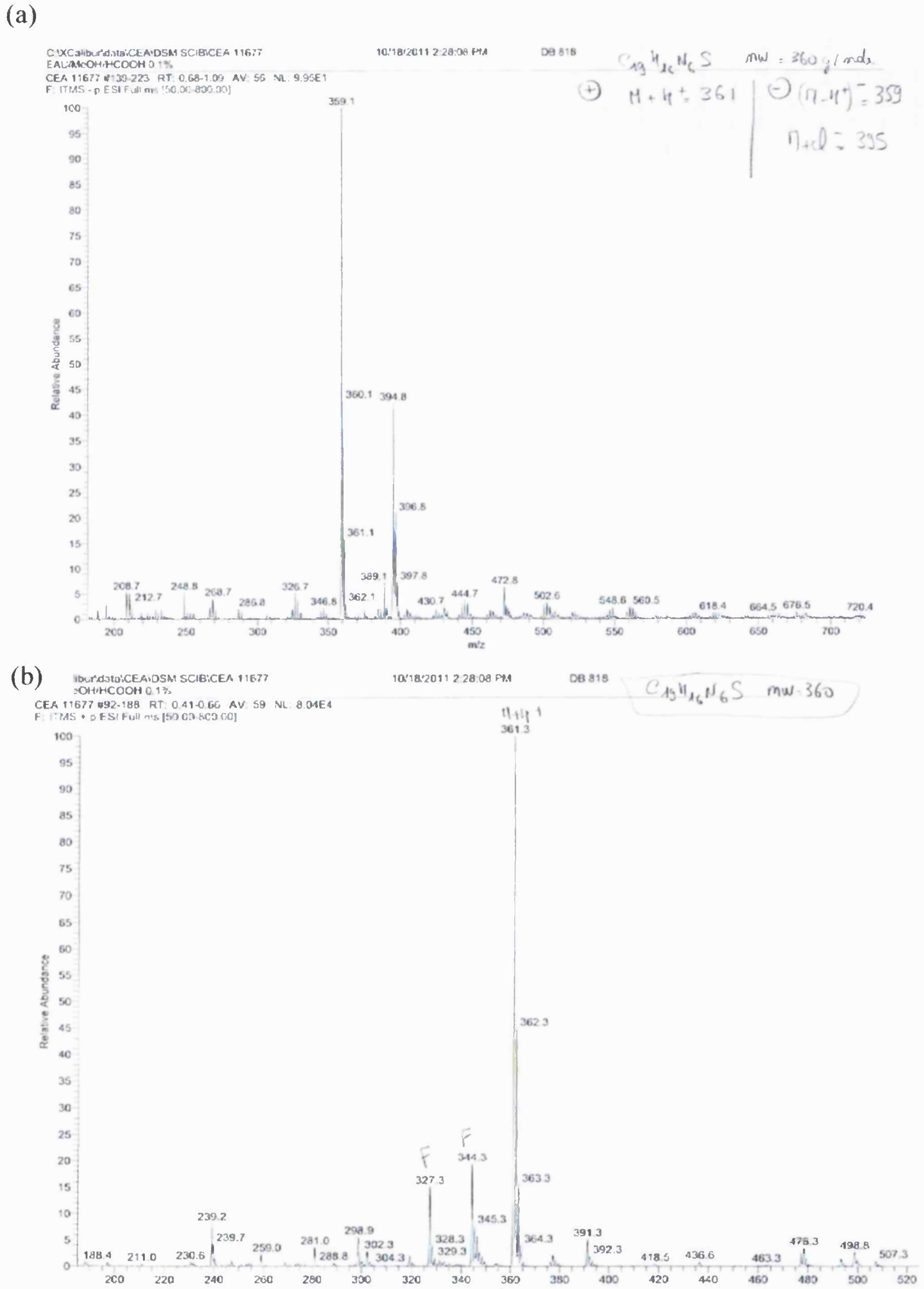


Figure 5.3. Mass spectrum of DB818 ligand obtained from (a) negative mode and (b) positive mode electron spray ionisation mass spectroscopy.

DNA sequence and Ligand	d(CGCAAATTTGCG)$_2$ and DB818 at ID 14-4	d(CGCAAATTTGCG)$_2$ and DB818 at ID 23-1
Data Collection		
Space group	$P2_12_12_1$	$P2_12_12_1$
Unit cell dimensions (Å,°)	a=24.69, b=40.47, c=64.38, $\alpha=\beta=\gamma=90.00$	a=24.79, b=40.59, c=64.34, $\alpha=\beta=\gamma=90.00$
Resolution (Å)	64.41-1.40 (1.48- 1.40)	40.59-2.00 (2.11-2.00)
Wavelength (Å)	0.98	1.77
No. measured reflections	72408 (5139)	117438 (17240)
No. unique reflections	13122 (1740)	4544 (634)
Completeness (%)	98.60 (93.0)	96.4 (94.5)
$I/\sigma(I)$ for the data set	18.10 (3.4)	35.9 (15.6)
Data redundancy	5.50 (3.00)	25.8 (27.2)
Rmerge (%)	4.80 (24.50)	6.6 (20.2)
Refinement		
Resolution range	17.43 – 1.40 (1.43-1.40)	
No. reflections	12409 (802)	
Rfactor (%)	19.81 (36.7)	
Rfree (%)	26.39 (43.0)	
RMS deviation		
Bond length (Å)	0.021	
Bond angle deviation (Å)	2.98	
Number of atoms		
DNA	486	
DB818 ligand	26	
Magnesium	1	
Water (full/half occupancy)	112 / 11	
Mean temperature factor of atoms (Å^2)		
DNA	28.38	
DB985 ligand	37.22	
Magnesium	24.08	
Waters	42.34	

Table 5.1. X-ray data collection statistics of the two data sets collected for DB818-A$_3$T$_3$ crystal on ID 14-4 and ID 23-1, along with the final refinement statistics of overall and high resolution shell. The high resolution shell statistics are shown in parentheses.

5.1.2. Results and discussion

The asymmetric unit of the crystal contains a duplex DNA (chain A and B), one molecule of DB818 bound within the central AAATTT tract of the minor groove, a fully hydrated magnesium ion and 112 full-occupancy and 11 half-occupancy water molecules. The hydrogen bonding of DB818 with DNA bases are explained in detail below (Figure 5.4). The amidine group attached to the phenyl end of the ligand has a poor $2F_o$-F_c electron density map (Figure 5.2.e) and could be only observed when the map was contoured to 0.7σ level.

Direct drug-DNA contact: The inner facing nitrogen atom of the amidine group attached to the benzimidazole ring of the ligand forms a direct hydrogen bond to O2 of thymine 20 (2.9Å) of chain B (Figure 5.4.a). The inner facing nitrogen atom of the benzimidazole ring forms a bifurcated hydrogen bond with O2 of thymine 7 (3.0Å) of chain A and O2 of thymine 19 (2.9Å) of chain B (Figure 5.4.a). The inner facing nitrogen atom of the amidine group attached to the phenyl ring of DB818 also forms a direct hydrogen bond with O2 of thymine 9 of chain A (3.1Å) (Figure 5.4.b).

Water mediated interactions: Both nitrogen atoms of the amidine group at the benzimidazole end of the ligand forms hydrogen bond with 0.5 occupancy water 175 A and 175 B. The nearest DNA atom to this water molecule is O4' of the sugar moiety of chain A adenine 6 (3.6Å) (Figure 5.4.a). The inner facing nitrogen atom of the amidine group at the phenyl end of the ligand forms hydrogen bond with water molecule water 41 (3.5Å). Water 41 forms hydrogen bond with N3 of guanine 10 (2.8Å) and O4' of sugar moiety of cytosine 11 (2.8Å) both belonging to chain A (Figure 5.4.b).

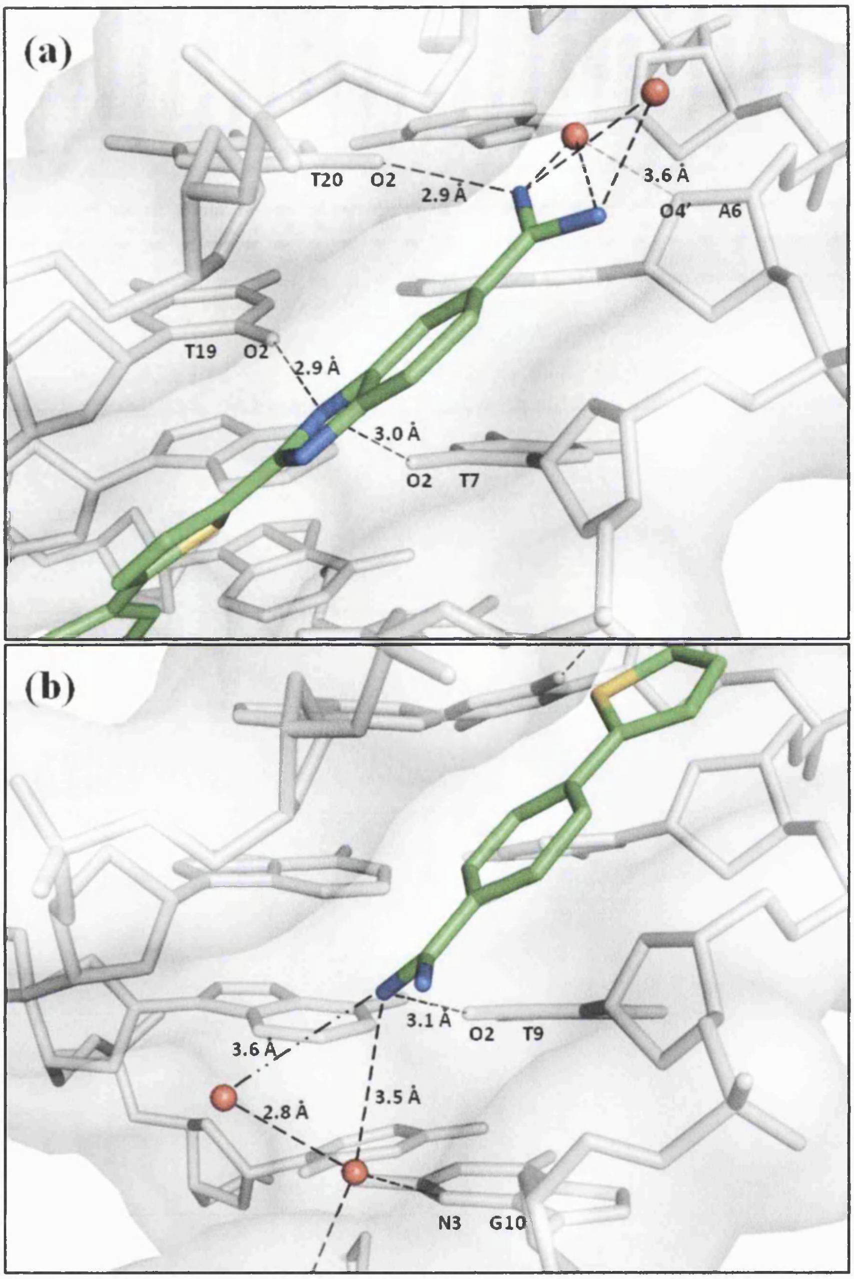


Figure 5.4. Hydrogen bonds between DB818 ligand and A_3T_3 DNA (a) A side view of the DNA minor groove showing the benzimidazole-amidine ends of the ligand, DB818. (b) A view into the minor groove at the phenyl-amidine end of the ligand. DB818 is shown as green stick model. Water molecules in the groove are shown as red spheres and DNA is shown as grey surface.

5.2. X-ray crystal structure of DB262 and d(CGCAAATTTGCG)$_2$

The ligand DB262 is structurally similar to furamidine (DB75), with the furan ring of DB75 replaced with a pyrrole ring in DB262. The crystal structure of DB75-A_2T_2 complex is available in the PDB at 2.2Å resolution, PDB code 227D (Laughton *et al.* 1996). In this section, the crystal structure of DB262 complexed with A_3T_3 DNA sequence is reported at 1.70Å resolution.

5.2.1. Materials and methods

To obtain the crystal structure of the DB262 DNA complex, several crystals were used. Two crystals of DB262 in complex with A_2T_2 DNA were used for X-ray data collection at ID 23-2 at ESRF. These crystals were obtained from condition 12 and condition 13 of Table 3.1. The structure solution of these datasets suggested no ligand binding in the minor groove of the DNA. Similar results were also obtained for some DB262-A_3T_3 DNA crystals grown in the same conditions.

However, two crystals of DB262 and A_3T_3 DNA complex provided diffraction data with ligand bound in the minor groove of the DNA. The crystallisation conditions and data collection strategy for these crystals are mentioned below.

5.2.1.1. Crystallisation

The crystals of DB262-A_3T_3 DNA used for the unsuccessful structure determination were obtained from the crystallisation conditions used for crystallising DB818 in complex with A_2T_2 (Mallena *et al.* 2004) and DB75 with A_2T_2 (Laughton *et al.* 1996). The data were collected at beamlines ID 23-2 and ID 29 at ESRF (Figure 5.5.a). After initial refinement of the structures it was evident from the electron density maps that the minor groove of DNA only had water molecules bound. Hence the data obtained from these crystals were not further pursued.

The crystal that provided the crystal structure of DB262-A_3T_3 complex was obtained from condition 4 of Table 3.1. The hanging drop contained 3µl each of 50mM $MgCl_2$, 30% MPD, 2.5mM spermine, 2mM double strand A_3T_3 DNA and 5mM DB262 ligand. The drop was equilibrated against 1ml of 50% MPD in reservoir. The crystals were obtained at 12°C and were of approximately (0.7 x 0.1 x 0.05) mm^3 volume (Figure 5.5.b).

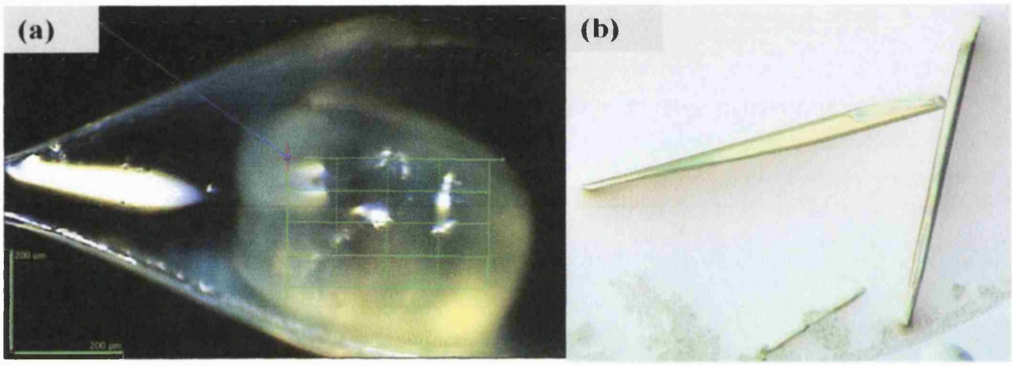


Figure 5.5. Crystals of the DB262–A_3T_3 complex.

5.2.1.2. Data collection

The crystal was flash-cooled in liquid nitrogen without any cryo-protection, and the X-ray diffraction data was collected on the ID 29 beamline (http://www.esrf.eu/UsersAndScience/Experiments/MX/About_our_beamlines/ID29) using the Pilatus 6M detector at the ESRF using X-rays of 0.95Å wavelength. 720 diffraction images were collected at 1.4Å resolution with oscillation degree of 0.25° and an exposure time of 0.3 seconds. The reflections were processed using XDS (Kabsch. 2010). The merged reflections in XDS format were then converted to mtz file format using the COMBAT program from the CCP4 suite (Collaborative Computational Project Number 4. 1994). This was then scaled and converted to structure factors using SCALA (Evans. 2006) and CTRUNCATE (Stein & Ballard. 2009) from the CCP4 suite

(Collaborative Computational Project Number 4. 1994). The data collection statistics for the structure are shown in Table 5.2.

5.1.1.3. Structure solution and refinement

PHASER from the CCP4 suite (McCoy *et al.* 2007) was used for molecular replacement to obtain the phasing information. The native A_3T_3 DNA sequences [PDB code 1S2R] (Woods *et al.* 2004) was used as molecular models for structure solution. This model was stripped off all water molecules and ions prior to molecular replacement.

First a rigid body refinement was performed and this was followed by restrained refinement using REFMAC5 (Murshudov *et al.* 1997). After the initial rigid body refinement the Rfactor and Rfree of the structure was 55.2% and 51.0% respectively. Following the restrained refinement the Rfactor and Rfree for the structure was 36.8% and 32.6% respectively. Water molecules were then added using the automated COOT water placement procedure and also placed manually by visual inspection of the $2F_o$-F_c and F_o-F_c maps. Water molecules that were less than 2.0Å apart were each assigned a partial occupancy of 0.5. Hydrated magnesium ion was placed manually on the basis of octahedral geometry. The R-factor and R-free at this stage for the structure was 26.3% and 21.8%.

A continuous electron density was observed in the minor groove of the DNA (Figure 5.6.a). The ligand was modelled into the electron density. However, the electron density for the ligand was not very clear and hence DB262-A_3T_3 was again crystallised with higher DNA-ligand ratio (1:4) in the drop solution.

Crystals were obtained within 2 weeks and data on this crystal were collected on beamline ID 23-1 at the ESRF. The data was processed and refined as mentioned earlier. The crystal diffracted to 1.48Å resolution. Although this crystallisation condition used higher concentration of DB262 and the crystal provided better diffraction resolution, the electron density map for the ligand in the minor groove was worse than that observed for the previous crystal (Figure 5.6.b).

Hence the structure was pursued with the data obtained for the previous crystal. The idealised model coordinates for the ligand were calculated using the Dundee PRODRG server (van Aalten *et al.* 1996). The position of the ligand was finalised (Figure 5.6.c) and the structure was completed by manually adding the required water molecules. The final Rfactor and Rfree were 23.1% and 17.8%, respectively. The final coordinates were validated using WHATIF web interface (Vriend. 1990) and analysed using w3DNA (Lu & Olson. 2003). The crystallographic refinement statistics for the structure are listed in Table 5.2.

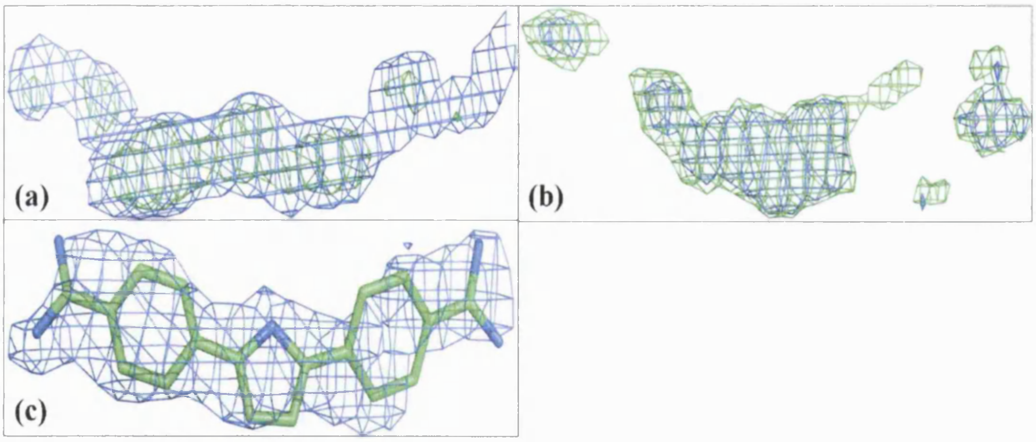


Figure 5.6. (a) The electron density in the minor groove of the DNA after first cycle of restrained refinement. The $2F_o$-F_c and F_o-F_c electron density maps are contoured at 1.0 σ level in blue and at 3.0 σ level in green, respectively. (b) The electron density map obtained from the crystal with higher concentration of DB262 ligand (c) The final electron density map of DB262 bound in minor groove of DNA.

DNA sequence and Ligand	$d(CGCGAATTCGCG)_2$ and DB262
Data Collection	
Space group	$P2_12_12_1$
Unit cell dimensions (Å,°)	a=23.76, b=39.14, c=62.75, $\alpha=\beta=\gamma=90.00$
Resolution (Å)	39.14-1.70 (1.79-1.70)
Wavelength (Å)	0.95
No. measured reflections	42511 (5648)
No. unique reflections	6682 (884)
Completeness (%)	97.9 (93.5)
$I/\sigma(I)$ for the data set	17.5 (6.6)
Data redundancy	6.4 (6.4)
Rmerge (%)	5.4 (23.0)
Refinement	
Resolution range (Å)	39.14-1.70 (1.74-1.70)
No. reflections	6365 (398)
Rfactor (%)	17.81 (21.40)
Rfree (%)	23.13 (25.90)
RMS deviation	
Bond length (Å)	0.021
Bond angle deviation (Å)	2.98
Number of atoms	
DNA	486
DB262 ligand	23
Magnesium	1
Water (full/half occupancy)	71/ 7
Mean temperature factor of atoms (Å^2)	
DNA	30.81
DB262 ligand	52.93
Magnesium	24.07
Waters	37.97

Table 5.2. X-ray data collection statistics for the complex of DB262-A_3T_3 on ID 29, along with the final refinement statistics of overall and high resolution shell. The high resolution shell statistics are shown in parentheses.

5.2.2. Results and discussion

The asymmetric unit of the crystal contains a duplex DNA (chain A and B), one molecule of DB262 bound within the central AAATTT tract of the minor groove, a fully hydrated magnesium ion and 71 full-occupancy and 7 half-occupancy water molecules. The hydrogen bonding of DB262 with DNA bases are explained in detail below (Figure 5.7).

Direct drug-DNA contact: The inner facing nitrogen atom of the amidine group which is located in between two A-T base (A5-T20 and A6-T19) forms a direct hydrogen bond to O4' belonging to the sugar moiety of adenine 6 (3.3Å) of chain A (Figure 5.7.a). The inner facing nitrogen atom of the central pyrrole ring of the ligand forms hydrogen bond with O2 of thymine 19 (3.3Å) of chain B. This nitrogen atom is also closely placed to the O2 of thymine 7 (3.6Å) of chain A and O4' atom of sugar moiety belonging to thymine 20 (3.6Å) of chain B (Figure 5.7.a). The inner facing nitrogen atom of the other amidine group forms a direct hydrogen bond with O2 of thymine 8 (3.4Å) and O4' of thymine 9 (3.3Å) of chain A (Figure 5.7.b).

Water mediated interactions: The inner facing nitrogen atom of the amidine group located in between two A-T base (A5-T20 and A6-T19) forms a hydrogen bond with water 113 (2.6Å) which further forms a hydrogen bond with O2 of thymine 21 (3.1Å) of chain B and N3 of adenine 5 (3.5Å) and O4' of adenine 6 (3.3Å) both belonging to chain A (Figure 5.7.a). The nitrogen atoms of the other amidine group of the ligand participate in a dense water network at this end of minor groove via water molecule 107 (3.0 and 3.1Å) (Figure 5.7.b). This water molecule hydrogen bonds with water 97 (2.7Å), which further interacts with water 108 (2.5Å), water 72 (2.8Å) and water 106 (3.0Å). Water 108 interacts with the N3 atom of adenine 17 (2.7Å) of chain B, whilst water 106 forms a hydrogen bond with N3 of guanine 10 (2.8Å) and O4' of cytosine 11

(2.7Å) both belonging to chain A of DNA. Water 72 forms a hydrogen bond with the O3' of guanine 14, chain B belonging to a symmetry related DNA molecule.

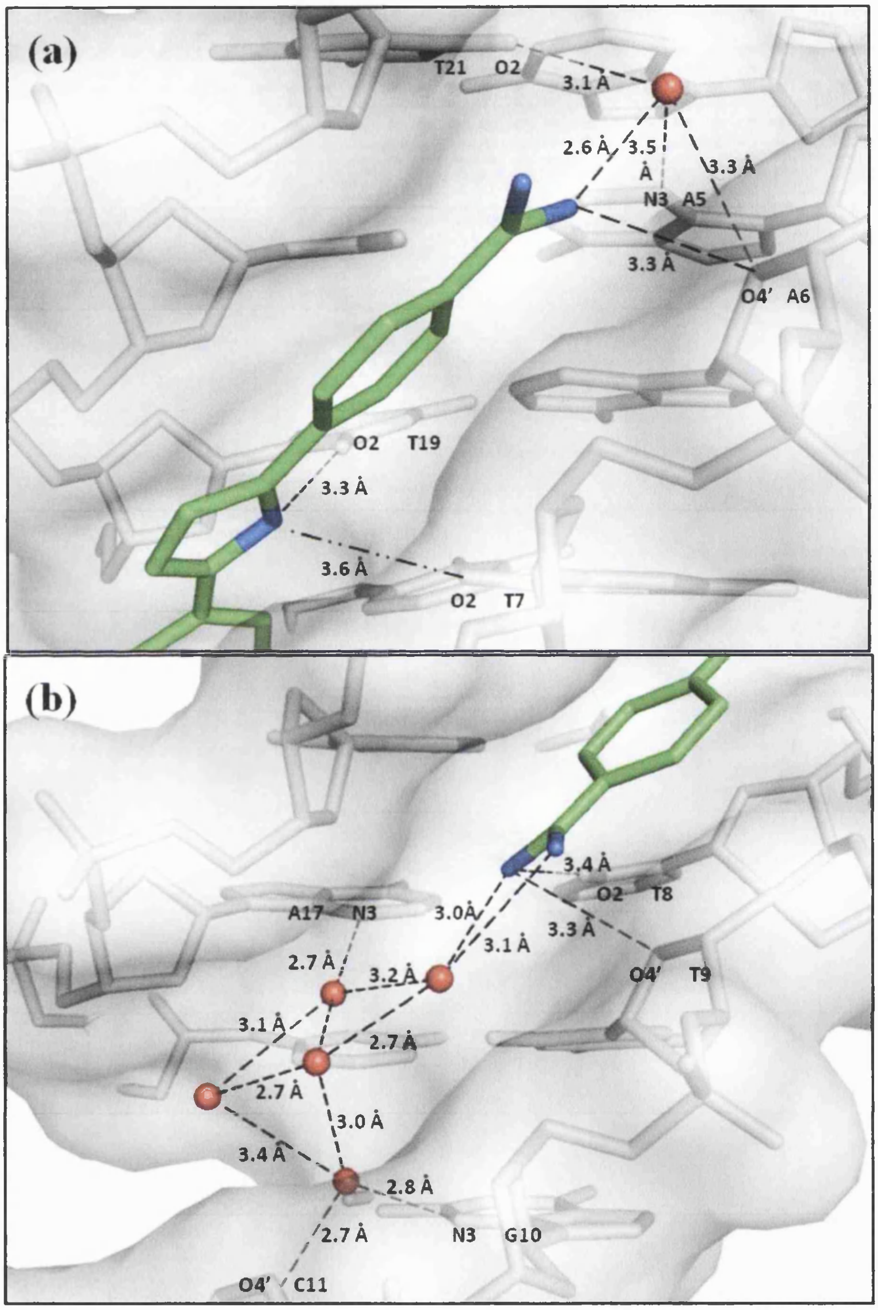


Figure 5.7. Hydrogen bond network between DB262 and A_3T_3 DNA (a) A side view of the DNA minor groove showing the amidine group of the ligand located between two A-T bases (A5-T20 and A6-T19) (b) A view into the minor groove showing the other amidine group of the ligand. DB262 is shown as green stick model. Water molecules in the groove are shown as red spheres and DNA is shown as grey surface.

5.3. Comparison with other ligand bound structures

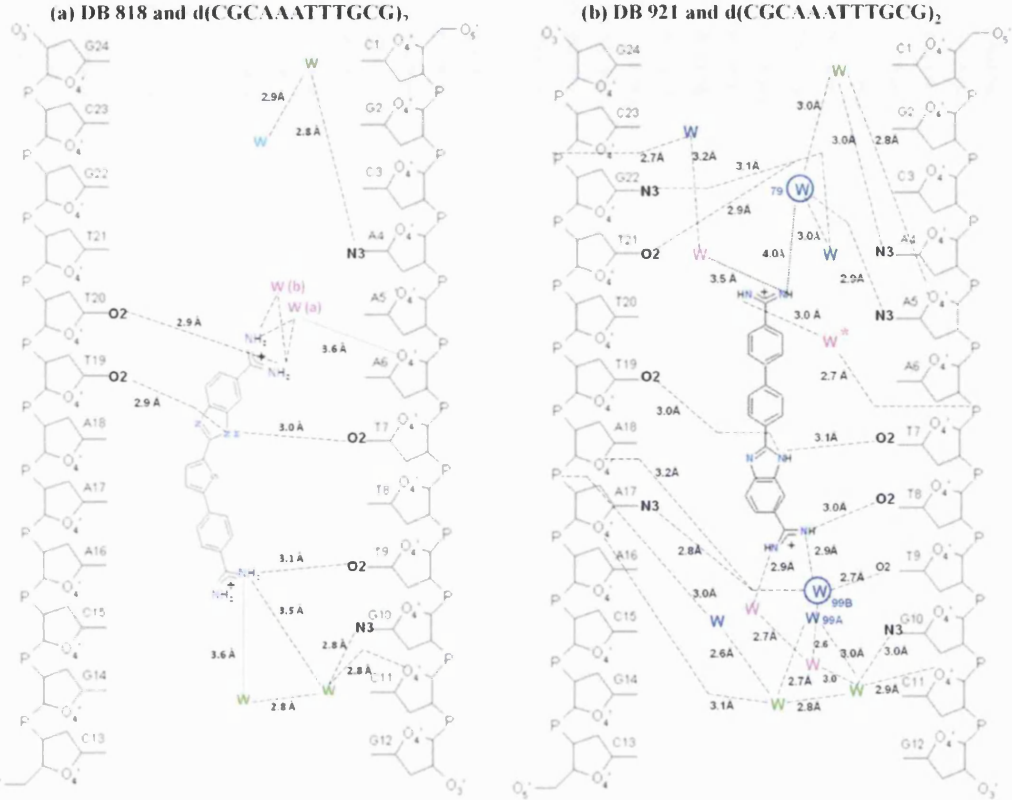
DB818, the iso-helical ligand is structurally similar to linear ligand DB921. The central ring of DB921 is a phenyl ring while the central ring of DB818 is a five member thiophene ring. This five member ring provides DB818 its curved morphology. There are a number of significant differences between the DB818-A_3T_3 DNA crystal structure and the DB921-A_3T_3 DNA structure reported in Chapter 4.

As seen from the structure of DB921 and A_3T_3 DNA, the ligand is extended between the base-pairs A4-T21 to T8-A17 in the minor groove of the DNA, whereas DB818 is shifted by 1 base-pair in the minor groove region and it extends between the A5-T20 to T9-A16 base-pair. Another significant difference in the binding of these ligands is in the position of the benzimidazole ring of the ligands. The benzimidazole ring of DB818 ligand is positioned at the 5' end of chain A while the benzimidazole ring of DB921 is positioned towards the 3' end of chain A. In spite of this "180° head-to-head flip" in the position of the two ligands, the inside facing nitrogen atoms of the benzimidazole ring are in the same position for both ligands. The position of the inner facing nitrogen atom of the benzimidazole ring is maintained in both structures due to the single base-pair shift in the binding of the ligands in the minor groove. Hence the inner facing nitrogen atom of both ligands shows the same hydrogen bonding ability with the O2 atom of thymine 7 of chain A and O2 of thymine 19 of chain B (Figure 5.8).

In both of the structures it was observed that the amidine group present at the 3' end of chain A forms a direct hydrogen bond with the thymine base the DNA, thymine 9 for DB818 and thymine 8 for DB921 both belonging to chain A. But the hydrogen bonding of the amidine group at the 5' end of DNA is different. The iso-helical ligand, DB818, can form a direct hydrogen bond with DNA (O2 of thymine 20 chain B); the linear ligand, DB921, cannot form any direct bonds with DNA at this end (Figure 5.8.b).

DB921 only forms water mediated interactions with the DNA. This difference in the hydrogen bonding ability of the ligands is directly due to their differences in the curvature of the two ligands. Due to the curved morphology of DB818 the amidine group at the 5' end of the DNA is placed deeper into the groove, whilst the linear morphology of DB921 causes the amidine group to protrude out of the minor groove compared to DB818.

It can also be noted that the water molecule 79 (Figure 5.8.b) observed in the DB921 structure and also in the crystal structure of other linear ligands (Chapter 4) is not observed in the DB818 bound DNA structure. This could be because the presence of this water molecule in the linear ligand structures ensures interaction of ligand and DNA at the 5' end of DNA while in case of iso-helical ligand DB818, a water molecule in this position does not contribute to an enhanced hydrogen bonding network. However, in the case of iso-helical ligand DB262, a water molecule is clearly visible in this position based on the electron density map (Figure 5.9.a). The presence of this water molecule in the DB262 crystal structure further ensures interactions between DB262 and DNA via the water molecule. The hydrogen bonding distance between the amidine group of DB262 and DNA is comparatively longer (3.3Å) than as observed with DB818 (2.9Å) at the 5' end of DNA (Figure 5.8.a.and 5.9.a). The water network observed in the minor groove of DB818 bound DNA structure is not as dense as other the DB921 bound structure especially at the 5' end of DNA chain A.


(a) DB 818 and d(CGCAAATTTGCG)$_2$

(b) DB 921 and d(CGCAAATTTGCG)$_2$

Figure 5.8. Schematic representation showing the hydrogen bonding networks of (a) DB818-A_3T_3 and (b) DB921-A_3T_3. Water molecules shown in green are present in both of the structures. Blue water molecules indicate that they are present in all of the linear ligand structures (DB985-A_2T_2, DB985-A_3T_3, and DB921-A_3T_3). Cyan water molecules are present in the structures of DB985-A_2T_2, DB985-A_3T_3 and DB818-A_3T_3. Water molecules represented in magenta indicates their presence only in the individual DB818-A_3T_3 and DB921-A_3T_3 structures.

DB262 is a smaller ligand compared to DB818. When the crystal structures of the two iso-helical ligands reported here are compared, it can be observed that due to this difference in size, while DB818 extends over 5 A-T base pairs from A5-T20 to T9-A16, DB262 only extends over 4 A-T base pairs involving A5-T20 to T8-A17. DB262 is structurally similar to DB75 (Furamidine) with the difference in their central furan ring in case of DB75 and pyrrole ring in case of DB262. While all these ligands (DB75, DB262 and DB818) are iso-helical in shape, the presence of the large sulphur atom in the central ring and the presence of the benzimidazole ring in DB818 endows it with a lower curvature (163.8°, or 16.2° away from linearity) while DB262 has a curvature of 155.2° and DB75 has the largest curvature of 147.2°. The increased curvature of DB75 can be seen as pushing the central region of the ligand away from the floor of minor groove, thereby preventing strong hydrogen bonding or van der Waals interactions between the aromatic rings of the ligand and DNA. DB818 with its curvature has a better fit into the minor groove of the DNA and also has a 30 times greater binding to DNA compared to DB75. DB818 has a binding constant of 2.8×10^8 M^{-1} while both DB75 and DB262 has similar binding constant of approximately 1×10^7 M^{-1} (Mallena *et al.* 2004; Munde *et al.* 2007).

From an early docking study of DB262 based on the crystal structure of DB75 it was thought that both DB75 and DB262 would have similar hydrogen bonding interactions with the DNA bases in the minor groove of the DNA (Munde *et al.* 2007). However, from the crystal structure of DB262 obtained in the work presented here, it is evident that there is a difference in the interaction of these ligands with DNA (Figure 5.9). DB262 is placed deeper into the minor groove compared to DB75 and hence the nitrogen atom of the central pyrrole ring can form a hydrogen bond with DNA (3.3Å apart) while the oxygen atom of the furan ring of DB75 is 3.8Å away from the DNA. This is the only significant but key difference in the crystal structure of these ligands. The amidine group of the ligands DB262 and DB75 have similar hydrogen bond forming ability.

However, because the electron density map obtained for DB262 (Figure 5.6.c) is not very clear and also the average B-factor for the ligand is very high at 52.93$Å^2$ (Table 5.2) it would be sensible to obtain better data from another crystal to confirm the position of the ligand DB262 in the groove. Also observed in the DB262 crystal structure is a dense water network in the minor groove towards the 3' end of DNA chain A which cannot be seen in the DB75 crystal structure (Figure 5.9).

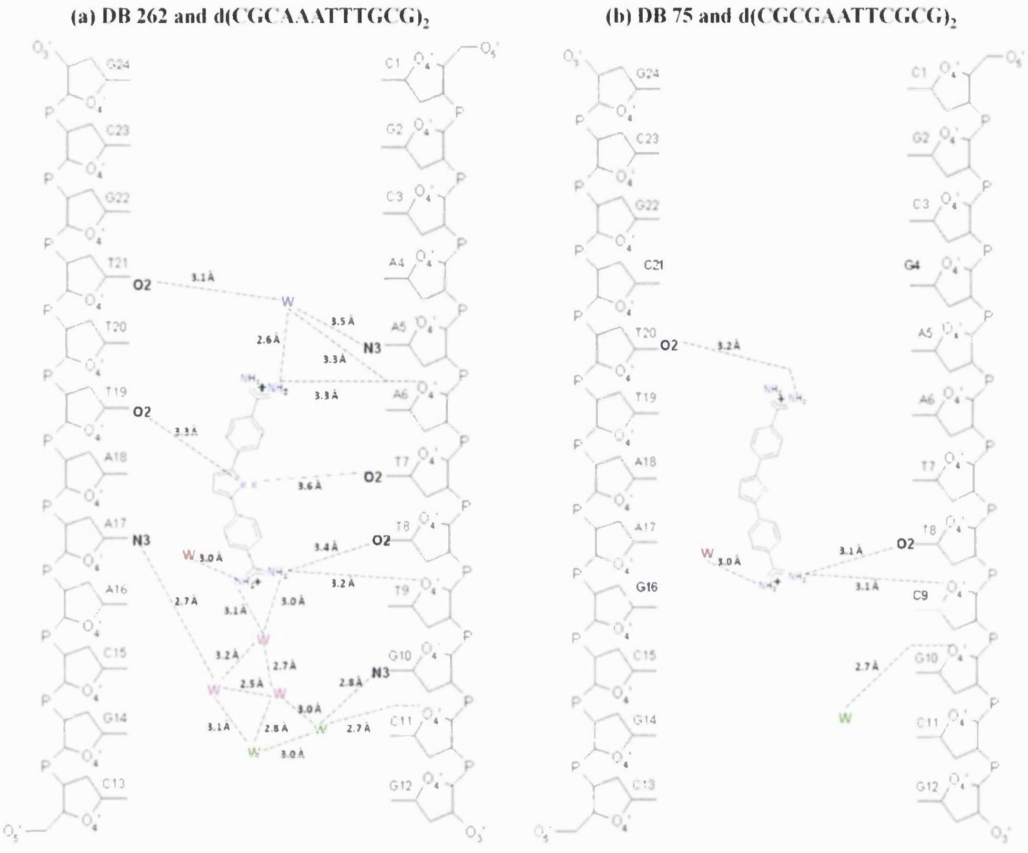


Figure 5.9. Schematic representation showing the hydrogen bonding of (a) DB262-A_3T_3 and (b) DB75-A_2T_2. Water molecules shown in red are present in both of the structures. Water molecules shown in green indicate their presence in all crystal structures studied here (linear ligands DB985 and DB921, and iso-helical ligands DB818, DB262 and DB75). Water molecules shown in blue are present in linear ligand structures and the DB262 structure, and waters shown in magenta indicates their presence only in the individual DB262-A_3T_3 and DB75-A_2T_2 structures.

5.4. Compared to the native B-DNA oligonucleotide structure

The native d(CGCAAATTTGCG)$_2$ (PDB code: 1S2R, 1.5Å resolution) (Woods *et al.* 2004) structure was used as the molecular model for the structure solution of DB818 and DB262 ligand bound DNA structures. These two structures are compared below to the native model to determine the changes in the DNA structure in presence of ligand. The w3DNA web interface was used to compare the changes in the minor groove width and local base-pair step parameters for the ligand bound and native structure (Lu & Olson. 2003).

The width of the minor groove in the central A-T region is approximately 9.7Å on average in the native DNA structure (PDB code: 1S2R). According to the w3DNA analysis the width of the minor groove changes to 10.4Å in the presence of DB818 and 10.1Å in the presence of DB262. There is a significant change in the tilt, roll and twist parameters of the DNA bases in the central A-T region in each case of the ligand bound structures (Table 5.3). The shift, slide and rise parameters are comparable in both native and ligand bound DNA sequences.

When the native structure is compared to the ligand bound structures, it is observed that the ligands displace the primary and secondary shells of hydration in the A-T tract of the DNA where the ligand is bound. The ligand DB818 displaces seven water molecules belonging to the primary shell of hydration and eight water molecules of the secondary shell of hydration from the central A-T tract of $d(CGCAAATTTGCG)_2$ native structure (PDB code 1S2R). DB262 also displaces seven water molecules of primary shell of hydration and eight water molecules belonging to the secondary shell of hydration.

Native A_3T_3 structure (PDB code: 1S2R)						
Step	Shift	Slide	Rise	Tilt	Roll	Twist
AA/TT	0.3	-0.0	3.4	0.3	0.0	37.7
AA/TT	0.1	-0.4	3.3	-0.9	-0.6	37.8
AT/AT	0.1	-0.6	3.3	0.8	-0.4	30.7
TT/AA	-0.3	0.3	3.3	1.6	-1.4	37.8
TT/AA	-0.0	0.1	3.4	1.7	-3.4	41.0
DB818-A_3T_3						
Step	Shift	Slide	Rise	Tilt	Roll	Twist
AA/TT	0.1	0.1	3.4	-1.6	2.3	36.6
AA/TT	0.2	-0.1	3.3	-2.1	2.9	36.2
AT/AT	-0.1	-0.5	3.2	0.0	0.9	31.6
TT/AA	-0.3	-0.2	3.3	2.0	1.8	35.5
TT/AA	-0.3	0.1	3.3	2.2	-2.5	40.0
DB262-A_3T_3						
Step	Shift	Slide	Rise	Tilt	Roll	Twist
AA/TT	0.2	0.3	3.3	-0.2	-1.5	37.3
AA/TT	0.1	-0.1	3.3	-3.5	3.2	37.6
AT/AT	0.2	-0.6	3.3	-0.6	1.1	30.8
TT/AA	-0.2	-0.3	3.3	0.9	2.6	36.7
TT/AA	-0.5	-0.1	3.2	3.3	-1.3	39.5

Table 5.3. The local base-pair step parameters in the central A-T tract of the native DNA and in the ligand bound structures.

5.5. Discussion

The crystal structures of the two iso-helical ligands DB818 and DB262 in complex with A_3T_3 DNA are reported here. The ligands are bound between the central A-T tract of the minor groove of the A_3T_3 DNA. DB818 extends over five base pairs while the smaller DB262 extends over four base pairs. The two amidine groups of both these iso-helical ligands form direct hydrogen bonds with DNA unlike the linear ligands (Chapter 4) which require water molecules to interact with DNA. The linear ligands DB921 and DB818 are structurally similar with differences in their central ring structure. However, this subtle change in the structure of the ligands results in significant difference in their

DNA binding ability and position of the ligand in the minor groove. In spite of these changes in the hydrogen bonding ability of these ligands, both have similar binding affinity to DNA: $K_a = 2.8 \times 10^8$ M^{-1} for DB818 and 2.7×10^8 M^{-1} for DB921 (Mallena *et al.* 2004; Miao *et al.* 2005). The other iso-helical ligand studied here, DB262, has an increased curvature compared to DB818 and the other linear ligands. This increased curvature of the ligand causes the central aromatic ring of the ligand to protrude from the minor groove. Due to this the nitrogen atom of the central pyrrole ring of DB262 is unable to form a bifurcated hydrogen bond with DNA where DB818 and other linear ligands are able to do so. This could be one of the major reasons for the reduced binding affinity (2.8×10^7 M^{-1}) of this ligand to DNA. From the crystal structure of the iso-helical ligands it can also be noted that these structures do not have a dense network of water molecules as observed with the linear ligands (Chapter 4).

From the early biophysical studies it is evident that all of the linear ligands and DB818 have a strong binding affinity for DNA and could have further biological relevance. Several synthetic analogues of DB75 have been designed. DB351 one such synthetic analogue has the central furan ring of DB75 changed to a thiophene ring, but this ligand has weaker DNA binding ability, whilst DB293 which has the phenyl ring of DB75 substituted for a benzimidazole ring it has similar binding affinity to DNA as DB75 (Mallena *et al.* 2004). When both these structural changes are made to the parent DB75 ligand, it results in DB818 which has 30 times stronger DNA binding affinity than both DB75 and DB293 (Mallena *et al.* 2004). This suggests that small structural changes in the ligand can cause significant change in the ligand-DNA binding. Therefore, although DB262 has a weaker binding affinity to DNA, it could be an important scaffold for further rational drug design to obtain better quality DNA binding drugs.

Chapter 6

The DNA Transcription Factor NF-κB

The DNA transcription factor NF-κB p50 (abbreviated hereafter as p50) belongs to the Rel family of eukaryotic transcription factors which regulates genes involved in process such as inflammation and the development of apoptosis (Speight *et al.* 2002). The other members of this family include the proteins p52, p65, c-rel and Rel B. The structural studies performed earlier on NF- κB shows that the protein wraps around DNA by binding to the guanine-cytosine rich zones in the major groove of the DNA (Chen *et al.* 1998; Cramer *et al.* 1997; Ghosh *et al.* 1995; Müller *et al.* 1995). This chapter describes the SPR and SAXS studies performed on the p50 protein to determine the effect of MGBLs on the interaction of p50 with DNA. It also describes the protein expression purification steps involved to obtain the protein necessary for the studies.

6.1. Protein expression and purification

The pMAS 107 expression vector (45ng/ml) was kindly provided by Dr. Darren Hart (EMBL, Grenoble, France) (Figure 6.1.a).

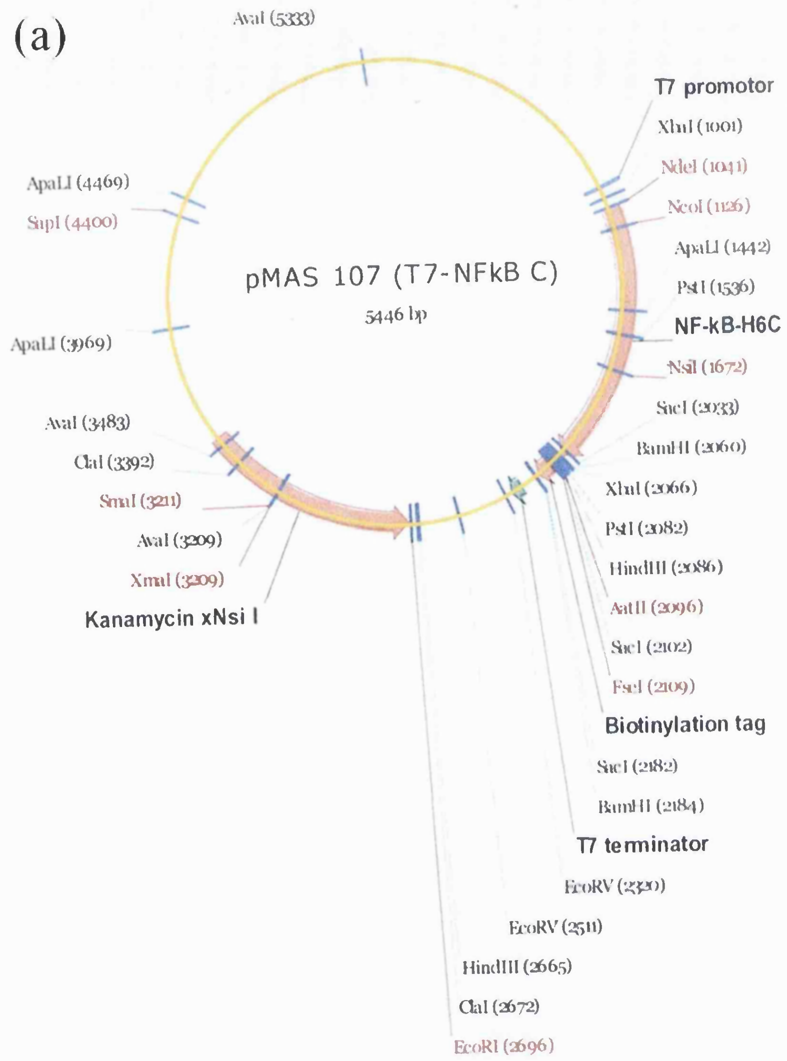
6.1.1. Plasmid preparation

The plasmid was transformed into commercially available One shot® TOP 10 chemically competent *Escherichia coli* (*E. coli*) cells (Invitrogen). The transformation was performed using the protocol provided by Invitrogen. 2µl of plasmid was gently mixed into a vial containing 50µl TOP10 cells and was left in an ice bath for 30 minutes. This mixture was then incubated for 30 seconds in a 42°C heat block and subsequently placed in an ice bath for 2 minutes. Then 450µl of Luria-Bertani medium (LB medium: 10g Bacto tryptone, 10g NaCl and 5g yeast in 1 litre distilled water;

mixture is then autoclaved) was added to the mixture and incubated for 1 hour at 37°C in a shaker at 200rpm. 25μl of this mixture was then spread on a LB-Agar plate containing kanamycin and incubated overnight at 30°C.

A single colony was isolated and transferred to a flask containing 50ml of LB medium with added kanamycin (50μg/ml). This was incubated overnight at 37°C at 200rpm. Cells were centrifuged for 10 minutes at 14,000rpm at 10°C using an ultra-centrifuge. The cell pellet obtained was used for plasmid preparation using the Wizard® *Plus* SV Minipreps DNA Purification System (miniprep kit) supplied by Promega. The cell pellet was resuspended in 250μl cell resuspension solution. Then 250μl of cell lysis solution was added and the Eppendorf tube inverted 4 times to mix, followed by addition of 10μl alkaline protease solution which was mixed by inverting the tube 4 times, and finally incubated for 5 minutes at room temperature. 350μl of neutralisation solution was added and the tube was again inverted 4 times to mix. This solution was then centrifuged for 10 minutes at room temperature at 13,000rpm using a bench top microcentrifuge (Eppendorf centrifuge). The clear supernatant was decanted onto a spin column which was inserted onto a collection tube (provided in the Promega miniprep kit) and then centrifuged for 1 minute at room temperature at 13,000rpm. The flow through was discarded and the spin column was inserted onto another collection tube. Then 750μl of wash solution with ethanol was added to spin column and centrifuged for 1 minute at 13,000rpm and the flow through discarded. This step was repeated with 250μl of wash solution with ethanol and centrifuged at 13,000rpm for 2 minutes. The spin column was transferred onto a sterile Eppendorf tube and 100μl of nuclease free water was added and centrifuged at 13,000rpm for 1 minute at room temperature. The DNA (92.5μg/ml; measured using a UV spectrophotometer at 260nm) was collected in the flow through.

(a)



(b)

```
         10         20         30         40         50         60
MADGPYLQIL EQPKQRGFRF RYVCEGPSHG GLPGASSEKN KKSYPQVKIC NYVGPAKVIV

         70         80         90        100        110        120
QLVTNGKNIH LHAHSLVGKH CEDGICTVTA GPKDMVVGFA NLGILHVTKK KVFETLEARM

        130        140        150        160        170        180
TEACIRGYNP GLLVHPDLAY LQAEGGGDRQ LGDREKELIR QAALQQTKEM DLSVVRLMFT

        190        200        210        220        230        240
AFLPDSTGSF TRRLEPVVSD AIYDSKAPNA SNLKIVRMDR TAGCVTGGEE IYLLCDKVQK

        250        260        270        280        290        300
DDIQIRFYEE EENGGVWEGF GDFSPTDVHR QFAIVFKTPK YKDINITKPA SVFVQLRRKS

        310        320        330
DLETSEPKPF LYYPEIKDKE EVQRKRQKGS SHHHHHHC
```

Figure 6.1. (a) pMAS 107 plasmid vector map with T7 promoter for p50 expression and (b) p50 protein sequence, as provided by Darren Hart's group.

The DNA obtained after the miniprep was tested by running a 1% agarose gel electrophoresis using ethidium bromide dye and 0.5X TBE as running buffer (109g Tris base, 55g boric acid and 4.65g EDTA in 2 litres distilled water). The miniprep DNA (1 µg) was digested using NdeI (restriction enzyme) and 10X NEBuffer 4 (Biolabs). This sample was loaded on the gel before and after incubation in a water bath at 37°C for 1 hour. The mini prep DNA, pMAS 107 plasmid and SmartLadder marker (Eurogentec) was loaded on the agarose gel. The gel was run at 5V/cm and the gel was observed under UV light. A single band was observed for the digested miniprep DNA sample between the 5000-4000 base pair (bp) bands of the SmartLadder marker (Figure 6.2). This confirmed that the plasmid obtained after the miniprep was the same as the vector plasmid (5446bp) provided.

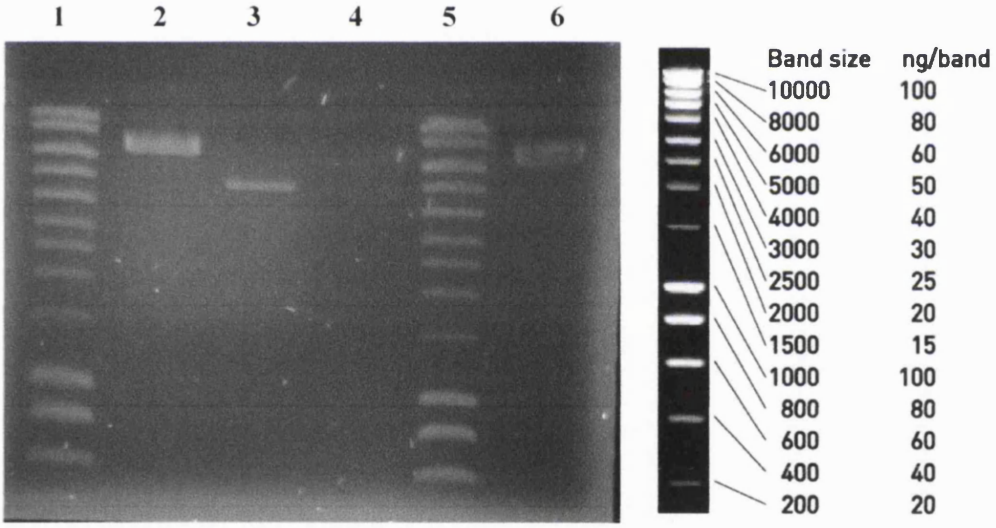


Figure 6.2. 1% agarose gel showing bands observed under UV light. Column (1) and (5) SmartLadder DNA marker, (2) miniprep DNA, (3) miniprep DNA with restriction enzyme incubated, (4) miniprep DNA with restriction enzyme without incubation and (6) pMAS 107 plasmid provided. Bands of SmartLadder DNA marker and the corresponding band size from 10000 to 200 bp.

6.1.2. Transformation

The plasmid obtained after miniprep (92.5µg/ml) was diluted 5 times, and 2µl of this diluted plasmid was transformed into *E. coli* BL21 (DE3) cells provided by Darren Hart's group, by electrophoretic shock. 10X LB media was added to the mixture and incubated for 1 hour at 37°C shaking at 250rpm. 50µl of this sample was spread on LB-agar plate containing kanamycin (50µg/ml) and chloramphenicol (170µg/ml) and incubated overnight at 37°C.

6.1.3. Protein expression

A single colony from the LB-agar plate containing kanamycin and chloramphenicol was isolated and inoculated into 5ml LB with kanamycin (50µg/ml) and chloramphenicol (170µg/ml). This flask was incubated overnight at 37°C with shaking at 150rpm. Then 2ml of this culture was transferred to fresh 30ml LB medium containing kanamycin and chloramphenicol and was incubated at 37°C with shaker at 150rpm for 10 hours. 1ml of this solution was inoculated into fresh 50ml LB medium (with kanamycin and chloramphenicol) and was left overnight for incubation at 30°C with shaking at 200rpm. From this overnight culture, 2ml solution was added to a litre LB medium with kanamycin (50µg/ml) and chloramphenicol (170µg/ml) in two conical flasks. These flasks were left on the shaker at 37°C until the OD_{600} reached 0.8 and then the culture was induced by adding 5ml of 10% w/v arabinose. These flasks were incubated overnight at 25°C with shaker at 200rpm. The cells were then spun down at 14,000rpm for 20mins at 10°C using Beckman Coulter Avanti J-20 XP centrifuge. The 4-5g of cell pellet was collected and stored at -80°C.

6.1.4. Cell lysis

Approximately 1.5g of the cell pellet was suspended in 30ml of resuspension buffer containing 50mM sodium phosphate buffer (pH 7.0), 300mM NaCl, 5mM imidazole (pH 7.0), 20% sucrose along with benzonase and protease inhibitor. The resuspended cells were disrupted at 1.3kBar at 4°C using a constant cell disruption system (Constant Systems). The lysate obtained was centrifuged at 14,000rpm for 40 minutes at 4°C using a JLA 14 rotor on a Beckman Coulter centrifuge. The supernatant was collected and the cell pellet was stored to check for presence of protein using SDS-PAGE gel (Figure 6.3, Gel 1, column 4). The supernatant was filtered and further purified to obtain protein.

6.1.5. SDS-PAGE gel

The samples were prepared using 10μl aliquots of fractions and 5μl of 5X sample buffer (10% SDS, 200mM dithiothreitol (DTT), 24% glycerol, 100mM Tris HCl pH 6.8, 2mg Coomassie G 250) and heated at 100°C for 4 minutes and spun down in a table-top centrifuge (Eppendorf centrifuge 5415D). 12% polyacrylamide gel was prepared and it was made with 6-7cm of running gel followed by the stacking gel. The running gel was prepared with 0.8% acrylamide, 3M Tris HCl pH 8.45, 0.1% SDS, glycerol and APS and TEMED used as the polymerising agent. The stacking gel contained a similar concentration of acrylamide and Tris HCL and 10% SDS and no glycerol. The samples were loaded on to the gel and run at 150V using a Power Pac 1000 (Bio-Rad) for 50-60 minutes. The gel was then stained with staining solution (Coomassie G 250, 10% acetic acid and 50% ethanol) and de-stained in a solution of 20% methanol and 10% acetic acid.

6.1.6. Protein purification

The p50 protein expressed was further purified using a two step purification to ensure the quality and monodispersity of the protein in order to carry out further studies.

6.1.6.1. Nickel column purification

The clear lysate after cell disruption was loaded on to 5ml HiTrap IMAC HP column (GE Healthcare) charged with Ni^{2+} metal ions at 4ml/min. The flow through (FT) was collected to check for presence of protein using SDS-PAGE gel (Figure 6.3; Gel 1, column 2). Following this the column was washed with 10 column volumes of wash buffer containing 50mM sodium phosphate (pH 7.0), 300mM NaCl and 10mM imidazole (pH 7.0). The wash flow through was also collected to check on the SDS-PAGE gel (Figure 6.3; Gel 1 column 3). The column was then loaded with elution buffer containing 50mM sodium phosphate (pH 7.0), 300mM NaCl, 250mM imidazole (pH 7.0) and 2mM 2-mercaptoethanol (BME). The first 40 fractions of 500µl volume each were collected in Eppendorf tubes. Small aliquots of each fraction were loaded on a SDS-PAGE gel to determine when the protein was eluted (Figure 6.3; Gel 1 column 7-13, Gel 2 and Gel 3). Following this, the column was washed with an excess of elution buffer (10 column volumes) which was again collected to check for presence of protein (Figure 6.3; Gel 3, column 8).

Based on the bands observed in the SDS-PAGE gel, fractions 8 to 40 were pooled and dialysed overnight against 5 litres of 50mM sodium phosphate buffer (pH 7), 300mM NaCl and 2mM BME using a dialysis tube of 10kDa molecular weight cut-off. Following this, buffer exchange was performed on the sample by overnight dialysis against 5 litres of 25mM Tris-HCl pH 7.5, 100mM NaCl and 1mM of DTT.

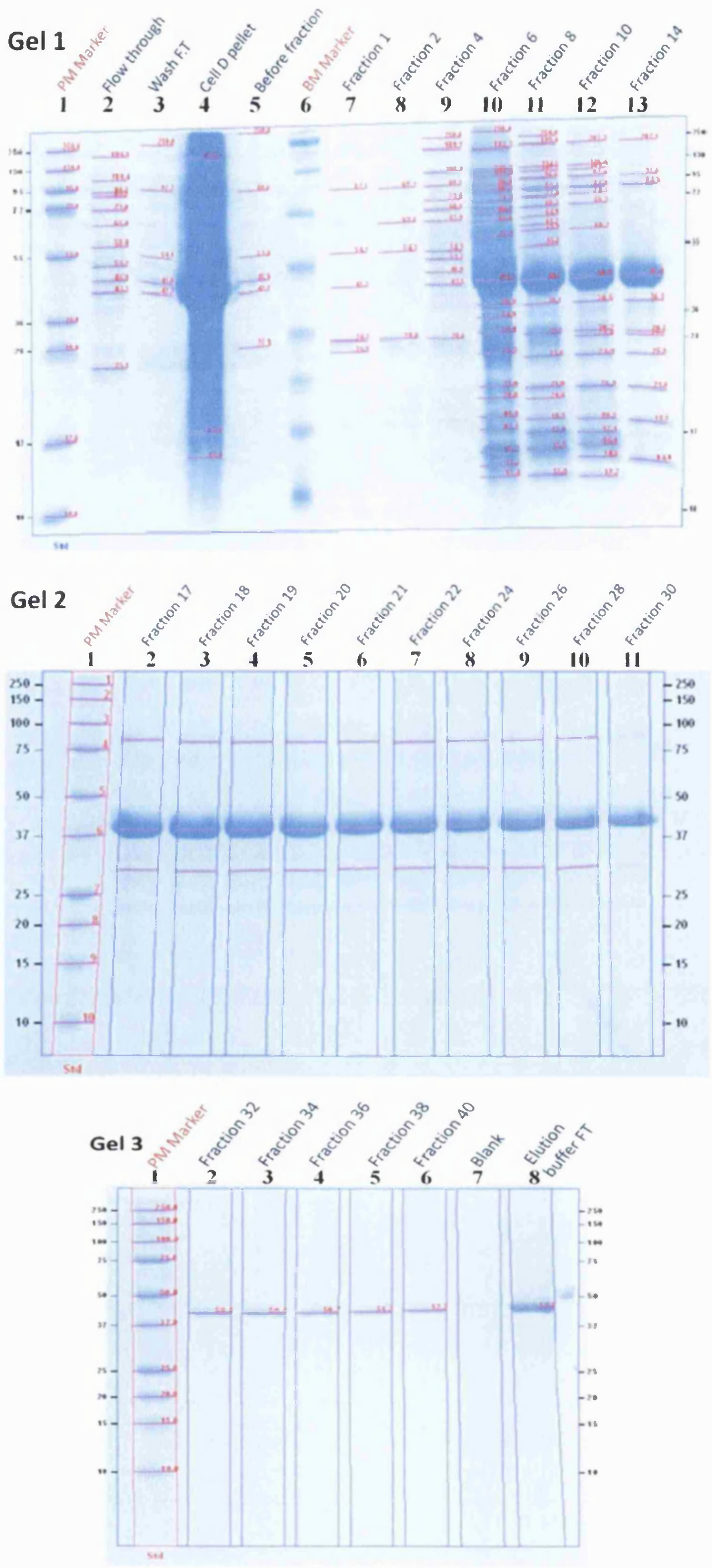


Figure 6.3. 12% SDS-PAGE gel of fraction collected after nickel column purification. The Precision Marker (PM) and Broad Range Marker (BM) from Bio-RAD are indicated in red.

6.1.6.2. Gel filtration

The buffer exchanged protein sample was concentrated to approximately 5ml volume using a centrifuge membrane filter of 10kDa molecular weight cut-off (Millipore) and further filtered using Spin-X centrifuge tube filters (Sigma-Aldrich). The sample was then further purified by gel filtration using a Superdex S200 5/15G column (GE Healthcare) attached to an ÄKTA purifier (GE Healthcare). The column was previously equilibrated with degassed distilled water 0.3ml/min followed by degassed sample buffer containing 25mM Tris-HCl (pH 7.5), 100mM NaCl and 1mM of DTT at 0.5ml/min at room temperature. Then 0.3ml of sample was injected onto the column at 1.5MPa pressure. The eluted protein samples were collected using a Fraction Collector Frac-950 (GE Healthcare) into 96 well plates. The presence of protein and nucleic acid in the sample was detected by UV absorbance at A_{280nm} and A_{254nm}, respectively (Figure 6.4.a). As before, small aliquots of the peak fractions were loaded on a SDS-PAGE gel to determine which fractions were to be pooled (Figure 6.4.b).

The remaining protein sample was loaded onto the Superdex 200 column using a 10ml Superloop (GE Healthcare). Multiple injection runs were set with the injection volume being maintained and fractions collected into 96 well plates both as before. The UV spectrum for the multiple injection runs is shown in Figure 6.4.c. All the samples from single injection and multiple injection runs were pooled together based on the SDS-PAGE gel bands observed for the first injection (Figure 6.4.b).

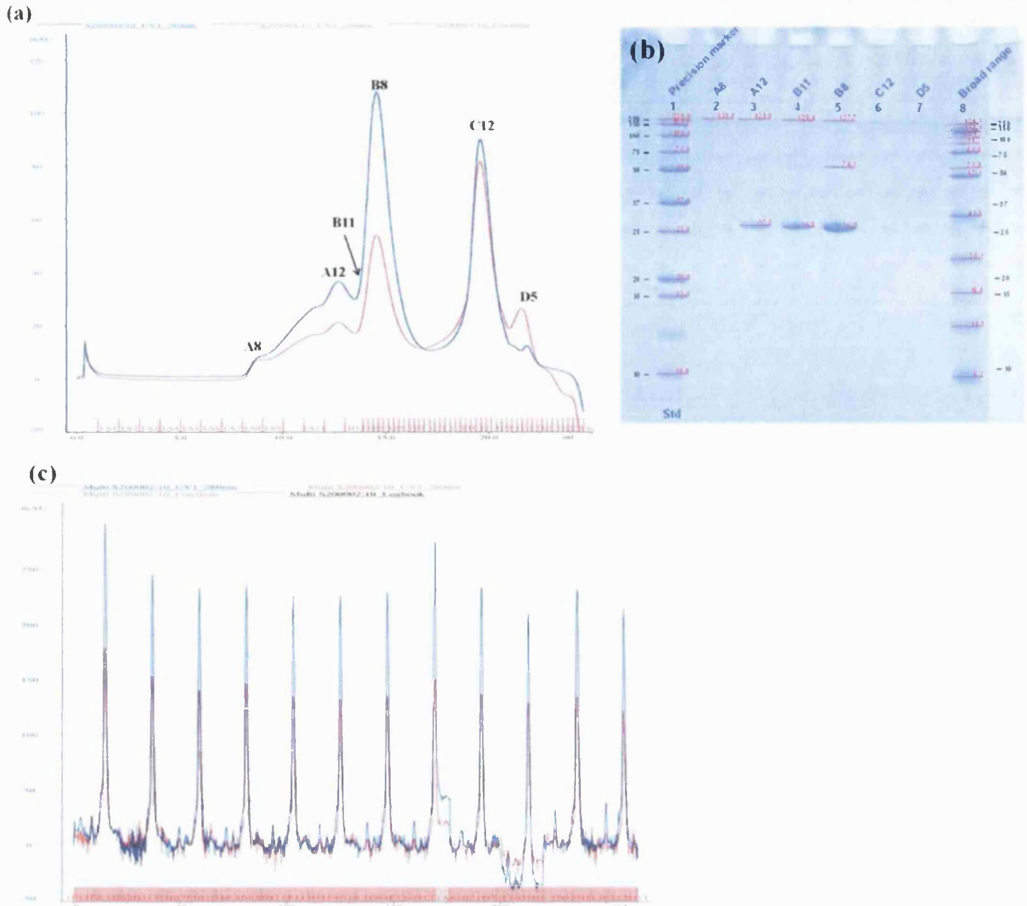


Figure 6.4. (a) UV absorption spectrum of single injection p50 protein after the gel filtration. The spectrum in blue is absorption measured at A_{280nm} for proteins and in red is absorption measured at A_{254nm} for nucleic acids. The peak fractions loaded on a SDS-PAGE gel are indicated in black. (b) 12% SDS-PAGE gel showing the bands of fraction eluted after the first injection of gel filtration on the Superdex 200 column. (c) UV absorption spectrum showing the peaks of the multiple injection run of gel filtration used to purify p50 protein.

6.1.7. Protein analysis

The protein fractions pooled together after gel filtration and the concentration of the protein was measured using a UV spectrophotometer at A_{280nm}; approximately 4ml of protein was obtained of 4.14mg/ml concentration. The protein was further analysed to check the purity of the sample obtained and to confirm the quality of the sample to carry out further experiments.

6.1.7.1. N-terminal sequencing

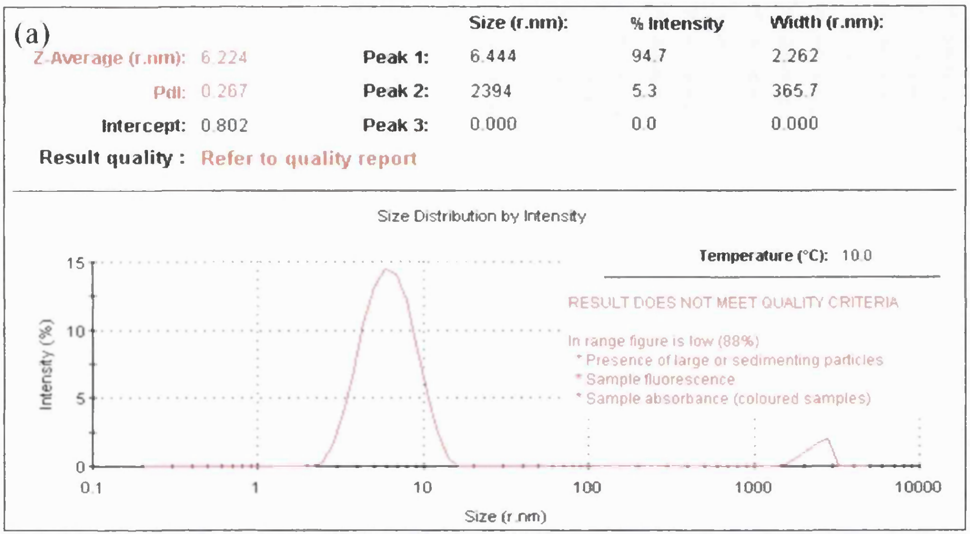
Approximately 30-40pmoles of protein sample was sent to the N-terminal sequencing platform at *Institut de Biologie Structurale* (IBS, Grenoble), one of the partner institutes of the Partnership Structural Biology (PSB, Grenoble), to determine the first 5 amino acid residues at the N-terminal end of the protein. The sequence found was ADGPY, in accordance with the sequence of the N-terminal end of p50 protein from the second residue, confirming that the protein expressed and purified is p50 protein. The p50 protein sequence is given in Figure 6.1.b.

6.1.7.2. Dynamic light scattering (DLS)

This technique was used to determine the quality of the protein sample in terms of its particle size distribution, polydispersity index and aggregation. Approximately 45μl of minimum 0.1mg/ml protein samples was needed for this technique. A Zetasizer Nano instrument (Malvern Ltd.) was used to perform the measurements. The first sample to be tested was the protein collected after the nickel column purification and after the first step of overnight dialysis to remove imidazole as described in section 6.1.6.1. The sample purity at this stage according to the DLS analysis was low (88%). The sample was found to be polydispersive with the presence of large or sedimenting particles (Figure 6.5.a). The protein sample obtained after gel filtration was also tested using DLS. According to the DLS analysis, the protein sample obtained after gel filtration still had a higher polydispersity index (Figure 6.5.b). Hence, the sample obtained after gel filtration was further centrifuged at 100,000rpm using an air driven ultracentrifuge AIRFUGE (Beckman Coulter) at room temperature for 5 minutes. This sample was also tested using DLS (Figure 6.5.c). This sample had lower polydispersity index (PdI), was monodispersive and therefore of sufficient quality to continue with further biophysical experiments.

The protein samples were purified using a two-step purification using a nickel column purification and gel filtration as described in section 6.1.6 followed by centrifugation using an AIRFUGE to obtain samples of sufficient quality for the SPR and SAXS experiments.

Figure 6.5. DLS report of the p50 protein sample (a) after nickel column purification and dialysis. The sample is not monodispersive and the polydispersity index (PdI) is high; (b) after gel filtration and (c) after Airfuge centrifugation.


(a)

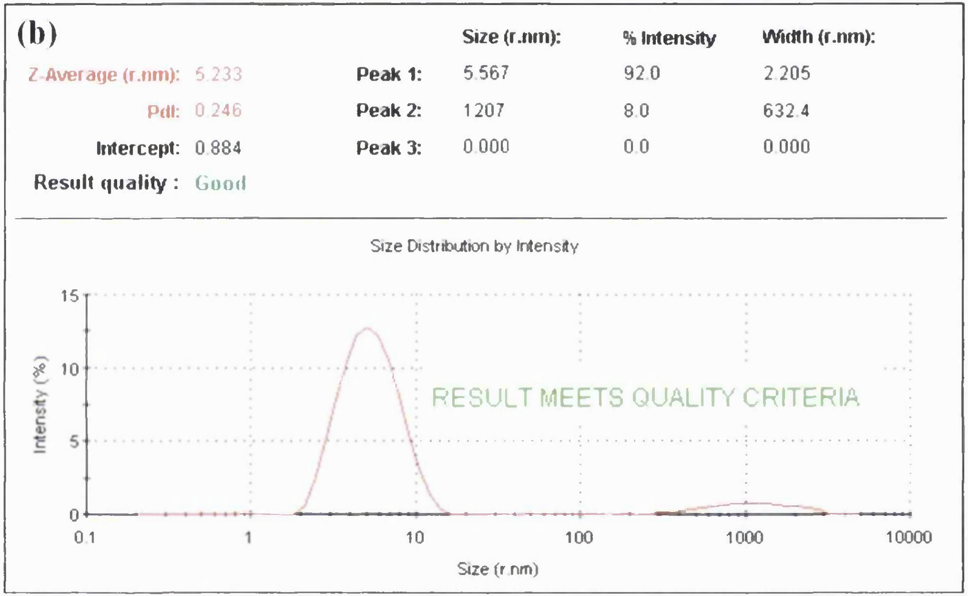
			Size (r.nm):	% Intensity	Width (r.nm):
Z-Average (r.nm):	6.224	Peak 1:	6.444	94.7	2.262
PdI:	0.267	Peak 2:	2394	5.3	365.7
Intercept:	0.802	Peak 3:	0.000	0.0	0.000
Result quality :	Refer to quality report				

Size Distribution by Intensity

Temperature (°C): 10.0

RESULT DOES NOT MEET QUALITY CRITERIA

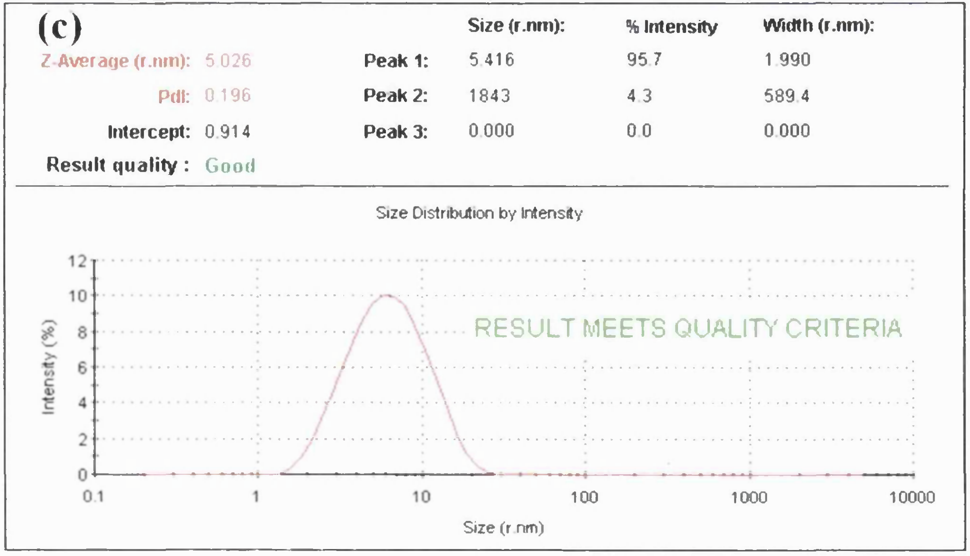
In range figure is low (88%)
- * Presence of large or sedimenting particles
- * Sample fluorescence
- * Sample absorbance (coloured samples)


(b)

			Size (r.nm):	% Intensity	Width (r.nm):
Z-Average (r.nm):	5.233	Peak 1:	5.567	92.0	2.205
PdI:	0.246	Peak 2:	1207	8.0	632.4
Intercept:	0.884	Peak 3:	0.000	0.0	0.000
Result quality :	Good				

Size Distribution by Intensity

RESULT MEETS QUALITY CRITERIA


(c)

			Size (r.nm):	% Intensity	Width (r.nm):
Z-Average (r.nm):	5.026	Peak 1:	5.416	95.7	1.990
PdI:	0.196	Peak 2:	1843	4.3	589.4
Intercept:	0.914	Peak 3:	0.000	0.0	0.000
Result quality :	Good				

Size Distribution by Intensity

RESULT MEETS QUALITY CRITERIA

6.2. Surface Plasmon Resonance (SPR) studies

This study was performed to determine the effect of MGBLs upon the interaction of p50 protein with DNA oligonucleotides. A previous study by Speight *et al.* indicated that distamycin A, a MGBL, affects the binding of p50 protein to DNA in a sequence specific manner, whilst Hoechst 33258, another MGBL, has minimal effect (Speight *et al.* 2002). A similar experiment was used in this work to determine the effect of the curved MGBLs berenil and DB75, and the linear MGBLs DB921 and DB985, and to compare their effects on p50 protein binding to four different DNA sequences.

6.2.1. Materials and methods

6.2.1.1. p50 protein

The protein used for this study was obtained as described in section 6.1. The protein was freshly prepared and stored in an ice bath throughout the experiment to minimise any sample degradation. The protein was present in the p50 buffer containing 25mM Tris-HCl (pH 7.5), 1mM DTT and 100mM NaCl.

6.2.1.2. DNA sequence

Five duplex DNA sequences were studied. Their central regions were designed to have a specific binding site for p50 protein and/or MGBL. The 30 base DNA sequences were obtained from IBA BioTAGnology. The five duplex sequences were made by annealing complementary oligonucleotides; one of each pair being a 5'-biotin labelled DNA sequences. The 5'-biotinylated DNA and the complementary non-biotinylated DNA were annealed together by heating in a water bath at 90°C for 15 minutes and then allowed to cool to room temperature overnight. The biotinylated and non-biotinylated sequences were annealed at a 1:2 ratio to ensure all of the biotinylated DNA sequences were double stranded. The DNA solutions were made up in STE buffer containing 10mM Tris-HCl (pH 7.4), 1mM EDTA and 100mM NaCl. All the sequences used for this study were identical except for the underlined regions:

a) Biotin Control DNA sequence

5'-biotinTEG-AGC TTC AGA GGG GAA ATT CCG AGA GTA CTG-3' and the complementary 5'-GAT CCA GTA CTC TCG GAA TTT CCC CTC TGA-3'. This sequence is also known as PRD*II* and has been studied previously with distamycin and Hoechst 33258 (Speight *et al.* 2002) and was used as a control sequence. The underlined sequence has the continuous guanine and cytosine region which forms the p50 binding site and also at the centre is the A-T rich minor groove ligand binding site with sequence AAATT.

b) Biotin Drug control DNA sequence

5'-biotinTEG-AGC TTC AGA GGG GAA GCT CCG AGA GTA CTG-3' and the complementary 5'-GAT CCA GTA CTC TCG GAG CTT CCC CTC TGA-3'. The underlined region in this sequence is very similar to the biotin control sequence except for the presence of G-C bases between the central A-T region. This sequence lacks the minor groove ligand binding region but still maintains the p50 binding site.

c) Biotin Nature DNA sequence

5'-biotinTEG-AGC TTC AGA TGG GAA TTC CCG AGA GTA CTG-3' and the complementary 5'-GAT CCA GTA CTC TCG GGA ATT CCC ATC TGA-3'. The underlined sequence was used for the crystal study of p50 with DNA; PDB code: 1NFK (Ghosh *et al.* 1995). The resulting structure was published in Nature and hence this sequence is referred to as the Biotin Nature sequence in this study.

d) Biotin A_2T_2 DNA sequence

5'-biotinTEG-AGC TTC AGA CGC GAA TTC GCG AGA GTA CTG-3' and the complementary 5'-GAT CCA GTA CTC TCG CGA ATT CGC GTC TGA-3'. The underlined sequence is the dodecamer sequence $d(CGCGAATTCGCG)_2$ which has been extensively crystallised with several MGBLs.

6.2.1.3. DNA binding ligands

The DNA binding ligands used for the study belong to the iso-helical and the linear groups of MGBLs. The iso-helical ligands studied were berenil and furamidine (DB75) and the linear ligands were DB921 and DB985. All the ligands were provided by collaborators, while berenil was ordered from Sigma Aldrich. The stock solutions of the ligands were made at 10mM concentration with 30mM sodium cacodylate (pH 6.5).

6.2.1.4. SPR assay

A Biacore X (GE Healthcare) instrument was used for the experiments. A sensor chip SA (GE Healthcare) was used to immobilise the biotinylated DNA onto the streptavidin surface of the chip. The protocol followed for the assay was same as described by Speight *et al.* The biotinylated duplex DNA at a concentration of 0.5nM was injected at 10μl/min across flow cell 2 (Fc2) of the chip until 20-25 response units (RU) had been immobilised. One response unit corresponds to a surface density of DNA of approximately 1pg/mm^2 (Stenberg *et al.* 1991). The first flow cell (Fc1) was used as a reference control for non-specific protein binding to the sensor chip matrix, bulk refractive index changes between the injected solution and the running buffer and baseline drift (Hart *et al.* 1999; Speight *et al.* 2002). The running buffer used for DNA immobilisation and later for protein injections contained 10mM Tris-HCl (pH 7.4), 0.2mM EDTA, 3mM DTT, 0.02% v/v Triton X-100, 10% v/v glycerol and 125mM KCl. The buffer was freshly prepared, filtered through a 0.22μm membrane and degassed prior to use.

The experiments were performed to study the effect of MGBLs on p50 and DNA interactions. The first SPR measurement was done in absence of ligand to obtain the binding affinity of p50 to DNA. The purified protein samples were diluted to a range of nanomolar (typically 2.5-20nM) concentrations and each sample injected for 300 seconds at a flow rate of 20μl/min over both Fc1 and Fc2. The protein sample was then replaced by running buffer at the same flow rate and the protein-DNA complex allowed to dissociate for 1000 seconds. During the injections of protein and running buffer (with and without MGBL – see below) sensograms were recorded. The surface of the sensor chip was then regenerated with an injection of running buffer supplemented with 2M NaCl for 45 seconds.

To measure the effect of the MGBLs on the protein-DNA interactions, experiments were performed as above but with the running buffer supplemented with ligand at 10µM concentration. This was done to ensure that the immobilised oligonucleotides were saturated with groove binding ligand prior to addition of the protein. The protein dilutions to be injected were prepared using the running buffer with the DNA binding ligand.

All the SPR experiments were performed at 20°C. The DNA immobilised on the sensor chip was low level (only 20-25RU) and a high flow rate (20µl/min) was used to run the protein sample and running buffer. This was done to minimise rebinding of the released protein to the free immobilised DNA in the dissociation phase (Karlsson & Fält. 1997). Each biotin duplex DNA was immobilised on a different sensor chip. To avoid any errors caused by contamination due to possible irreversible ligand binding to DNA, each MGBL was also studied on a different sensor chip. Since each ligand was measured on a different chip, measurements of protein-DNA interaction were performed on all sensor chips, thereby avoiding any discrepancies due to different RU of immobilised DNA on the sensor chip.

The kinetic analyses of the sensograms were performed using BIAeval 3.0 analysis software based on an algorithm for numerical integration.

6.2.2. Results with biotin control DNA

6.2.2.1. In the presence of iso-helical ligand, DB75

The chip was immobilised with 20µl of 0.5nM duplex biotin control DNA across Fc2 and obtained 22RU. The binding kinetics of p50 to this sequence was measured by injecting several concentrations of p50 over the flow cells (Figure 6.6.a).

The kinetic data (Figure 6.6.a) were fitted globally using the 1:1 Langmuir model using algorithms supplied with the BIAeval 3.0 software analysis package. Global fitting of the data gave $k_a = 6.84 \times 10^5 M^{-1}s^{-1}$ (S.E(k_a) = 6.35×10^3) , $k_d = 2.99 \times 10^{-4}s^{-1}$ (S.E(k_d) = 2.02×10^{-6}), KA = $2.29 \times 10^9 M^{-1}$, KD = 4.37×10^{-10}M and $Chi^2 = 3.4$. The residual plot (Figure 6.6.b) shows that there is a strong difference between the fitted curve and the experimental curves indicating that the model was more or less inappropriate for the experimental data. For this reason the association and dissociation phases of the binding curves were analysed separately using local fitting. The improved residual plot for this fit is shown in Figure 6.6.c and the kinetics of this local fit is shown in Table 6.1. Thereafter both global and separate local fitting models were performed for all analyses.

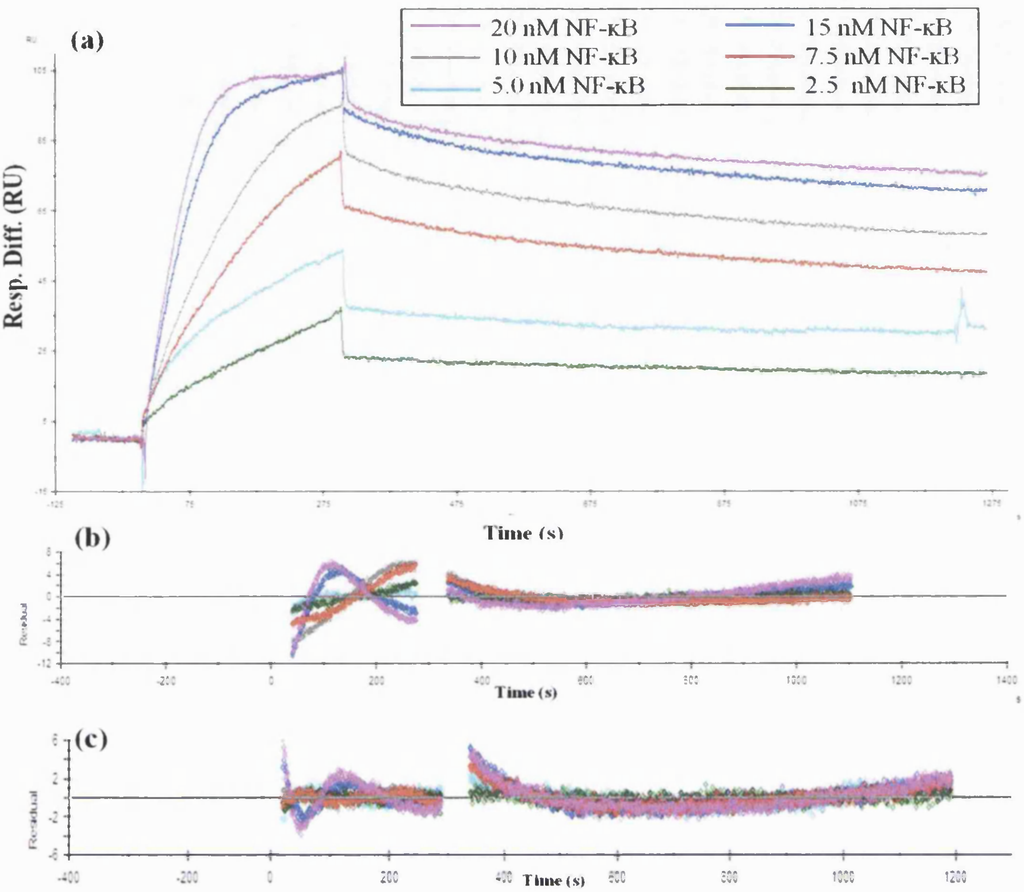


Figure 6.6. (a) Sensograms of 20, 15, 10, 7.5, 5, 2.5 nM p50 binding to 22 RU of immobilised biotin control DNA. Residual plot showing the systematic deviation in the association and dissociation phase of the experimental curve and (b) global fitted curve and (c) separate local fitted curve.

In the next measurement the binding affinity of p50 to this DNA in the presence of the iso-helical ligand DB75 (furamidine) was tested. The immobilised DNA was first saturated with DB75 by flowing the running buffer with DB75 (prepared as described earlier) over the two flow cells, and then the protein samples were injected (Figure 6.7.a). Global fitting of this data gave $k_a = 5.63 \times 10^5 M^{-1}s^{-1}$ ($S.E(k_a) = 7.33 \times 10^3$), $k_d=9.48 \times 10^{-4}s^{-1}$ ($S.E(k_d) = 2.91 \times 10^{-6}$), $KA = 5.94 \times 10^8 M^{-1}$, $KD = 1.68 \times 10^{-9}M$ and $Chi^2=2.61$.

The sensograms obtained for different concentrations of p50 protein binding to DNA in presence of DB75 were compared with the sensograms obtained for protein and DNA interaction and shown in Figure 6.7.b. In the presence of the ligand DB75, there is an increased dissociation of the p50 and DNA (Figure 6.7.b), confirmed by their dissociation constants: in absence of DB75: KD = 4.37 x 10^{-10}M, in presence of DB75: KD = 1.68 x 10^{-9}M. This effect is more pronounced at higher concentration of protein as seen in Figure 6.7.b. and Table 6.1.

Concentration		**k_a**	**k_d**	**KA**	**KD**	**Chi^2**
Protein	**DB75**	**($M^{-1}s^{-1}$)**	**(s^{-1})**	**(M^{-1})**	**(M)**	
20 nM	0 μM	1.09 x 10^{6}	2.04 x 10^{-4}	5.33 x 10^{9}	1.88 x 10^{-10}	0.874
	10 μM	8.69 x 10^{5}	7.9 x 10^{-4}	1.1 x 10^{9}	9.09 x 10^{-10}	0.233
15 nM	0 μM	1.07 x 10^{6}	2.45 x 10^{-4}	4.35 x 10^{9}	2.3 x 10^{-10}	
	10 μM	6.44 x 10^{5}	8.5 x 10^{-4}	7.58 x 10^{8}	1.32 x 10^{-9}	
10 nM	0 μM	5.79 x 10^{5}	3.03 x 10^{-4}	1.91 x 10^{9}	5.23 x 10^{-10}	
	10 μM	4.99 x 10^{5}	7.25 x 10^{-4}	6.88 x 10^{8}	1.45 x 10^{-9}	
7.5 nM	0 μM	4.78 x 10^{5}	3.1 x 10^{-4}	1.54 x 10^{9}	6.48 x 10^{-10}	
	10 μM	7.59 x 10^{5}	8.63 x 10^{-4}	8.8 x 10^{8}	1.14 x 10^{-9}	
5 nM	0 μM	9.79 x 10^{5}	2.07 x 10^{-4}	4.73 x 10^{9}	2.11 x 10^{-10}	
	10 μM	1.26 x 10^{6}	6.06 x 10^{-4}	2.08 x 10^{9}	4.81 x 10^{-10}	
2.5 nM	0 μM	7.16 x 10^{5}	2.27 x 10^{-4}	3.15 x 10^{9}	3.17 x 10^{-10}	

Table 6.1. Kinetic constants (k_a and k_d) and calculated association and dissociation constants (KA and KD) for binding of p50 to biotin control DNA in presence and absence of DB75. Calculated by separate analyses of the association and dissociation phases using local fit.

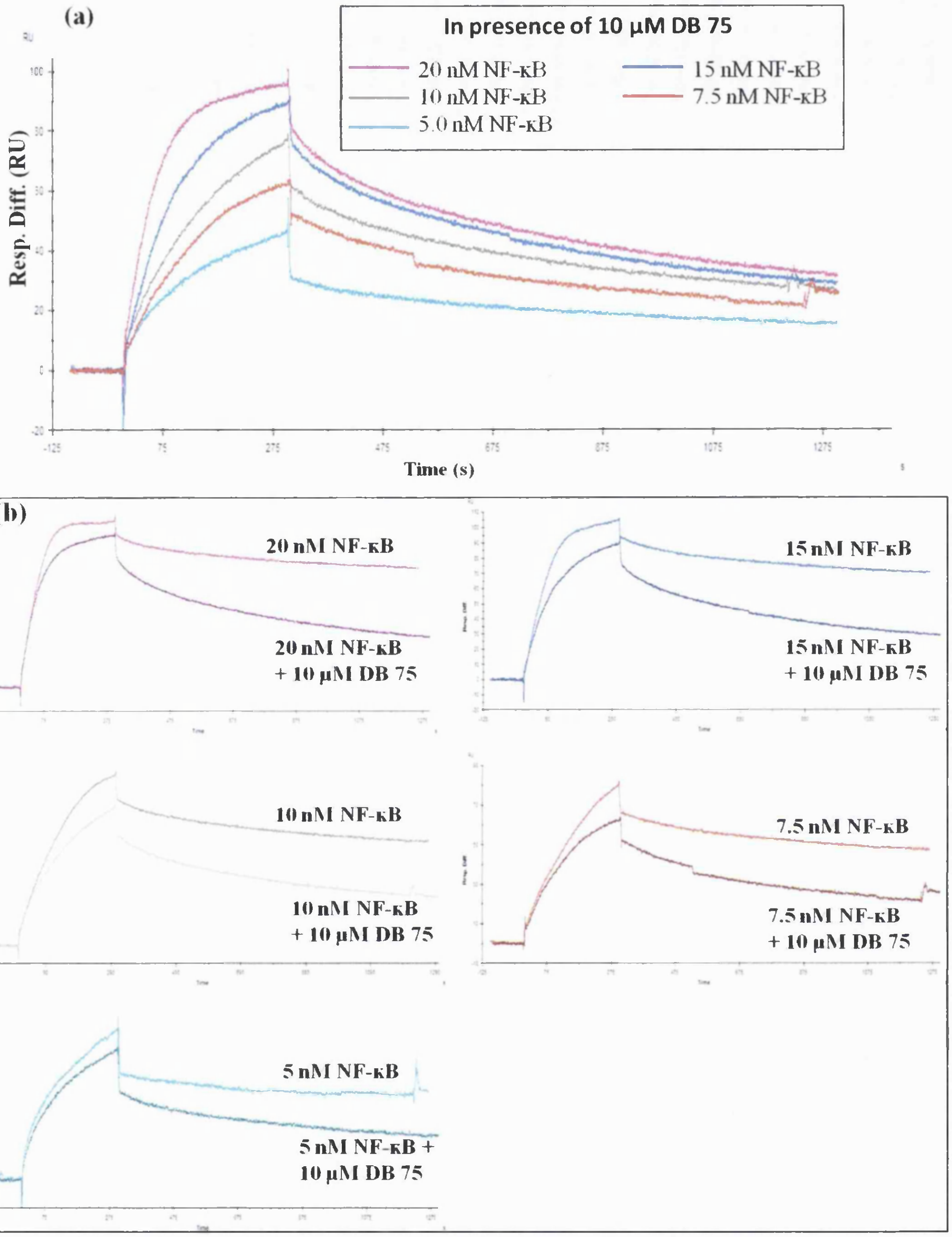


Figure 6.7. (a) Sensograms of 20, 15, 10, 7.5 and 5 nM p50 binding to 22 RU of immobilised biotin control DNA in the presence of iso-helical MGBL DB75. (b) For visual comparison the sensograms are plotted in the presence and absence of iso-helical ligand DB75.

6.2.2.2. In the presence of iso-helical ligand, berenil

The new chip used for berenil was also immobilised with 20µl of 0.5nM biotin control DNA and obtained 24RU.

As before, the binding affinity study of p50 and DNA in the absence of berenil was performed first (Figure 6.8.a). Global fitting of this data gave $k_a = 9.54 \times 10^5 M^{-1}s^{-1}$ ($S.E(k_a) = 8.51 \times 10^3$), $k_d = 4.35 \times 10^{-4} s^{-1}$ ($S.E(k_d) = 2.28 \times 10^{-6}$), $KA = 2.19 \times 10^9 M^{-1}$, $KD = 4.56 \times 10^{-10} M$ and $Chi^2 = 2.41$. The residual plot indicates a good fit of the experimental and fitted curves (Figure 6.8.b). The association and dissociation phases were analysed separately (Table 6.2). The binding affinity of protein to DNA was then studied in presence of berenil (Figure 6.9.a). Global fitting for this data gave $k_a = 4.09 \times 10^5 M^{-1}s^{-1}$ ($S.E(k_a) = 1.3 \times 10^3$), $k_d = 1.62 \times 10^{-3} s^{-1}$ ($S.E(k_d) = 1.41 \times 10^{-3}$), $KA = 2.52 \times 10^8 M^{-1}$, $KD = 3.96 \times 10^{-9} M$ and $Chi^2 = 9.53$. From the sensogram and the kinetic analysis, it can be observed that in presence of berenil both association and dissociation of the protein and DNA is affected (Figure 6.9.b and Table 6.2).

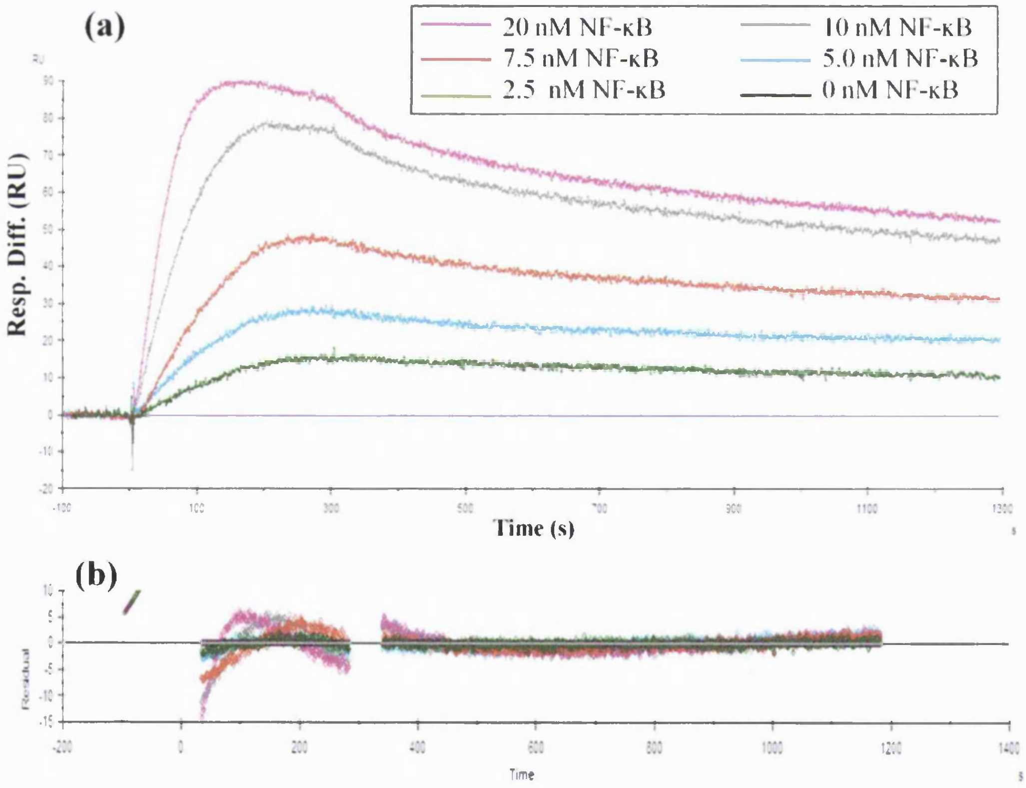


Figure 6.8. (a) Sensograms of 20, 10, 7.5, 5 and 2.5 nM p50 binding to immobilised biotin control DNA (24 RU). (b) Residual plot showing the fit of association and dissociation phase of the fitted curve and the experimental curve.

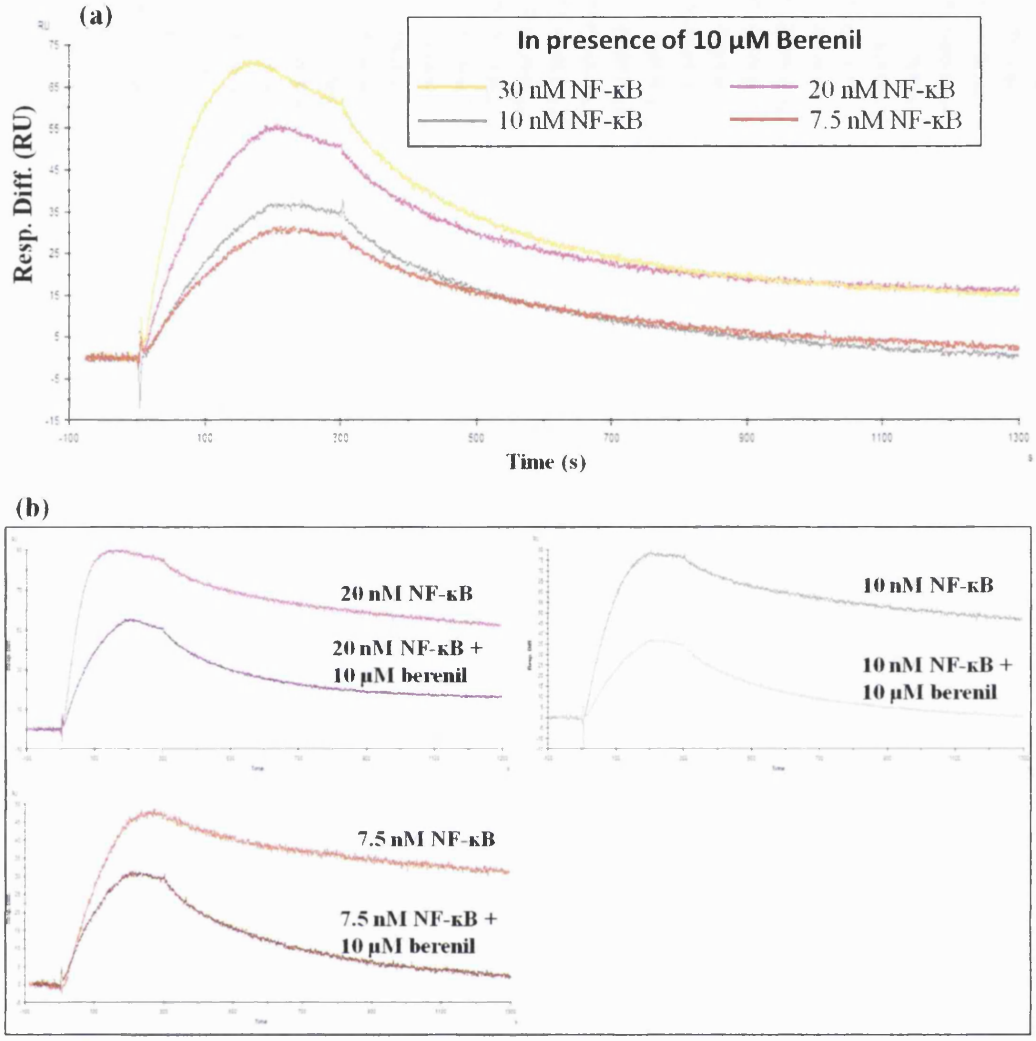


Figure 6.9. (a) Sensograms of 30, 20, 10 and 7.5 nM p50 binding to immobilised biotin control DNA (24 RU) in presence of berenil. (b) Comparing sensograms in presence and absence of berenil.

Concentration		**k_a**	**k_d**	**KA**	**KD**	**Chi^2**
Protein	**Berenil**	**($M^{-1}s^{-1}$)**	**(s^{-1})**	**(M^{-1})**	**(M)**	
30 nM	10 μM	7.87×10^5	1.04×10^{-3}	7.57×10^8	1.32×10^{-9}	3.73
20 nM	0 μM	1.59×10^6	3.86×10^{-4}	4.11×10^9	2.43×10^{-10}	1.42
	10 μM	6.68×10^5	8.31×10^{-4}	8.04×10^8	1.24×10^{-9}	
10 nM	0 μM	1.53×10^6	3.97×10^{-4}	3.85×10^9	2.59×10^{-10}	
	10 μM	8.38×10^5	2.88×10^{-3}	2.91×10^8	3.44×10^{-9}	
7.5 nM	0 μM	1.18×10^6	3.49×10^{-4}	3.38×10^9	2.96×10^{-10}	
	10 μM	1.24×10^6	2.28×10^{-3}	5.43×10^8	1.84×10^{-9}	
5 nM	0 μM	1.72×10^6	2.68×10^{-4}	6.42×10^9	1.56×10^{-10}	
2.5 nM	0 μM	2.56×10^6	3.43×10^{-4}	7.47×10^9	1.34×10^{-10}	

Table 6.2. Kinetic constants (k_a and k_d) and calculated association and dissociation constants (KA and KD) for binding of p50 to biotin control DNA in presence and absence of berenil.

6.2.2.3. In the presence of linear ligand, DB921

On a new chip 40μl of 0.5nM duplex biotin control DNA was immobilised onto the sensor chip and this resulted in 20RU. The binding kinetics of protein and DNA was first analysed for the amount of DNA immobilised on the chip in absences of DB921 (Figure 6.10.a). Global fitting of this data using the 1:1 Langmuir model gave $k_a = 8.85 \times 10^5 M^{-1}s^{-1}$ (S.E(k_a) = 8.88×10^3), $k_d = 6.81 \times 10^{-4} s^{-1}$ (S.E(k_a) = 6.81×10^{-6}), KA = $1.3 \times 10^9 M^{-1}$, KD = 7.7×10^{-10}M and Chi^2 = 2.23. The data were also analysed by separate analyses of the association and dissociation phases using local fit (Table 6.3).

The ligand binding site on immobilised biotin control DNA was saturated with DB921 by flowing running buffer with 10μM DB921 followed by injecting different concentrations of protein. It was observed that in the presence of DB921 there was no significant binding of the protein to DNA (Figure 6.10.b). This was also observed at high concentrations of p50 (40 and 100nM). Since no binding was observed, calculation of rate constants was not possible.

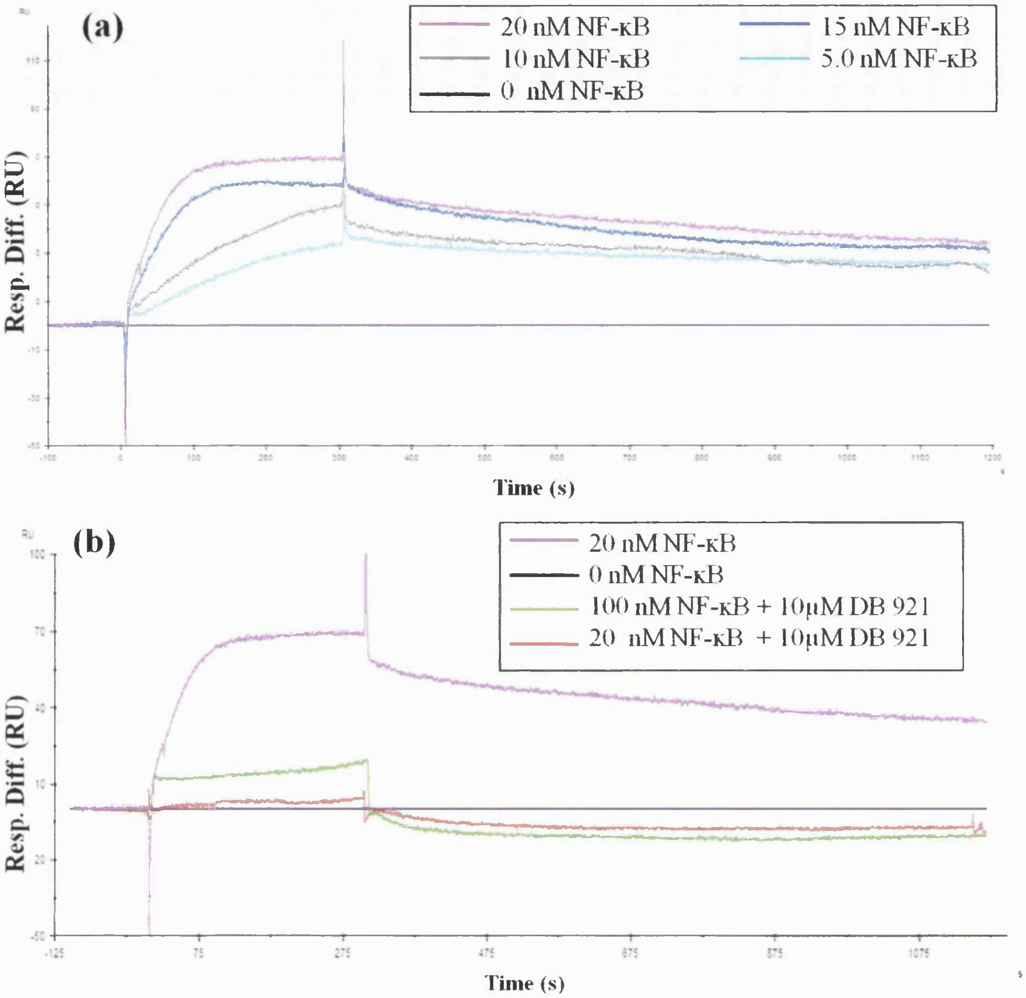


Figure 6.10. (a) Sensograms of 20, 15, 10 and 5 nM p50 binding to immobilised biotin control DNA (19 RU). (b) Sensograms of p50 binding to biotin control DNA in presence of ligand DB921 compared with binding in absence of DB921. In the presence of ligand DB921, there is no binding of protein to DNA.

Protein	**k_a ($M^{-1}s^{-1}$)**	**k_d (s^{-1})**	**KA (M^{-1})**	**KD (M)**	**Chi^2**
20 nM	1.25×10^6	5.13×10^{-4}	2.43×10^9	4.11×10^{-10}	0.571
15 nM	1.42×10^6	6.57×10^{-4}	2.17×10^9	4.61×10^{-10}	
10 nM	3.97×10^5	6.03×10^{-4}	6.58×10^8	1.52×10^{-9}	
5 nM	7.52×10^5	3.63×10^{-4}	2.07×10^9	4.83×10^{-10}	

Table 6.3. Kinetic constants (k_a and k_d) and calculated association and dissociation constants (KA and KD) for binding of p50 to biotin control DNA.

As mentioned earlier, from the thermodynamic studies performed, it is evident that the linear ligand DB921 has a greater DNA binding affinity compared to the other iso-helical ligand studied (Miao *et al.* 2005). This could be the reason for the complete loss of interaction between p50 and DNA in the presence of 10µM DB921 while at same concentration of the iso-helical ligands DB75 and berenil, the protein interacts with DNA but this interaction is diminished but not eliminated by the presence of ligands (Figure 6.7.b and 6.9.b).

Hence, a titration analysis was performed with DB921, where different concentrations of DB921 were added to the running buffer to verify whether the effect of DB921 on protein-DNA interaction was concentration dependent. The old sensor chip could not reproduce a similar RU as observed before in absence of DB921 (Figure 6.10.a) after DB921 had been introduced. The results were not reproducible even after several injections of regeneration buffer (running buffer with 2M NaCl). This suggests some form of irreversible binding of DB921 and hence the titration study was performed on a new sensor chip.

The new chip was immobilised with 40µl of 0.5nM biotin control DNA followed by 30µl of 1nM DNA to obtain 22RU. Several injections of p50 concentration were repeated to obtain a reproducible curve. For this titration experiment, a high concentration of protein (40nM) was chosen. Following this, the concentration of DB921 (20nM, 200nM, 1µM, 5µM, 7.5µM and 10µM) was varied in the running buffer (Figure 6.11). The complex was allowed to dissociate for 500 seconds for this study.

From this titration study it was observed that the change in interaction of protein-DNA in the presence of DB921 is concentration dependent. At low concentration (20nM) of DB921 there is no change in the interaction of protein-DNA. As the concentration of DB921 is increased (200nM, 1µM) there is a small change in association phase and dissociation phase of the curves. As the concentration is further increased (5µM) there is rapid decrease in association and dissociation phase. At the intermediate concentrations (200nM, 1μM and 5μM) of DB921 the changes in interaction of protein-DNA is similar to that observed for berenil at a 10μM concentration. As the concentration of DB921 is further increased to 7.5μM and 10μM, there is rapid dissociation of protein from DNA which results in no binding of protein at the end of the dissociation phase (Figure 6.11).

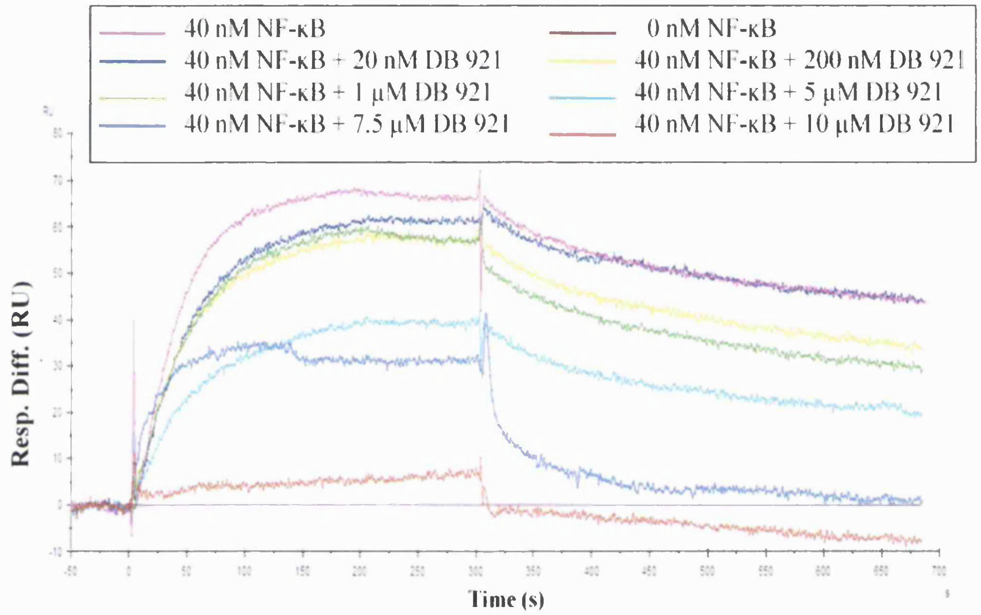


Figure 6.11. Sensograms showing the changes in interaction of p50 and biotin control DNA with increasing concentration of DB921. The binding of p50 to DNA is strongly affected by higher concentration of DB921 (7.5 and 10μM).

6.2.2.4. In the presence of the linear ligand DB985

Following the interesting results for DB921, another linear ligand very similar to DB921 was studied, DB985. Since there was a limited supply of DB985, the effect of interaction of protein and DNA in the presence of this ligand was only done for a small range of protein concentrations. Due to this the full kinetic analysis for this ligand was not possible due to insufficient data. The protein-DNA complex was allowed to dissociate for only 200-500 seconds in presence of the ligand.

Before introducing the ligand, the protein-DNA interaction was studied in the absence of DB985. Following this the immobilised DNA was saturated with ligand and several protein dilutions, prepared with running buffer supplemented with 10μM DB985 was injected to study the protein-DNA interaction . It was observed that in presence of this ligand there was a rapid dissociation of protein bound to DNA while the association of the protein was not affected (Figure 6.12). From visual comparison with the sensograms of other ligands, DB985 has a increased dissociation rate compared to DB75 and berenil (Figure 6.7.b and 6.9.b).

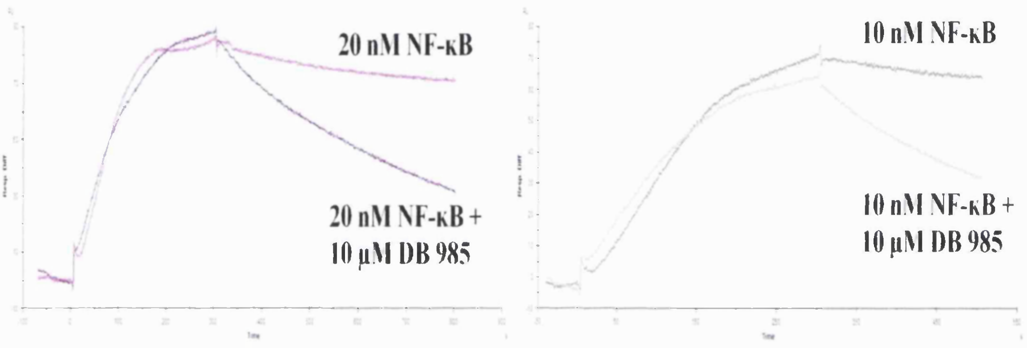


Figure 6.12. Sensograms of association and dissociation phase of 20 and 10 nM of p50 binding to biotin control DNA. The sensograms are plotted in the presence and the absence of linear ligand DB985.

6.2.3. Results with biotin drug control DNA

The sequence 5'-biotinTEG-AGC TTC AGA GGG GAA GCT CCG AGA GTA CTG-3' and the complementary 5'-GAT CCA GTA CTC TCG GAG CTT CCC CTC TGA-3' was chosen to ascertain whether the changes in the protein-DNA interaction in the presence of MGBL as observed earlier (Section 6.2.2) were actually caused by the drug binding to a DNA sequence or due to a non-specific binding of the ligand to the protein or DNA. Hence this DNA sequence was designed with a G-C base pair between the A-T region to disrupt the binding of MGBLs. Hence the absence of the continuous A-T region in this sequence causes it to lack the MGBL site but to nonetheless maintain the p50 binding site. For this study DB75 (iso-helical ligand) and DB921 (linear ligand) were used.

The DNA was immobilised by injecting 0.5nM duplex biotin drug control (BDC) DNA and obtained 22RU and 20RU for the two sensor chips used for DB75 and DB985 binding study respectively. The protein dilutions without ligands were then injected. High non-specific binding of the protein to Fc1 was observed both in presence and absence of MGBLs. Hence the final sensogram (Fc2-Fc1; i.e. Fc2 corrected with subtraction of Fc1) shows a dip at the start of association phase in the protein-DNA binding curve (Figure 6.13.a). It was observed that the presence of 10μM DB75 had no significant effect on protein-DNA interaction (Figure 6.13.a). This confirms that the early observation of DB75 causing increased dissociation of p50 from biotin control DNA (Section 6.2.2.1) was as a result of DB75 binding to the DNA. This effect was abolished when the DNA binding site of the ligand was modified as in BDC DNA. Although in the presence of linear ligand DB921 the binding of p50 to BDC DNA was also inhibited (Figure 6.13.b), this effect was same as seen with biotin control DNA (Section 6.2.2.3) in spite of BDC DNA not having a binding site for MGBL. This indicates that this ligand probably has another mode of binding or interaction to DNA or protein which further causes the inhibition of p50-DNA interaction.

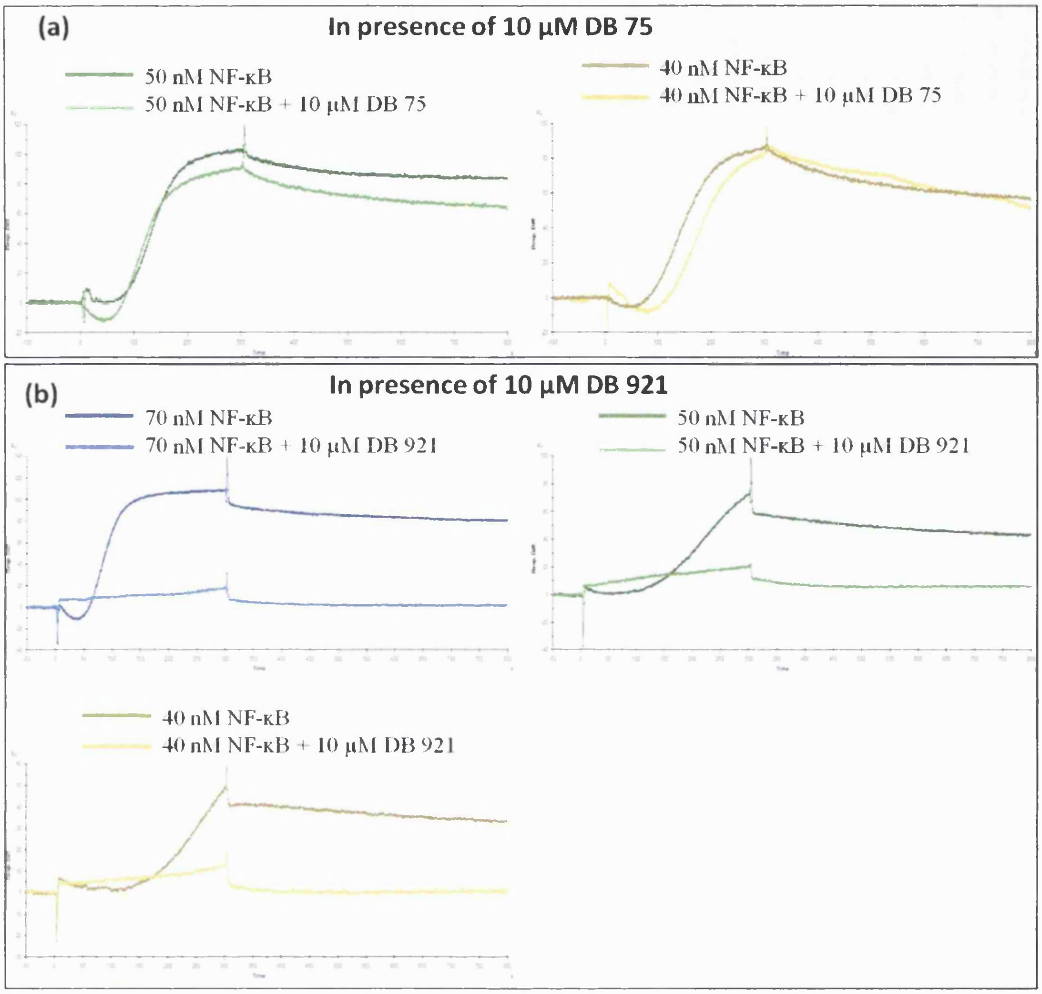


Figure 6.13. Sensograms showing p50 binding to biotin drug control DNA in the presence and absence of (a) iso-helical ligand DB75 and (b) linear ligand DB921. The dip at the beginning of the association phase indicates non-specific binding of protein.

6.2.4. Results with biotin Nature DNA

The biotin Nature DNA was immobilised on the sensor chip by injecting 1nM of duplex DNA which gave 20.8RU. Different concentrations of p50 protein were injected to determine the binding kinetics of protein to this DNA sequence (Figure 6.14.a). The sensograms were analysed using a 1:1 Langmuir model and a global fitting method to obtain $k_a = 2.32 \times 10^3 M^{-1}s^{-1}$ (S.E(k_a) = 43.6), $k_d = 3.36 \times 10^{-4} s^{-1}$ (S.E(k_a) = 3.28×10^{-6}), KA = $6.91 \times 10^6 M^{-1}$, KD = 1.47×10^{-7}M and Chi^2 = 10.7. This data was also globally fitted by separate analysis of association and dissociation phase which gave

$k_a = 2.86 \times 10^4 M^{-1}s^{-1}$ (S.E(k_a) = 1.37×10^3), $k_d = 8.29 \times 10^{-3}s^{-1}$ (S.E(k_d) = 1.01×10^{-4}), KA = $3.45 \times 10^6 M^{-1}$, KD = $2.9 \times 10^{-7}M$ and $Chi^2 = 2.29$. This method of data analysis provided an improved Chi value and also a fit of the experimental and fitted curves. The kinetic study indicates that p50 binds less strongly to this, biotin Nature DNA as compared to the biotin control DNA (Section 6.2.2).

The binding of the protein to this DNA was only tested in presence of the iso-helical ligands DB75 and berenil. The immobilised DNA was saturated with DB75 by flowing running buffer supplemented with 10μM of DB75. Protein samples were then injected and the binding kinetics was studied in presence of DB75. Global fitting of the data resulted in $k_a = 2.28 \times 10^5 M^{-1}s^{-1}$ (S.E(k_a) = 3.76×10^3), $k_d = 1.17 \times 10^{-3}s^{-1}$ (S.E(k_d) = 4.79×10^{-6}), KA = $1.95 \times 10^8 M^{-1}$, KD = $5.13 \times 10^{-9}M$ and $Chi^2 = 6.91$. This data was also analysed using the separate analysis of association and dissociation phase. This gave $k_a = 3.48 \times 10^5 M^{-1}s^{-1}$ (S.E(k_a) = 3.8×10^3), $k_d = 1.38 \times 10^{-7}s^{-1}$ (S.E(k_d) = 1.00×10^{-5}), KA = $2.52 \times 10^{11} M^{-1}$, KD = $3.96 \times 10^{-12}M$ and $Chi^2 = 5.77$. The Chi value obtained using both data analysis techniques were high. Hence in this case, a visual comparison of the sensograms in the presence and absence of the ligand was used to better assess the effect of ligand on protein-DNA interaction (Figure 6.14.b).

Following this, the effect of berenil on protein-DNA interaction was studied in the same way as mentioned for DB75.

The global fitting of the data resulted in $k_a = 1.47 \times 10^5 M^{-1}s^{-1}$ (S.E(k_a) = 3.22×10^3), $k_d = 1.2 \times 10^{-3}s^{-1}$ (S.E(k_d) = 4.35×10^{-6}), KA = $1.22 \times 10^8 M^{-1}$, KD = $8.18 \times 10^{-9}M$ and $Chi^2 = 2.97$. Separate analysis of this data gave $k_a = 2 \times 10^5 M^{-1}s^{-1}$ (S.E(k_a) = 6.3×10^3), $k_d = 2.73 \times 10^{-3}s^{-1}$ (S.E(k_d) = 2.75×10^{-4}), KA = $7.31 \times 10^7 M^{-1}$, KD = $1.37 \times 10^{-8}M$ and $Chi^2 = 2.02$. From the kinetic constants it is evident that berenil affects the dissociation rate of the protein-DNA interaction. This is also seen in Figure 6.14.b.

At low concentrations of protein (40 and 35nM) both DB75 and berenil increased the rate of dissociation of protein from the sequence, the trends being broadly similar in both cases. At higher concentration of protein (50nM), DB75 increases the rate of dissociation while berenil reduces both association and dissociation rate (Figure 6.14.b).

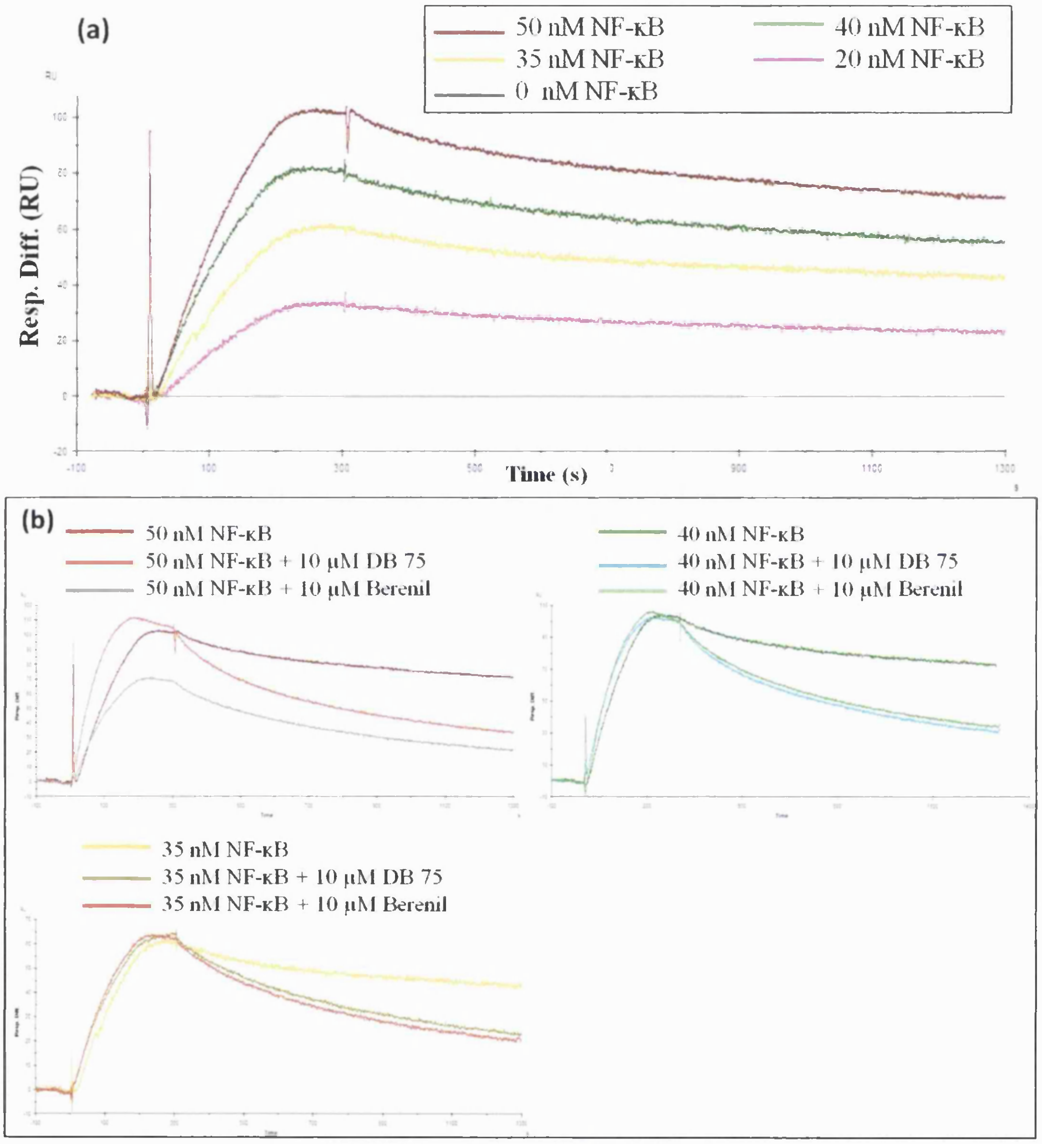
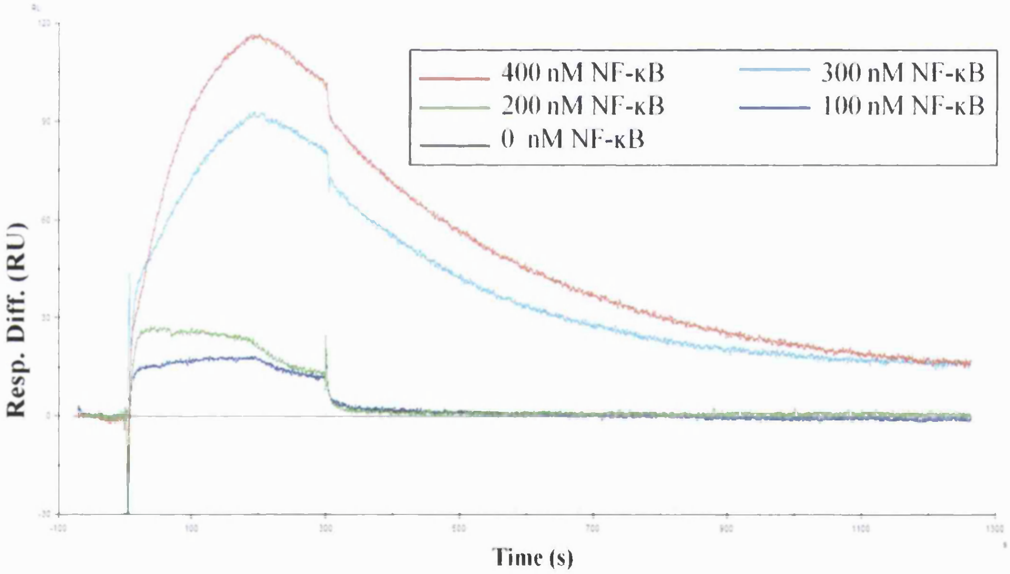


Figure 6.14. (a) Sensograms of 50, 40, 35, 20 and 0nM p50 binding to 21RU of immobilised biotin nature DNA.(b) Sensograms showing the binding of 50, 40 and 35nM of p50 to biotin nature DNA in the presence and absence of iso-helical ligands DB75 and berenil.

6.2.5. Results with Biotin A_2T_2 DNA

As this sequence has been studied extensively with MGBLs for crystallisation as well as for thermodynamic studies, it was chosen to study the effect of protein interaction with DNA.

The DNA was immobilised on the sensor chip by injecting 30µl of 0.5nM of DNA to obtain 24.5RU. Injections of p50 at 100 and 200nM concentrations showed no binding of the protein to DNA. Although for injections at 300 and 400nM concentrations there is a high association RU, the dissociation phase over 1000 seconds with a rapid rate of dissociation leads to negligible effective binding of protein to DNA (Figure 6.15). Due to this insignificant binding, measurements on the interaction of protein-DNA in the presence of MGBLs were not performed.



Fgure 6.15. Sensograms showing binding of 400, 300, 200 and 100nM p50 to biotin A_2T_2 DNA.

6.2.6. Discussion

In the presence of the iso-helical MGBLs DB75 and berenil, there is an increase in dissociation constant of the p50–DNA complex. In the presence of DB75 there is not a significant change in association rate constant of the same complex (k_a=6.84 x $10^5M^{-1}s^{-1}$

in absence of ligand and k_a=5.63 x $10^5M^{-1}s^{-1}$ in the presence of ligand), but the dissociation rate constant increases in the presence of DB75 (k_d=2.99 x $10^{-4}M^{-1}s^{-1}$ in absence of ligand and k_d=9.48 x $10^{-4}M^{-1}s^{-1}$ in presence of ligand). This in turn causes the dissociation constant change from 4.37 x 10^{-10}M to 1.68 x 10^{-9}M.

In the presence of berenil, there is a decrease in association rate constant (k_a=9.54 x $10^5M^{-1}s^{-1}$ in absence of ligand and k_a=4.09 x $10^5M^{-1}s^{-1}$in presence of ligand), an increase in dissociation rate constant (k_d=4.35 x $10^{-4}M^{-1}s^{-1}$ in absence of ligand and k_d=1.62 x $10^{-3}M^{-1}s^{-1}$ in presence of ligand), and changes in the dissociation constant from 4.56 x 10^{-10}M to 3.96 x 10^{-9}M in the presence of berenil.

In the case of both iso-helical ligands there is an increase in dissociation of the protein-DNA complex, while DB75 does not affect the association of the protein-DNA complex, berenil also decreases the association of protein and DNA.

The two linear ligands studied, DB921 and DB985, are structurally similar. However, both these ligands have a different effect on the protein-DNA interaction. DB985 increases the rate of protein-DNA dissociation, more than both iso-helical ligands, while DB921 completely perturbs the protein-DNA interaction at 10μM concentration. DB921 at lower concentrations (5 to 7.5nM) has a similar effect as seen with other minor groove ligands.

These differences in the p50-DNA interaction in presence of different ligands could be due to different ligand-DNA binding affinities. Iso-helical ligands berenil and DB75 have weaker binding constants while the linear ligands DB921 and DB985 have stronger binding to DNA.

The biotin drug control DNA sequence was chosen to determine whether the changes in protein-DNA interaction seen in the presence of MGBLs were due to a particular DNA-MGBL interaction. This sequence was used a negative control since it lacks the necessary groove-binding site. This sequence was tested with DB75 (iso-helical ligand)

and DB921 (linear ligand). In the presence of DB75 there was no difference observed in the case of NF-κB with the biotin drug control DNA sequence, although in the presence of DB921 there is complete loss of binding. This surprising result indicates that DB921 probably has a different mode of binding able to alter the p50-DNA interaction even in the absence of a ligand binding site in the DNA sequence.

The p50-biotin Nature DNA tests showed the protein to have a weaker binding affinity to this sequence in comparison to the biotin control DNA sequence. In the presence of both iso-helical ligands there is an increase in dissociation rate of protein from DNA.

With biotin control and biotin Native DNA sequence it was observed that presence of minor groove ligand causes weakening to p50-DNA interaction. From an earlier gel shift assay it has been shown that p50 and the groove binding ligand Hoechst 33258 can bind simultaneously to the same DNA duplex (Speight *et al.* 2002). It has also been reported that binding of p50 to PRD*II* DNA sequence causes a 25° bend towards the major groove of the DNA, while the PRD*II* sequence inherently has a 20° bend towards the minor groove (Thanos & Maniatis. 1992). This indicates that p50 bends the DNA to maximise DNA interactions. A number of studies have shown that MGBLs bind to the B-conformation of DNA and stabilises the DNA in this form. This conformational change in DNA after binding to protein could be the reason for reduced interaction of protein-DNA in the presence of MGBLs. As suggested earlier by Speight *et al.* it could be possible that the MGBL stabilises the DNA in an alternate conformation which causes a less optimal binding site for protein, leading to less tightly bound p50-DNA complex and increased association rate (Speight *et al.* 2002).

6.3. Small Angle X-ray scattering (SAXS) studies

From the SPR studies performed it became evident that the presence of MGBLs has a significant effect on the interaction between p50 and DNA. The SPR study directly demonstrates the existence of competitive binding between protein and drugs with DNA, although the two have different binding sites on the DNA (Section 6.2). As mentioned earlier, the DNA undergoes conformational change on binding with NF-κB. The changes in the protein-DNA interaction observed by SPR could be due to the inability of DNA to undergo the conformational change required in the presence of MGBLs. However, this remains unproven, with no information from high resolution crystallographic studies on tertiary complexes between MGBLs, DNA and NF-κB. A bio-SAXS experiment was therefore planned to follow, albeit at lower resolution, the changes in the structural envelope of the protein in the presence and absence of DNA and MGBLs.

Since the relative scattering power of the relatively small DNA oligonucleotide is weak compared to the p50 protein, it would be essentially impossible to detect any changes in the DNA structure itself. The aims of the bio-SAXS study were therefore to (a) to establish the shape/molecular envelope of the protein alone and to compare this with that predicted by the available crystal structure, (b) to perform a similar analysis following nucleic acid binding and (c) to follow changes in the protein-DNA complex following the addition of MGBLs.

6.3.1. Materials and methods

6.3.1.1. Sample preparation

The protein samples were prepared as described in section 6.1. To ensure the quality and monodispersity of protein, the samples were freshly prepared and subjected to quality control by DLS measurements prior to the bio-SAXS experiments. The p50 solutions were measured at low concentrations to avoid the aggregation observed previously at high concentrations. The protein samples were prepared in buffer containing 25mM Tris-HCl (pH 7.5), 1mM DTT and 100mM NaCl. The protein samples were concentrated, filtered and AIRFUGE centrifuged just prior to data collection.

The DNA sequence d(TGG GAA TTC CC)$_2$ was used for this experiment. This self-complementary sequence contains both G-C rich protein binding site and A-T rich ligand binding sites. The sequence was annealed by heating in a water bath at 85°C for 15 minutes and then allowed to cool to room temperature overnight. The concentration of the annealed DNA sequence was measured by absorption at A_{260nm}; 4.7mM (31.33mg/ml) was obtained. DNA stock solution was prepared in STE buffer. The DNA MGBLs studied were same as those used for the SPR studies: the iso-helical ligands berenil and DB75, and linear ligands DB921 and DB985. Stock solutions at 10mM of the ligands were made in 30mM sodium cacodylate (pH 6.5). For some experiments higher concentrations of ligands (up to 10mM) were required. This was prepared by adding 5% DMSO to stock solution (20mM ligand solution) to ensure solubility. The p50-DNA-ligand solutions were prepared by first mixing DNA (6.2μM) and ligand (variable concentrations) followed by addition of protein (0.5mg/ml).

6.3.1.2. SAXS data collection

Bio-SAXS data were collected at ESRF, Grenoble using the ID 14-3 beamline (http://www.esrf.eu/UsersAndScience/Experiments/MX/About_our_beamlines/BM29). The sample to detector distance was maintained at 1.0m and the wavelength of the X-rays was 0.93Å. Samples were contained in a 1.8mm diameter quartz capillary which was maintained at 20°C. To limit the extent of radiation damage different areas of the sample in the capillary were exposed to X-rays by flowing the sample in the capillary. The scattering intensities were collected on a Pilatus 1M detector.

Between every sample measurement, the buffer solution was measured to provide a reference for later corrections. For each sample and buffer run, 10 scattering images were collected and averaged by the beamline processing software to obtain a SAXS scattering curve. The averaged buffer background runs were subtracted from the individual sample runs.

Bio-SAXS spectra of the protein samples were measured at three different concentrations (2, 1 and 0.5mg/ml). The lowest concentration of protein (0.5mg/ml) which had no inter-particular association was chosen to measure bio-SAXS spectra of the protein-DNA and protein-DNA-ligand complexes.

The DNA in all the complexes was set at 10% molar excess to the protein. Different ligand concentrations were used between 1µM to 10mM. All solutions of the complex were freshly prepared before data collection and maintained in an ice bath till data collection at 20°C.

6.3.1.3. SAXS data analysis

The instrument was calibrated before data collection on sample using approximately 5mg/ml of bovine albumin serum (BSA) and water. The forward intensity at zero angle for BSA was $I(0)$ = 14.36, MM ~79kDa and for water; $I(0)$ = 3.55. The scattering curves were analysed using the following programmes:

a) **PRIMUS and GNOM**

The background was subtracted using the PRIMUS (Konarev *et al.* 2003) followed by determination of Guinier region to calculate the radius of gyration (R_g) and forward intensity at zero angle ($I(0)$). GNOM (Semenyuk & Svergun. 1991) was used to obtain pair distance distribution functions (p(R)) and R_g. The maximum dimension (D_{max})

value was manually adjusted to obtain R_g values from GNOM in agreement with the R_g values from the Guinier analysis.

b) Ab initio modelling

DAMMIF (Franke & Svergun. 2009) in fast mode was performed on GNOM output file to build *ab initio* molecular envelopes.

c) Rigid body fitting

CRYSOL was used to rigid body fit SAXS scattering curves with a simulated scattering curve generated using the known X-ray crystal structure of the p50 and p50-DNA complex (Svergun *et al.* 1995).

d) Multi-component fitting

OLIGOMER (Svergun *et al.* 1994-2004) was used to determine the volume fraction of different components in the SAXS curves obtained. The form factor containing the scattering intensities of each components were created using FFMAKER. For the OLIGOMER analysis of the p50 alone scattering curve, the form factor included - monomer and dimer PDB coordinates. OLIGOMER analysis of p50-DNA complex scattering curve, the form factor included three components – p50-DNA (PDB 1NFK), p50 alone (SAXS scattering curve was used) and DNA (SAXS scattering curve of DNA alone). The p50-DNA-ligand scattering curves analysis included two components in the form factor – p50 alone (SAXS scattering curve of protein alone) and p50-DNA (PDB code: 1NFK).

6.3.2. SAXS results on p50 alone and on the p50-DNA complex

a) PRIMUS and GNOM analysis. Analysis of the scattering curves for the three different concentrations of protein (Figure 6.16.a and 6.16.b) showed the R_g values had no significant dependence on protein concentration and were all 3.81 nm within ± 0.05

nm. This indicates that the protein samples were highly monodispersed with insignificant aggregation or intermolecular interaction in this concentration range. The Guinier and GNOM analysis results for the protein alone at low concentration (Figure 6.17) and p50-DNA complex are listed Table 6.4.

The results show that the p50-DNA complex has a smaller size, indicated by the lower R_g and D_{max} values of the complex, compared to the protein alone. The molar mass (MM) of the p50-DNA complex is larger than protein alone but the Porod volume (V_p) is smaller for the complex. This analysis therefore confirms that the p50-DNA complex has a more compact structure compared to protein alone. This is also confirmed by the differences in the shapes of the pair distance distribution, *p*(R), function of protein alone and protein-DNA complex (Figure 6.18).

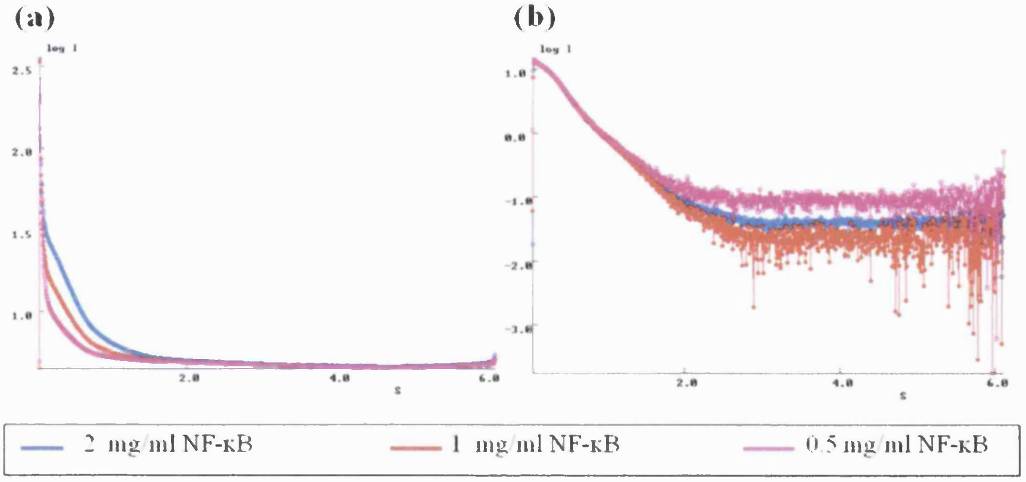


Figure 6.16. SAXS scattering curves of p50 protein alone, Intensity (log I) v/s q (Å^{-1}) recorded at 2, 1, 0.5mg/ml. (a) Experimental curves and (b) Subtracted curves which was further used for Guinier analysis.

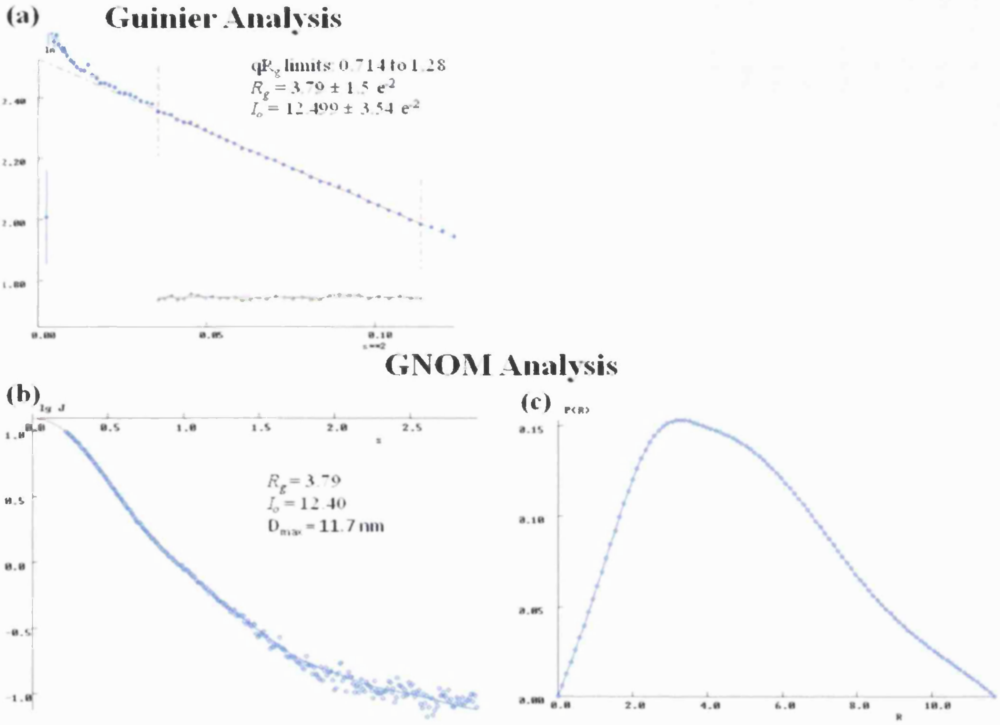


Figure 6.17. SAXS analysis of 0.5mg/ml of p50 sample using (a) Guinier and GNOM analyses. The GNOM analysis provides (b) experimental and calculated curves and (c) *p*(R) function.

Sample	Concentration		R_g (nm)	$I(0)$	MM (kDa)	V_p (nm^3)	D_{max} (nm)
	Protein	DNA					
p50 alone	0.5mg/ml	0μM	3.79	12.13	67	132	11.7
p50-DNA	0.5mg/ml	6.21μM	3.44	13.30	73	112	10.8

Table 6.4. PRIMUS and GNOM parameters of the p50 alone and p50-DNA scattering curves.

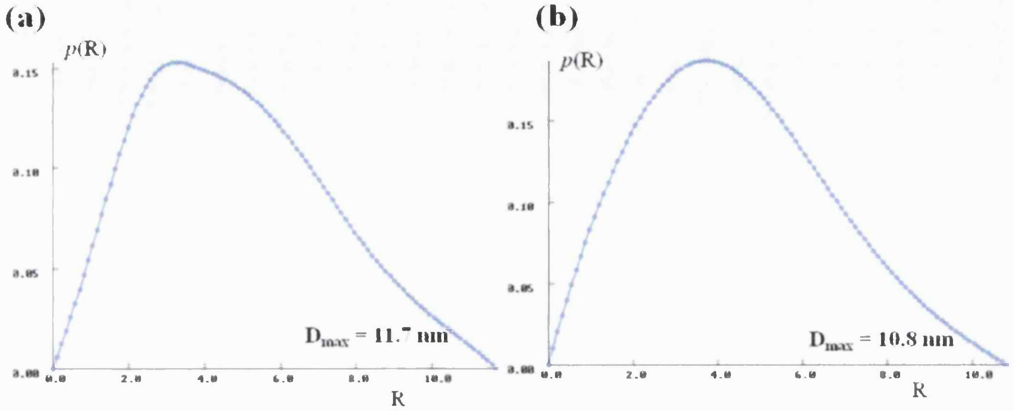


Figure 6.18. *p*(R) function of (a) p50 protein alone and (b) p50–DNA complex.

b) *Ab initio* modelling

DAMMIF modelling of p50 protein alone was comparatively difficult which could be due to high flexibility of the protein observed in solution or due to the presence of a mixture of p50 monomer and dimer species in solution. In the presence of DNA, models were obtained more easily. The *ab initio* models obtained for both protein alone and protein-DNA complex were compared to the X-ray crystal structure of p50-DNA (PDB code: 1NFK; resolution: 2.30 Å) as shown in Figure 6.19. From the *ab initio* models generated for the protein alone and protein-DNA complex, it can be seen that the p50-DNA complex model is more compact and closely resembles the X-ray crystal structure, although some parts of the crystal structure are not visible in the SAXS model of p50-DNA complex (Figure 6.19.b).

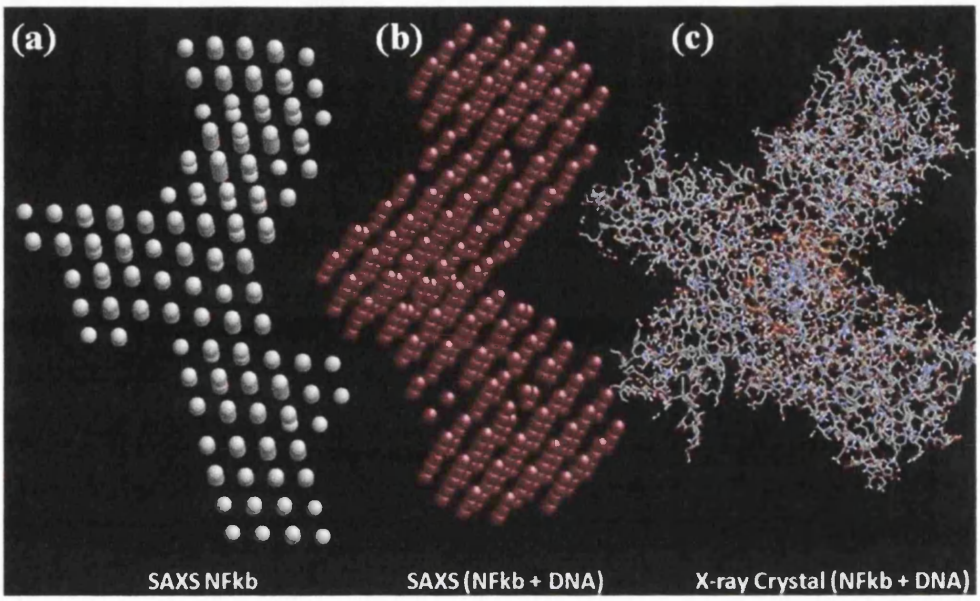


Figure 6.19. SAXS *ab initio* model generated by using DAMMIF of (a) p50 protein alone and (b) p50 and DNA complex. These models are compared with (c) X-ray crystal structure of p50 bound to DNA (PDB code: 1NFK).

c) Rigid body fitting

Several CRYSOL analyses were performed to fit the p50 alone SAXS scattering curve. These analyses are listed below-

- **CRYSOL fit with shorter p50 crystal structure (PDB code 1BFS):** The p50 protein used for SAXS analysis had 338 amino acids. However, crystal structure of p50 alone was only available for short p50 construct (106 to 115 amino acids). The SAXS scattering curve of p50 alone was fitted with 1BFS crystal structure (106 amino acid residue), the CRYSOL fit was poor with a Chi^2 value of 6.4 (Figure 6.20.a).
- **CRYSOL fit with p65 protein crystal structure (PDB code 1NFI)**: p65 homodimer, belonging to the NF-κB family has 40% structural homology to p50 protein. The CRYSOL fit with this crystal structure (306 amino acid residue) also has a poor fit with a Chi^2 value of 5.1 (Figure 6.20.b).

- **CRYSOL fit with p50-DNA complex crystal structure (PDB code 1NFK):** From the *ab initio* modelling study, it was observed that the model obtained for p50 alone scattering curve was comparable to this structure. From CRYSOL analysis it was observed that the p50 alone SAXS scattering curve had an acceptable fit with a Chi^2 value to 3.8 (Figure 6.20.c).
- **CRYSOL fit with monomer and dimer structure:** The monomer and dimer coordinates for this CRYSOL analysis were extracted from 1NFK PDB coordinates by deleting the DNA molecule for CRYSOL analysis as crystal structure was not available. The CRYSOL fit with monomer coordinate has a poor fit with a n/a Chi^2 value (Figure 6.20.d), whilst the fit with dimer coordinate had a good fit with Chi2 value of 2.8 (Figure 6.20.e).

The experimental SAXS scattering curve of p50-DNA complex was CRSYOL fitted against the p50-DNA crystal structure (1NFK) and it had a good fit with Chi^2 value of 2.6 (Figure 6.20.f).

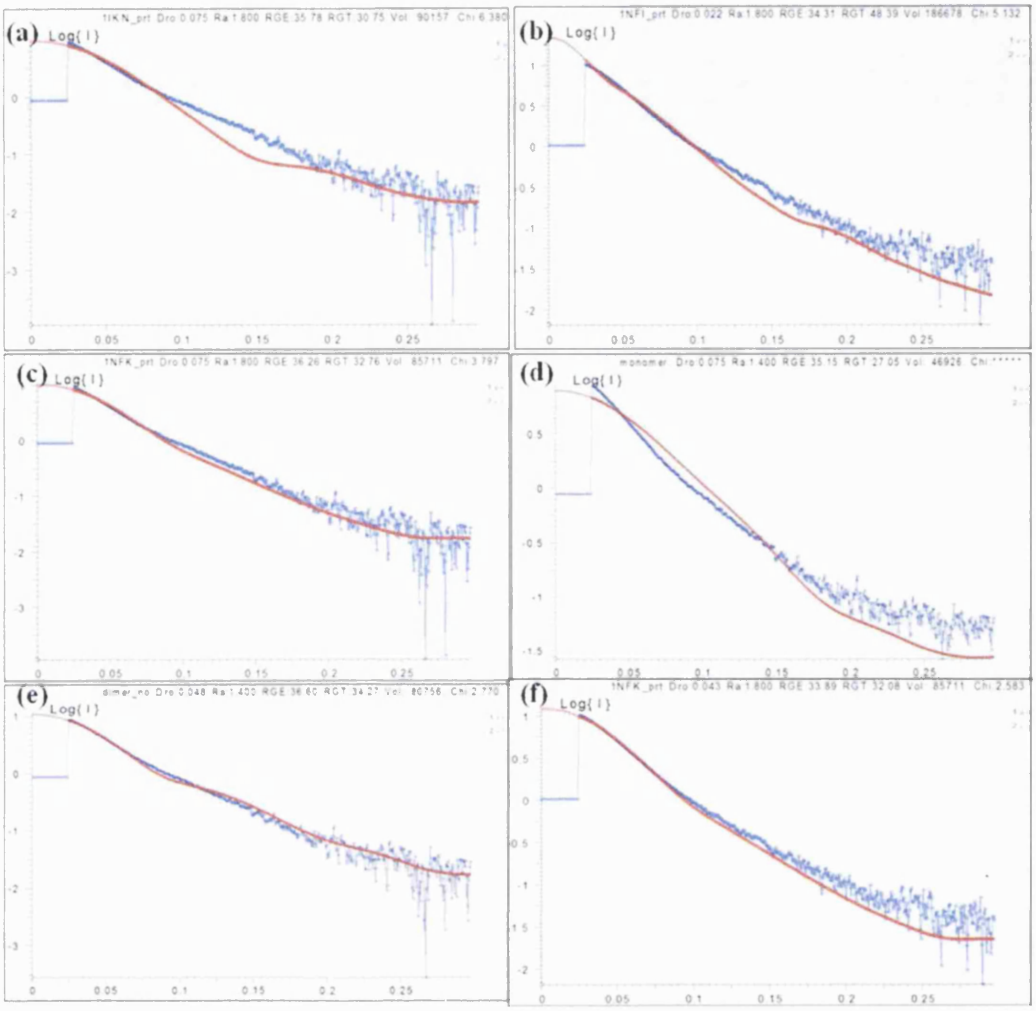


Figure 6.20. CRYSOL fit of SAXS scattering curve of p50 alone with X-ray crystal structure of (a) short p50 protein construct (PDB 1BFS), (b) p65 protein alone (PDB code: 1NFI), (c) p50-DNA (PDB code: 1NFK), (d) p50 monomer coordinates and (e) p50 dimer coordinates. (f) CRSOL fit of SAXS scattering curve of p50-DNA complex with p50-DNA crystal structure (1NFK).The experimental SAXS scattering curves are shown in blue and the crystal structures are shown in red.

d) Multi-component fitting

Whilst the protein elutes as a monomer during the protein purification step (see Section 6.1.6), p50 exists naturally as a dimer and could be possibly present as a mixture of both monomer and dimer species in solution. The OLIGOMER analysis on SAXS scattering curve of protein alone showed that this solution contained a mixture of monomer (7.8%) and dimer (92.2%). The SAXS scattering curve of the p50-DNA complex was also analysed by OLIGOMER and it showed 71% of protein-DNA complex, 0% protein alone and 29% excess DNA.

6.3.3. SAXS results in the presence of minor groove binding ligands

The p50-DNA-ligand scattering curves were analysed using PRIMUS and GNOM although this analysis was not possible at higher concentration of ligands as the components became aggregated. The scattering curves were further assessed using OLIGOMER. The OLIGOMER analysis was a helpful tool to determine the changes in the solution in presence of increasing concentration of ligands. OLIGOMER analysis was performed using a form factor containing two components: protein alone and the protein-DNA complex. The form factors were created using (a) the SAXS scattering curve of protein alone (as the OLIGOMER analyses indicated that the protein was present as a mixture of monomer and dimer) and (b) the X-ray crystal structure of the p50-DNA complex (PDB code 1NFK).

6.3.3.1. In the presence of iso-helical ligand, berenil

a) PRIMUS and GNOM analysis. Several concentrations of berenil ranging between 10mM and 0.5μM were tested. Protein and DNA concentrations were kept constant. The PRIMUS and GNOM analyses of the most interesting berenil concentrations are shown in Table 6.5. However, it was observed that in presence of MGBLs the quality of the SAXS curves were poor which could be due to the poor polydispersed ligand solutions. From this analysis it is apparent that with changes in berenil concentration there are significant changes in the R_g, $I(0)$ and MM values.

b) Multi-component fitting. The changes observed with PRIMUS and GNOM analysis have an impact upon the OLIGOMER analysis. To verify if the changes observed in the protein-DNA solution in the presence of berenil were actually caused by berenil, scattering curves were measured for the protein-berenil solution and berenil alone. These scattering curves were only analysed using OLIGOMER. The results of this OLIGOMER analysis of the protein-DNA-berenil solution and protein-berenil solution are listed in Table 6.6.

Berenil concentration	R_g (nm)	$I(0)$	D_{max} (nm)	MM (kDa)	V_p (nm^3)	Quality (%)
10mM	3.83	10.90	13.4	49	38	74
2.0mM	4.13	12.13	14.5	55	171	37 (low)
1.6mM	3.94	11.98	13.8	54	109	12 (low)
1.4mM	4.12	12.86	13.5	58	167	18 (low)
1.0mM	3.75	13.31	12.3	60	149	48 (low)
1.0μM	3.74	14.18	13.1	64	151	94

Table 6.5. PRIMUS and GNOM parameters of the p50 protein and DNA in presence of Berenil.

Berenil conc (mM)	Volume fraction of each component (%)		
	Protein	Protein + DNA	Chi^2
	Protein - DNA- Berenil		
2	100	0	2.4
1.6	85	15	1.9
1.4	57	43	2.6
1	25	75	3.0
0.01	0	100	3.4
	Protein – Berenil		
2	97	3	-
1.6	100	0	-
1.4	100	0	-
1	97	3	-
0.25	95	5	-

Table 6.6. OLIGOMER analyses of SAXS scattering curves measured for p50-DNA-berenil and p50-berenil at varying concentrations of berenil.

The OLIGOMER analysis clearly demonstrates that with an increasing concentration of berenil, the volume fraction of protein increases while the protein-DNA volume fraction decreases. This confirms that the protein-DNA complex dissociates as the concentration of berenil increases. While the OLIGOMER analysis on the p50-berenil scattering curve shown no significant change in volume fraction with changing berenil concentration. This indicates that the change in p50-DNA complex fraction is caused due to interaction of berenil with DNA and it has no significant effect on p50.

6.3.3.2. In the presence of isohelical ligand, DB75

Similar to the study performed on berenil, the SAXS scattering curves of p50-DNA-DB75 and p50-DB75 were collected at different concentrations of DB75 ranging from

1mM to 10μM. The results for PRIMUS and GNOM analyses (Table 6.7) and OLIGOMER analysis (Table 6.8) of these scattering curves are shown.

DB75 concentration	R_g (nm)	$I(0)$	D_{max} (nm)	MM (kDa)	V_p (nm^3)	Quality (%)
1mM	5.51	16.59	19.3	75	346	4 (low)
500μM	5.37	16.08	18.4	73	278	11 (low)
300 μM	4.03	14.41	13.8	65	170	40 (low)
100 μM	3.75	13.60	12.6	61	148	89
10 μM	3.45	12.98	12.1	59	138	81

Table 6.7. PRIMUS and GNOM parameters of the p50 protein and DNA in presence of DB75.

DB75 conc (μM)	Volume fraction of each component (%)		
	Protein	Protein + DNA	Chi^2
	Protein - DNA- DB75		
1000	100	0	4.0
500	100	0	2.6
300	50	50	1.6
100	15	85	1.0
10	0	100	1.9
	Protein – DB75		
1000	100	0	0.9
500	100	0	2.1
300	89	11	1.2
10	49	51	1.8

Table 6.8. OLIGOMER analysis of SAXS scattering curve of p50-DNA-DB75 and p50-DB75 at varying concentration of DB75.

From the PRIMUS and GNOM analysis it is observed that R_g, $I(0)$, MM, V_p, D_{max} values decrease with decreasing concentration of DB75. At high concentration of DB75 the quality of the curves was low, which could be due to presence of aggregates or large mixtures. The OLIGOMER analyses on the protein-DNA-DB75 scattering curve follows the same trend as observed in presence of berenil. The volume fraction of protein-DNA decreases with increasing concentration of DB75.

6.3.3.3. In the presence of linear ligand, DB985

The scattering curves of protein-DNA-DB985 and protein-DB985 were collected, where the protein and DNA concentration were stable and ligand concentration was varied in the range of 500μM to 10μM. The PRIMUS and GNOM analysis results are listed in Table 6.9. At high concentration of this ligand the scattering curves could not be processed satisfactorily. Hence only the analyses performed at low concentrations are listed. OLIGOMER analysis results are listed in Table 6.10.

From the OLIGOMER analysis of p50-DNA-DB985, it is observed that with increasing concentration of DB985 the volume fraction of protein-DNA complex decreases in solution. The OLIGOMER analyses on the protein-DB985 curves reveal no significant change in the volume fraction of any component at varying concentration of DB985. This confirms that the changes in the protein-DNA interaction in solution occurring at high concentration of ligand are caused by DNA-DB985 interaction and that DB985 has no effect on the protein in absence of DNA.

DB985 concentration	R_g (nm)	$I(0)$	D_{max} (nm)	MM (kDa)	V_p (nm^3)	Quality (%)
40μM	4.15	18.07	13.6	93	120	41
30 μM	3.84	16.64	13.4	86	120	62
20 μM	4.28	17.13	13.4	90	173	78
10 μM	3.70	14.28	13	65	150	90

Table 6.9. PRIMUS and GNOM parameters of the p50 protein and DNA in the presence of DB985.

DB985 conc (μM)	Volume fraction of each component (%)		
	Protein	Protein + DNA	Chi^2
	Protein - DNA- DB985		
70	100	0	4.8
50	93	7	3.5
40	48	52	-
30	45	55	-
20	30	70	-
10	0	100	0.9
	Protein – DB985		
70	100	0	-
50	100	0	-
40	100	0	-
30	100	0	-
20	100	0	-

Table 6.10. OLIGOMER analysis of SAXS scattering curve of p50-DNA-DB985 and p50-DB985 at varying concentrations of DB985.

6.3.3.4. In the presence of linear ligand, DB921

The SAXS scattering curves were collected from protein-DNA-DB921; protein-DB921 and DB921 at varying concentrations of ligand (10mM to 1μM). The PRIMUS and GNOM analysis are listed in Table 6.11. and the OLIGOMER analysis is in Table 6.12. The PRIMUS and GNOM analysis could only be made at lower concentrations of DB921. At the higher DB921 concentrations the scattering pattern resembled to a well ordered globular system. The OLIGOMER analysis for the scattering curves of p50-DNA-DB921; p50-DB921 and DB921 alone were performed. From the OLIGOMER analysis of the p50-DNA-DB921 complex system it is apparent that the volume fraction of p50-DNA complex decreases while the protein component increases with increasing concentration of DB921. This indicates that DB921 causes dissociation of p50-DNA complex at higher concentration, although no significant conclusion can be drawn from the OLIGOMER analysis of the protein-DB921 system.

DB921 concentration	R_g (nm)	$I(0)$	D_{max} (nm)	MM (kDa)	V_p (nm^3)	Quality (%)
40µM	2.49	13	12.7	59	112	43
30 µM	3.12	13.66	13.5	62	160	52
20 µM	3.56	14.04	12.2	63	151	80
10 µM	3.48	13.10	12.2	59	146	81
1 µM	3.50	13.40	11.7	61	148	87

Table 6.11. PRIMUS and GNOM parameters of the p50 protein and DNA in the presence of DB921.

DB921 conc (µM)	Volume fraction of each component (%)		
	Protein	Protein + DNA	Chi^2
	Protein - DNA- DB921		
100	100	0	6.8
50	90	10	2.6
30	60	40	1.7
20	20	80	1.2
10	3	97	0.9
	Protein – DB921		
100	-	-	-
50	100	0	-
30	-	-	-
20	-	-	-

Table 6.12. OLIGOMER analysis of SAXS scattering curves of p50-DNA-DB921 and p50-DB921 at varying concentrations of DB921.

As mentioned earlier, at higher concentrations of DB921 (10mM - 100µM) the scattering curves of the p50-DNA-DB921; p50-DB921 and DB921 alone resembled that of a well ordered spherical or globular structure (Figure 6.21). The scattering intensities of these curves were higher than those observed for protein alone or the protein-DNA complex solution. An initial Guinier analysis indicated that the radius of gyration for the solution could not be assessed as it was very large and beyond the limit of the SAXS set-up.

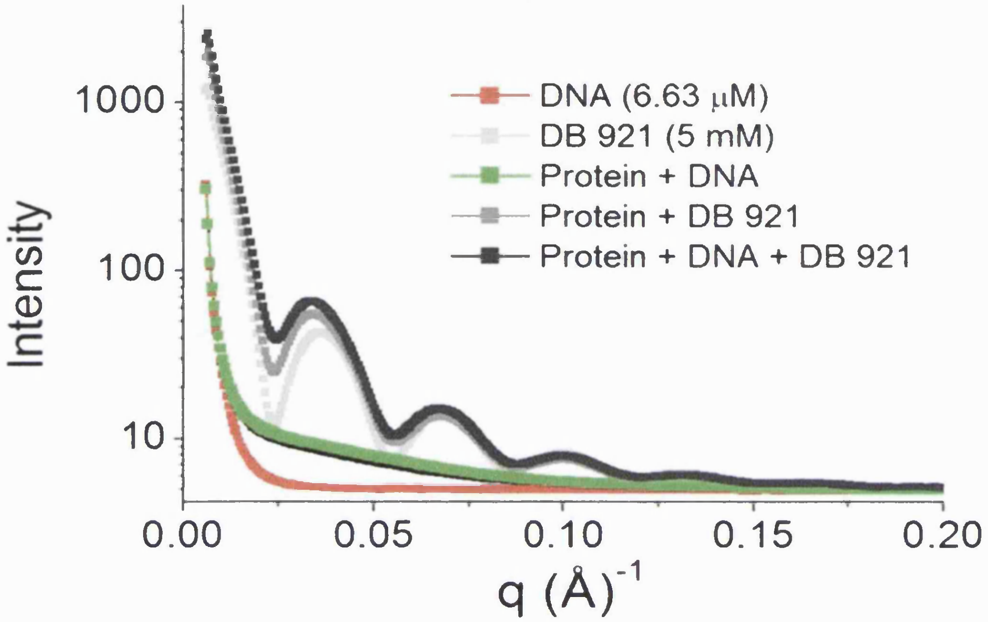


Figure 6.21. SAXS scattering curves of p50-DNA-DB921; p50-DB921 and DB921 alone, all shown in shades of grey. These curves are compared against the scattering curves obtained from p50-DNA complex solution and DNA alone. The protein, DNA and DB921 concentration in all solutions were 0.5mg/ml, 6.6μM and 5mM respectively.

To obtain a general overview of the size and shape of these molecules a complementary technique, negative staining electron microscopy was used. The microscopy was performed at IBS, Grenoble using the mica-carbon flotation technique and sodium silico tungstate (SST) at 2% as dye. The four solutions provided for testing with negative stain were (a) protein alone at 0.5mg/ml (b) DB921 alone at 100μM, (c) protein (0.5mg/ml) and DB921 (100μM) and (d) protein-DNA-DB921 (where protein and DB921 concentration were the same as above and the DNA concentration was 6.2μM). The SAXS scattering curves for these solutions are shown in Figure 6.22. These samples were diluted 10 times for negative stain analysis. The 2D images obtained for the samples are shown in Figure 6.23.

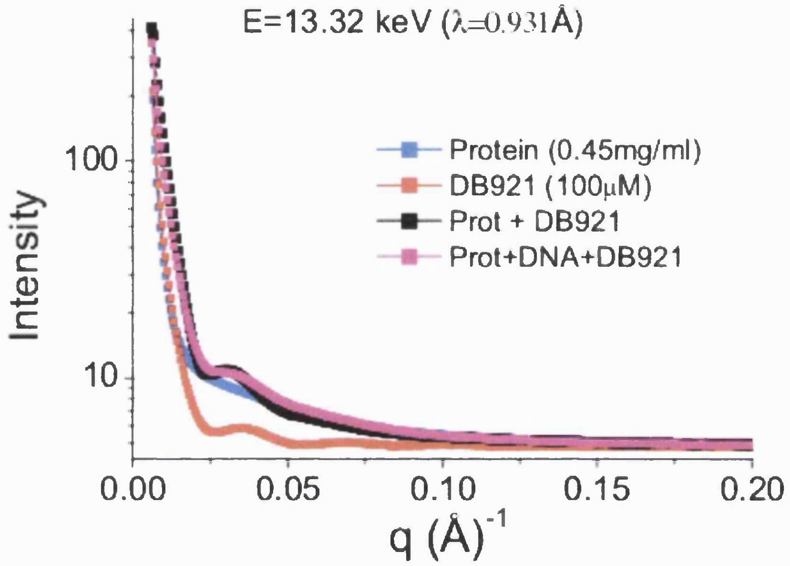


Figure 6.22. SAXS scattering curves of samples used for negative stain study.

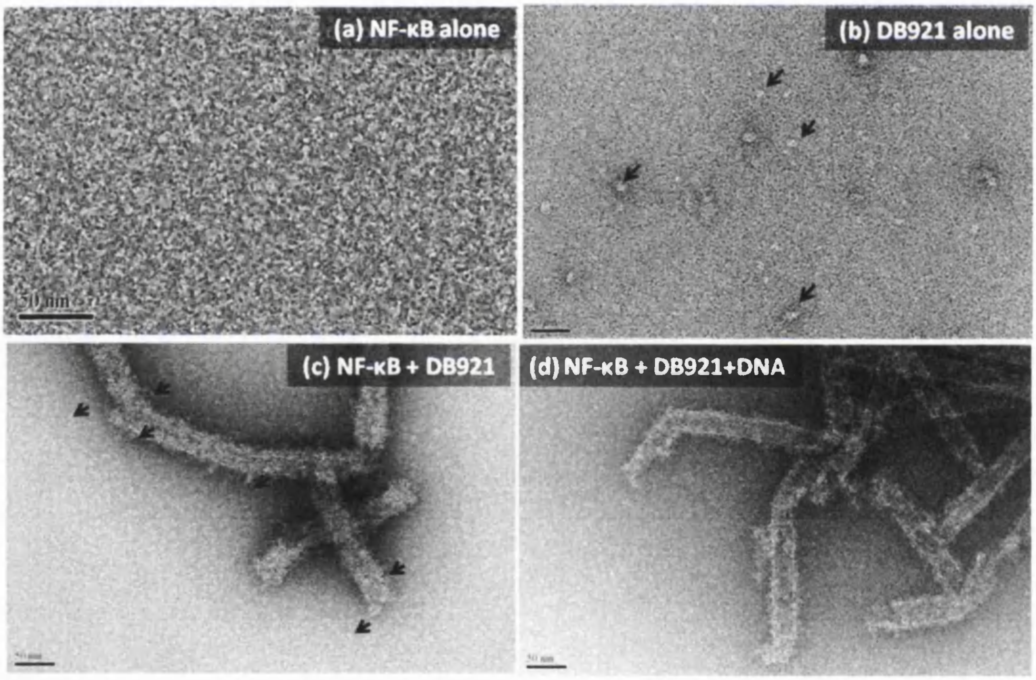


Figure 6.23. Negative stain images obtained for (a) p50 protein alone, (b) DB921 ligand alone, (c) protein-DB921 and (d) protein-DNA-DB921 complex solution.

The negative stain result for protein alone sample shows a monodispersed solution with small white spots representing the protein (Figure 6.23.a). The size and shape of the protein could not be assessed. The results for DB921 solution showed the presence of globular particles (Figure 6.23.b). These particles were not homogenous and were surprisingly larger in size than for the protein alone. The globular particles were approximately 15nm in diameter. For the protein-DB921 sample, the grid was covered with "nanotubes" of approximate diameter 40nm (Figure 6.23.c). It was also observed that some protein or DB921 particles appeared to be attached to the tubes. A similar result was obtained for the protein-DNA-DB921 sample (Figure 6.23.d), but this sample had less tubes when compared to the protein-DB921 sample.

Although negative stain analysis was performed on a single concentration of DB921 (100μM), it suggests an unusual self-assembly behaviour of the transcription factor in the presence of DB921 at high concentration. However, the globular shape of DB921 alone suggests that the ligand may itself have some self-assembly properties. In this case the self-assembly of DB921 is observed in absence of protein and/or DNA which could lead to the conclusion that the self-assembly phenomenon is caused by the buffers used in sample preparation. However, this cannot be confirmed until further experiments are performed.

6.3.4. Discussion

The SAXS analysis of the protein alone and protein-DNA complexes shows a significant difference in their R_g values. The smaller R_g value ($R_g =$ 3.44nm) of the protein-DNA complex solution indicates a higher molecular weight, it is more compact and has a smaller size compared to protein alone ($R_g =$ 3.79nm). This difference in R_g values could be due to the protein solution alone being more flexible but in the presence of DNA the transcription factor, p50 binds to DNA which hinders the protein flexibility and forms a more compact structure. The *ab initio* model generated for protein alone and protein-DNA complex also show similar results. The CRSYOL analysis of the SAXS scattering curve of protein alone and X-ray crystal structure of smaller construct of p50 dimer without DNA (PDB: 1BFS) and p65 dimer with DNA (PDB: 1NFI) showed poor fit while comparison with p50 dimer bound to DNA (PDB: 1NFK) has a superior fit with a Chi^2 value of 3.8. The OLIGOMER analysis on the SAXS scattering curve of protein alone showed that the protein was present as a mixture of p50 monomer (8%) and dimer (92%). The CRYSOL fit also implies that the experimental scattering curve of protein alone resembles the crystal structure coordinates of dimer which was obtained from the p50 dimer-DNA complex structure (PDB: 1NFK). However, this p50 dimer structure coordinate from 1NFK would be biased as the original crystal structure was solved in presence of DNA. The experimental curve of protein-DNA complex solution was also analysed with CRYSOL and it provided a good fit with the X-ray crystal structure of p50 dimer-DNA with a Chi^2 value of 2.6 (PDB: 1NFK). This shows that the crystal structure and structure of protein-DNA complex in solution are similar.

The effect of MGBLs on the protein-DNA interaction was also studied. The OLIGOMER analysis was a better method to determine the effects of ligand as satisfactory model could not be obtained in presence of ligands. From the OLIGOMER analyses it was evident that the all the ligands used in the study; isohelical ligands berenil and DB75 and linear ligands DB985 and DB921, had significant effects on protein-DNA interactions. From the OLIGOMER analysis it was evident that with increasing concentration of the MGBLs the volume fraction of protein-DNA complex component decreases while the protein alone component increases. This shows that the protein-DNA interaction is affected by the presence of MGBLs. The OLIGOMER analysis for all the MGBLs is summarised in Figure 6.24.

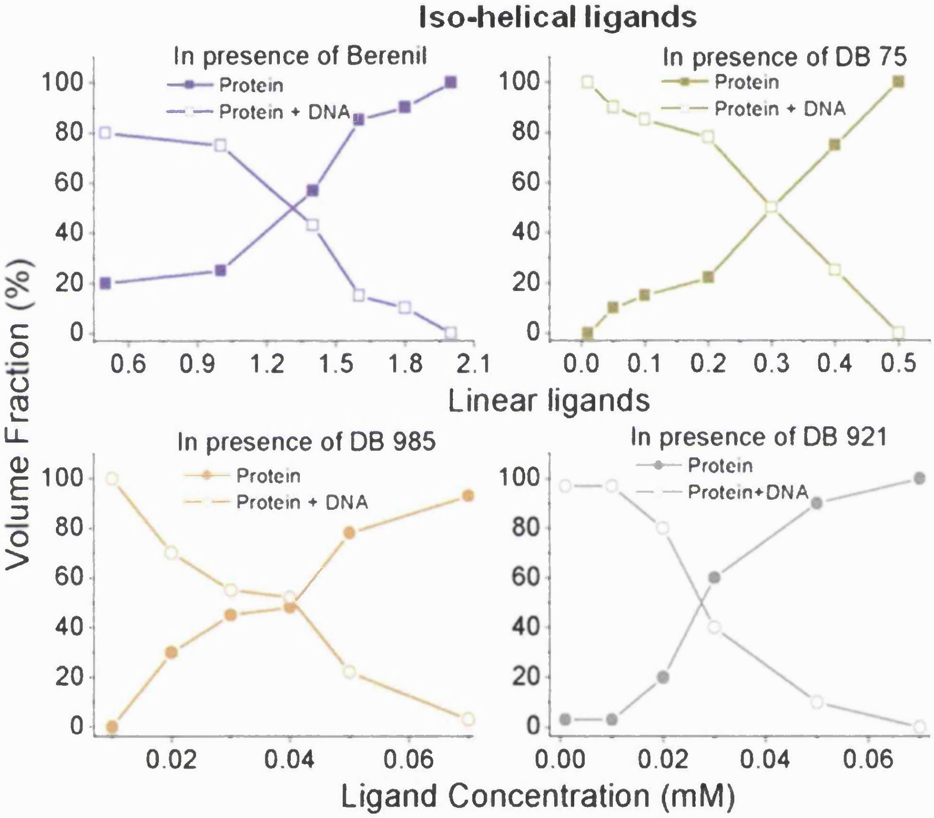


Figure 6.24. The results for the OLIGOMER analysis done on SAXS scattering curves of protein-DNA solution is presence of varying concentration of MGBLs; iso-helical ligands, berenil and DB75 and linear ligands, DB985 and DB921.

It is evident from Figure 6.24, that all the ligands have similar effect on the volume fractions of protein and protein-DNA in presence of ligand. Although it can also be observed that the concentration at which these ligand exhibits their effect on protein-

DNA interaction are different. While the iso-helical ligand, berenil is effective between 2mM and 250μM another iso-helical ligand DB75 affects the protein-DNA interaction within the range of 1mM to 100μM. In comparison to the iso-helical ligands, both linear ligands show activity between 100μM to 10μM concentration range. If inhibition of the transcription activity of the transcription factor, NF-κB is one of the mechanisms of action of MGBLs, then the linear ligands would be more promising scaffold to design new drugs. As the linear ligands are active at lower concentration range and would need to be administered at low dose and hence reducing the toxicity issues involved with several of these MGBLs. But at the same time the linear ligands have a small therapeutic window in which they are active as compared to iso-helical ligands.

The rare self-assembly occurrence of protein and DB921 observed could be a reason for SPR results obtained for DB921 which shows complete inhibition of protein-DNA interaction with both biotin control DNA (Section 6.2.2.3) and biotin drug control DNA (Section 6.2.3).

However, additional work such as crystallographic study of the tertiary complex (p50-DNA-ligand), further electron microscopy and SAXS studies and also contrast variation study using SANS study can be investigated.

Chapter 7

Conclusion and Future work

The minor groove of duplex DNA is a major target site for many small molecule ligands. Minor groove binding ligands (MGBLs) with fairly good sequence discrimination are of interest as potential therapeutic agents for many human diseases. Several MGBLs have been previously studied extensively by structural and biophysical techniques.

In this research we have studied two classes of diamidine derivative ligands: the traditional iso-helical ligands and the more novel linear ligands. The ligands studies were chosen based on their structural features. High resolution X-ray crystallographic studies have been performed on several MGBLs in complex with two DNA sequences, d(CGCGAATTCGCG)$_2$ (A_2T_2) and d(CGCAAATTTGCG)$_2$ (A_3T_3). The crystal structures of the MGBLs show that all the ligands bind in the AT rich region of the minor groove of the DNA.

The X-ray crystal structure of linear ligands, DB985 and DB921, form direct hydrogen bonds with the bases of the DNA via the nitrogen atom of the benzimidazole ring of the ligands and the amidine group at the benzimidazole end of the ligand. However, the linear shape of the ligand causes a decrease in curvature of the ligand making the amidine group at the phenyl end to protrude out of the minor groove and prohibiting direct contacts with the DNA bases. This limitation is overcome by the conserved water molecules present at the phenyl end of the ligand in the minor groove of the DNA. The water molecules participate in indirect contact of the ligand with bases of the DNA, hence providing effectively a curved mode of binding which matches the classical iso-helical ligand binding mode.

The crystal structures of the iso-helical ligands, DB818 and DB262, however, show that both the amidine groups of this ligands form direct hydrogen bonds with DNA. DB818 forms bifurcated hydrogen bonds with the DNA via the nitrogen atom of the benzimidazole rings of the ligand, which is also observed with the linear ligand structures. DB262 has an increased curvature which causes the central aromatic rings of the ligand to protrude from the minor groove. Due to this the nitrogen atom of the central pyrrole ring of DB262 is unable to form a bifurcated hydrogen bond with DNA.

The biophysical studies that have been performed earlier (Mallena *et al.* 2004; Miao *et al.* 2005) indicate that linear ligand DB921 ($K_a = 2.7 \times 10^8 M^{-1}$) and iso-helical ligand DB818 ($K_a = 2.8 \times 10^8 M^{-1}$) have similar binding affinity to DNA. From the structural studies of the linear ligands, DB921 and DB985, they have similar hydrogen bonding patterns and the water network is also stable in the groove. However, DB985 has a less strong binding affinity to DNA ($K_a = 6.1 \times 10^7 M^{-1}$, data provided by collaborators). DB262, an iso-helical ligand, on the other hand has the least binding affinity ($K_a = 2.8 \times 10^7 M^{-1}$, Munde *et al.* 2007) compared to the other ligands studied here. The crystal structure of DB262 and d(CGCAAATTTGCG)$_2$ does lack some hydrogen bonds that are observed with the linear ligands and iso-helical ligand DB818; this could explain the lower binding affinity of DB262.

From comparing the ligand bound structure with the native DNA structure it is observed that the stabilisation of the ligands in the minor groove occurs through the displacement of ordered water from the minor groove, as well as interactions with the remaining minor-groove water molecules. Therefore the water molecules play a very important role in maintaining the stability and flexibility of DNA. Information on hydration and protonation states and on the way in which these contribute to the stability of the DNA in presence of MGBLs is of key interest for an understanding of the details of the drug binding. This makes the DNA-ligand complexes good candidates for neutron crystallography studies. Crystallisation trials were performed in deuterated conditions to obtain large volume crystals suitable for neutron diffraction studies. The iso-helical ligands DB75 and berenil, and linear ligands DB921 and DB985 are prime candidates for neutron study. After extensive screening and optimisations of crystallisation conditions, large crystals of DB75 and DB985 in complex with A_3T_3 DNA were grown and were tested for neutron diffraction. These crystals provided good neutron diffraction results but could not be collected due to unexpected unavailability of the neutron beamlines at the ILL.

The detailed mechanism of action of some of these MGBLs is still unproven, although they are known to be effective inhibitors of a number of minor and major groove binding protein-DNA interactions. In order to further our understanding of the biological mechanism of action of MGBLs, biophysical studies were undertaken with the DNA major groove binding transcription factor, NF-κB. This transcription factor binds to the continuous guanine and cytosine ends of the major groove leaving the adenine-thymine rich minor groove exposed to other molecules. Two ligands belonging to each iso-helical (berenil and DB75) and linear (DB985 and DB921) class of ligands were chosen for SPR and SAXS studies. The research was conducted to study the change in interaction of NF-κB to DNA in presence of these MGBLs. With the SPR study it was observed that the iso-helical ligands, berenil and DB75, caused significant change in the NF-κB DNA interaction. The presence of DB75 did not affect the association of the protein to DNA, however, it increased the rate of dissociation; whilst in presence of berenil there was a decrease in association rate constant and increase in dissociation rate. The two linear ligands studied, DB921 and DB985, had different effects on the protein-DNA interaction. The presence of DB985 increased the rate of protein-DNA dissociation, more than the two iso-helical ligands, while DB921 completely perturbs the protein-DNA interaction at similar concentrations as used for other ligands. This difference in the results obtained for structurally similar linear ligands could be due to stronger DNA binding affinity of DB921 ($K_a = 2.7 \times 10^8 M^{-1}$) compared to DB985 ($K_a = 6.1 \times 10^7 M^{-1}$). To determine that the effects observed in the interaction of protein-DNA were due to DNA-ligand interactions, a negative control SPR study was performed where a DNA sequence was chosen with no ligand binding site. This sequence was tested with DB75 (iso-helical ligand) and DB921 (linear ligand). It was observed that in the presence of DB75 there was no difference in NF-κB and DNA interaction, although in the presence of DB921 there was still complete loss of protein-DNA interaction.

From early structural studies of the NF-κB and DNA complex it has been seen that binding of the protein to DNA causes a change in DNA curvature, while crystallographic studies of the DNA-ligand complex have shown that MGBLs bind to the B-conformation of DNA and stabilises the DNA in this form. A bio-SAXS experiment was therefore performed in this work to follow the changes in the structural envelope of the protein-DNA in the presence of MGBLs. The SAXS analysis of provided a smaller R_g value ($R_g = 3.44$ nm) for the protein-DNA complex while giving a larger R_g value ($R_g = 3.79$ nm) for the protein alone. This difference in R_g values could be due to the transcription factor, NF-κB, alone in solution being more flexible, but in the presence of DNA the protein binds to DNA which hinders the protein flexibility and forms a more compact structure. The CRYSOL and OLIGOMER analysis of the scattering curve of protein alone indicate that the protein was present as a mixture of monomer and dimer in solution. The *ab initio* model of protein-DNA complex in presence of MGBLs could not be obtained due the poor quality of the sample solution at high concentration of MGBLs and therefore OLIGOMER analysis was used. Using the simple OLIGOMER analysis involving only the protein-DNA and protein alone component it was observed that with increasing concentration of the MGBLs the volume fraction of protein-DNA complex component decreases while the protein component increases. This clearly indicates that MGBLs causes a change in the protein-DNA interaction which further supports the results obtained from the SPR analysis. It was also observed from SAXS analysis that the linear ligands DB985 and DB921 affected the protein-DNA interaction at a lower concentration range than the iso-helical ligands berenil and DB75. However, at higher concentrations of DB921 an unusual self-assembly was observed in SAXS and further electron microscopy negative stain studies were performed. The electron microscopy result suggests that this ligand caused the protein to self-assemble which indicates that DB921 can also interact with the protein. This shows that DB921 probably has a different mode of binding and hence alters the binding to protein-DNA even in absence of a ligand binding site in the DNA sequence as observed in the negative control SPR study.

7.1. Future work

Several MGBLs were chosen based on their structural features to obtain crystal structure of these ligands in complex with DNA. While the crystal structures of most of the shortlisted ligands have been obtained, the ligands DB685 and DB832A are yet to be successfully studied. The X-ray crystal structure of these ligands in complex with DNA could not be determined as the crystals were not of sufficient quality. The crystallisation conditions of these ligands needs to be further optimised to obtain better quality crystals with ligand bound in the structure. Crystals of DB75 and DB985 in complex with $d(CGCAAATTTGCG)_2$ have been obtained of sufficient volume for neutron studies. As, unfortunately, the data could not be collected for this research, the crystals remain available for future neutron data collection.

The SPR and SAXS results suggest that MGBLs have significant effect on the interaction of transcription factor, NF-κB, and DNA. However, no structural envelopes could be obtained from the SAXS studies performed. Crystallisation studies of the tertiary complex protein-DNA-MGBLs would provide important structural information on the cause of the change in protein-DNA interaction in the presence of MGBLs. From an earlier gel shift assay it has been shown that protein and the MGBL Hoechst 33258 can bind simultaneously to the same DNA duplex (Speight *et al.* 2002), therefore it should be possible to obtain the tertiary complex. The crystal structure of the tertiary complex compared with the structure of protein-DNA complex would help to determine the reasons for the change in their interactions. More electron microscopy studies could also be performed on other ligands to determine if these ligands also posses the self-assembly ability at higher concentrations as observed with DB921.

From the research conducted, a different biological significance of these MGBLs has been revealed. It has also helped to realise that the traditional iso-helical ligands are not the only strong DNA binders but linear ligands also hold promise too. This further shows that this class of ligand shows potential and needs to be further studied and optimised to develop better drugs with improved binding efficacy and biological significance.

Appendix I

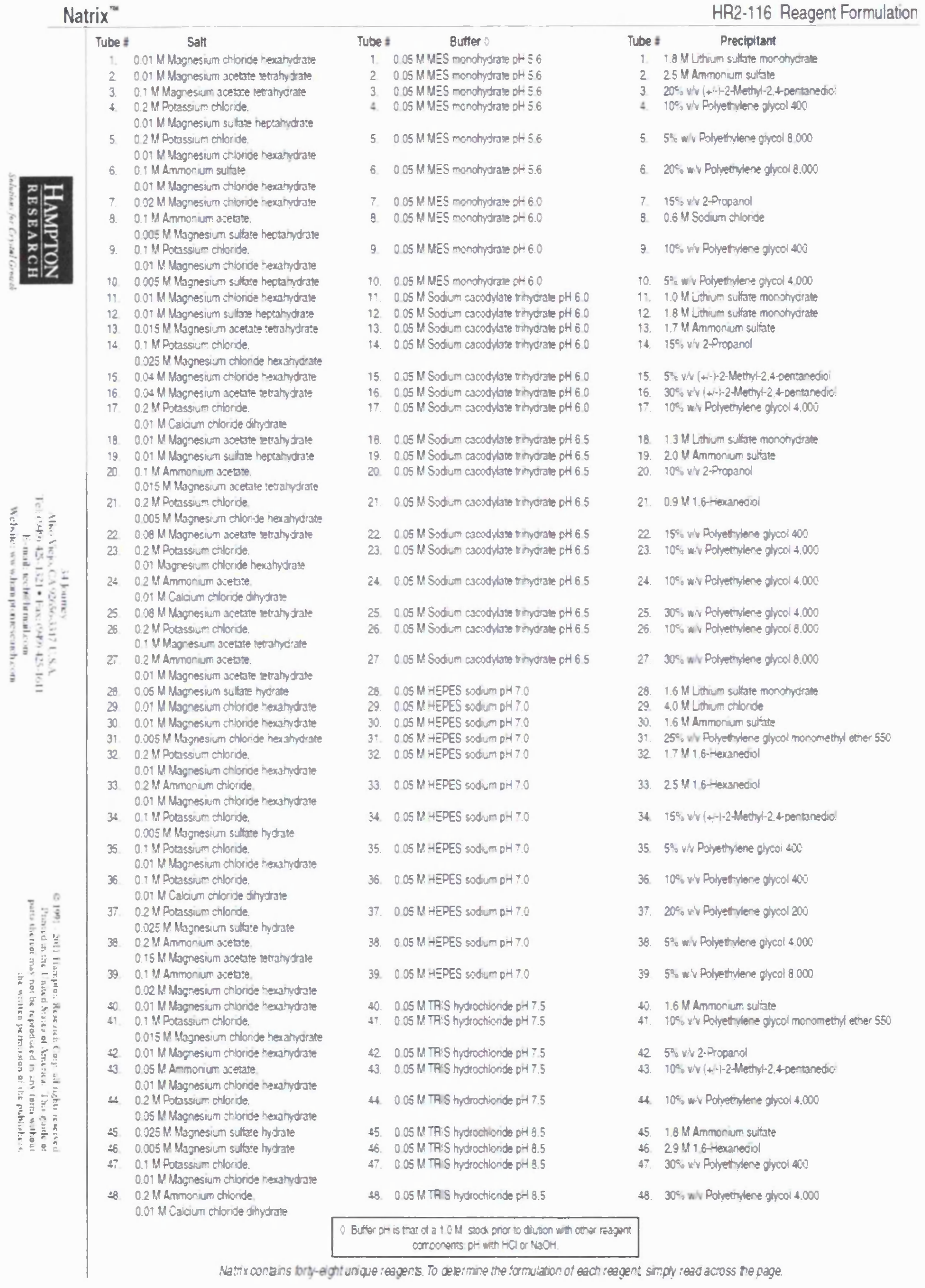
Nucleic Acid Mini Screen™ HR2-118 Reagent Formulation

Tube Number	Precipitant	Tube Number	Buffer	Tube Number	Polyamine	Tube Number	Monovalent Ion	Tube Number	Divalent Ion
1.	10% v/v MPD	1.	40 mM Na Cacodylate pH 5.5	1.	20 mM Cobalt Hexamine	1.	None	1.	20 mM Magnesium Chloride
2.	10% v/v MPD	2.	40 mM Na Cacodylate pH 5.5	2.	20 mM Cobalt Hexamine	2.	80 mM Sodium Chloride	2.	20 mM Magnesium Chloride
3.	10% v/v MPD	3.	40 mM Na Cacodylate pH 5.5	3.	20 mM Cobalt Hexamine	3.	12 mM Sodium Chloride / 80 mM Potassium Chloride	3.	None
4.	10% v/v MPD	4.	40 mM Na Cacodylate pH 5.5	4.	20 mM Cobalt Hexamine	4.	40 mM Lithium Chloride	4.	20 mM Magnesium Chloride
5.	10% v/v MPD	5.	40 mM Na Cacodylate pH 6.0	5.	12 mM Spermine tetra-HCl	5.	80 mM Potassium Chloride	5.	20 mM Magnesium Chloride
6.	10% v/v MPD	6.	40 mM Na Cacodylate pH 6.0	6.	12 mM Spermine tetra-HCl	6.	80 mM Potassium Chloride	6.	None
7.	10% v/v MPD	7.	40 mM Na Cacodylate pH 6.0	7.	12 mM Spermine tetra-HCl	7.	80 mM Sodium Chloride	7.	20 mM Magnesium Chloride
8.	10% v/v MPD	8.	40 mM Na Cacodylate pH 6.0	8.	12 mM Spermine tetra-HCl	8.	80 mM Sodium Chloride	8.	None
9.	10% v/v MPD	9.	40 mM Na Cacodylate pH 6.0	9.	12 mM Spermine tetra-HCl	9.	80 mM Sodium Chloride / 12 mM Potassium Chloride	9.	20 mM Magnesium Chloride
10.	10% v/v MPD	10.	40 mM Na Cacodylate pH 6.0	10.	12 mM Spermine tetra-HCl	10.	12 mM Sodium Chloride / 80 mM Potassium Chloride	10.	None
11.	10% v/v MPD	11.	40 mM Na Cacodylate pH 6.0	11.	12 mM Spermine tetra-HCl	11.	80 mM Sodium Chloride	11.	20 mM Barium Chloride
12.	10% v/v MPD	12.	40 mM Na Cacodylate pH 6.0	12.	12 mM Spermine tetra-HCl	12.	80 mM Potassium Chloride	12.	20 mM Barium Chloride
13.	10% v/v MPD	13.	40 mM Na Cacodylate pH 6.0	13.	12 mM Spermine tetra-HCl	13.	None	13.	80 mM Strontium Chloride
14.	10% v/v MPD	14.	40 mM Na Cacodylate pH 7.0	14.	12 mM Spermine tetra-HCl	14.	80 mM Potassium Chloride	14.	20 mM Magnesium Chloride
15.	10% v/v MPD	15.	40 mM Na Cacodylate pH 7.0	15.	12 mM Spermine tetra-HCl	15.	80 mM Potassium Chloride	15.	None
16.	10% v/v MPD	16.	40 mM Na Cacodylate pH 7.0	16.	12 mM Spermine tetra-HCl	16.	80 mM Sodium Chloride	16.	20 mM Magnesium Chloride
17.	10% v/v MPD	17.	40 mM Na Cacodylate pH 7.0	17.	12 mM Spermine tetra-HCl	17.	80 mM Sodium Chloride	17.	None
18.	10% v/v MPD	18.	40 mM Na Cacodylate pH 7.0	18.	12 mM Spermine tetra-HCl	18.	80 mM Sodium Chloride / 12 mM Potassium Chloride	18.	20 mM Magnesium Chloride
19.	10% v/v MPD	19.	40 mM Na Cacodylate pH 7.0	19.	12 mM Spermine tetra-HCl	19.	12 mM Sodium Chloride / 80 mM Potassium Chloride	19.	None
20.	10% v/v MPD	20.	40 mM Na Cacodylate pH 7.0	20.	12 mM Spermine tetra-HCl	20.	80 mM Sodium Chloride	20.	20 mM Barium Chloride
21.	10% v/v MPD	21.	40 mM Na Cacodylate pH 7.0	21.	12 mM Spermine tetra-HCl	21.	80 mM Potassium Chloride	21.	20 mM Barium Chloride
22.	10% v/v MPD	22.	40 mM Na Cacodylate pH 7.0	22.	12 mM Spermine tetra-HCl	22.	40 mM Lithium Chloride	22.	80 mM Strontium Chloride / 20 mM Magnesium Chloride
23.	10% v/v MPD	23.	40 mM Na Cacodylate pH 7.0	23.	12 mM Spermine tetra-HCl	23.	40 mM Lithium Chloride	23.	80 mM Strontium Chloride
24.	10% v/v MPD	24.	40 mM Na Cacodylate pH 7.0	24.	12 mM Spermine tetra-HCl	24.	None	24.	80 mM Strontium Chloride / 20 mM Magnesium Chloride

Nucleic Acid Mini Screen contains twenty-four unique reagents. To determine the formulation of each reagent, simply read across the page.

Nucleic Acid Mini (NAM) Screen formulation from Hampton Research

Appendix II



Natrix™

HR2-116 Reagent Formulation

Tube #	Salt	Tube #	Buffer ◊	Tube #	Precipitant
1.	0.01 M Magnesium chloride hexahydrate	1.	0.05 M MES monohydrate pH 5.6	1.	1.8 M Lithium sulfate monohydrate
2.	0.01 M Magnesium acetate tetrahydrate	2.	0.05 M MES monohydrate pH 5.6	2.	2.5 M Ammonium sulfate
3.	0.1 M Magnesium acetate tetrahydrate	3.	0.05 M MES monohydrate pH 5.6	3.	20% v/v (+/-)-2-Methyl-2,4-pentanediol
4.	0.2 M Potassium chloride, 0.01 M Magnesium sulfate heptahydrate	4.	0.05 M MES monohydrate pH 5.6	4.	10% v/v Polyethylene glycol 400
5.	0.2 M Potassium chloride, 0.01 M Magnesium chloride hexahydrate	5.	0.05 M MES monohydrate pH 5.6	5.	5% w/v Polyethylene glycol 8,000
6.	0.1 M Ammonium sulfate, 0.01 M Magnesium chloride hexahydrate	6.	0.05 M MES monohydrate pH 5.6	6.	20% w/v Polyethylene glycol 8,000
7.	0.02 M Magnesium chloride hexahydrate	7.	0.05 M MES monohydrate pH 6.0	7.	15% v/v 2-Propanol
8.	0.1 M Ammonium acetate, 0.005 M Magnesium sulfate heptahydrate	8.	0.05 M MES monohydrate pH 6.0	8.	0.6 M Sodium chloride
9.	0.1 M Potassium chloride, 0.01 M Magnesium chloride hexahydrate	9.	0.05 M MES monohydrate pH 6.0	9.	10% v/v Polyethylene glycol 400
10.	0.005 M Magnesium sulfate heptahydrate	10.	0.05 M MES monohydrate pH 6.0	10.	5% w/v Polyethylene glycol 4,000
11.	0.01 M Magnesium chloride hexahydrate	11.	0.05 M Sodium cacodylate trihydrate pH 6.0	11.	1.0 M Lithium sulfate monohydrate
12.	0.01 M Magnesium sulfate heptahydrate	12.	0.05 M Sodium cacodylate trihydrate pH 6.0	12.	1.8 M Lithium sulfate monohydrate
13.	0.015 M Magnesium acetate tetrahydrate	13.	0.05 M Sodium cacodylate trihydrate pH 6.0	13.	1.7 M Ammonium sulfate
14.	0.1 M Potassium chloride, 0.025 M Magnesium chloride hexahydrate	14.	0.05 M Sodium cacodylate trihydrate pH 6.0	14.	15% v/v 2-Propanol
15.	0.04 M Magnesium chloride hexahydrate	15.	0.05 M Sodium cacodylate trihydrate pH 6.0	15.	5% v/v (+/-)-2-Methyl-2,4-pentanediol
16.	0.04 M Magnesium acetate tetrahydrate	16.	0.05 M Sodium cacodylate trihydrate pH 6.0	16.	30% v/v (+/-)-2-Methyl-2,4-pentanediol
17.	0.2 M Potassium chloride, 0.01 M Calcium chloride dihydrate	17.	0.05 M Sodium cacodylate trihydrate pH 6.0	17.	10% w/v Polyethylene glycol 4,000
18.	0.01 M Magnesium acetate tetrahydrate	18.	0.05 M Sodium cacodylate trihydrate pH 6.5	18.	1.3 M Lithium sulfate monohydrate
19.	0.01 M Magnesium sulfate heptahydrate	19.	0.05 M Sodium cacodylate trihydrate pH 6.5	19.	2.0 M Ammonium sulfate
20.	0.1 M Ammonium acetate, 0.015 M Magnesium acetate tetrahydrate	20.	0.05 M Sodium cacodylate trihydrate pH 6.5	20.	10% v/v 2-Propanol
21.	0.2 M Potassium chloride, 0.005 M Magnesium chloride hexahydrate	21.	0.05 M Sodium cacodylate trihydrate pH 6.5	21.	0.9 M 1,6-Hexanediol
22.	0.08 M Magnesium acetate tetrahydrate	22.	0.05 M Sodium cacodylate trihydrate pH 6.5	22.	15% v/v Polyethylene glycol 400
23.	0.2 M Potassium chloride, 0.01 Magnesium chloride hexahydrate	23.	0.05 M Sodium cacodylate trihydrate pH 6.5	23.	10% w/v Polyethylene glycol 4,000
24.	0.2 M Ammonium acetate, 0.01 M Calcium chloride dihydrate	24.	0.05 M Sodium cacodylate trihydrate pH 6.5	24.	10% w/v Polyethylene glycol 4,000
25.	0.08 M Magnesium acetate tetrahydrate	25.	0.05 M Sodium cacodylate trihydrate pH 6.5	25.	30% w/v Polyethylene glycol 4,000
26.	0.2 M Potassium chloride, 0.1 M Magnesium acetate tetrahydrate	26.	0.05 M Sodium cacodylate trihydrate pH 6.5	26.	10% w/v Polyethylene glycol 8,000
27.	0.2 M Ammonium acetate, 0.01 M Magnesium acetate tetrahydrate	27.	0.05 M Sodium cacodylate trihydrate pH 6.5	27.	30% w/v Polyethylene glycol 8,000
28.	0.05 M Magnesium sulfate hydrate	28.	0.05 M HEPES sodium pH 7.0	28.	1.6 M Lithium sulfate monohydrate
29.	0.01 M Magnesium chloride hexahydrate	29.	0.05 M HEPES sodium pH 7.0	29.	4.0 M Lithium chloride
30.	0.01 M Magnesium chloride hexahydrate	30.	0.05 M HEPES sodium pH 7.0	30.	1.6 M Ammonium sulfate
31.	0.005 M Magnesium chloride hexahydrate	31.	0.05 M HEPES sodium pH 7.0	31.	25% v/v Polyethylene glycol monomethyl ether 550
32.	0.2 M Potassium chloride, 0.01 M Magnesium chloride hexahydrate	32.	0.05 M HEPES sodium pH 7.0	32.	1.7 M 1,6-Hexanediol
33.	0.2 M Ammonium chloride, 0.01 M Magnesium chloride hexahydrate	33.	0.05 M HEPES sodium pH 7.0	33.	2.5 M 1,6-Hexanediol
34.	0.1 M Potassium chloride, 0.005 M Magnesium sulfate hydrate	34.	0.05 M HEPES sodium pH 7.0	34.	15% v/v (+/-)-2-Methyl-2,4-pentanediol
35.	0.1 M Potassium chloride, 0.01 M Magnesium chloride hexahydrate	35.	0.05 M HEPES sodium pH 7.0	35.	5% v/v Polyethylene glycol 400
36.	0.1 M Potassium chloride, 0.01 M Calcium chloride dihydrate	36.	0.05 M HEPES sodium pH 7.0	36.	10% v/v Polyethylene glycol 400
37.	0.2 M Potassium chloride, 0.025 M Magnesium sulfate hydrate	37.	0.05 M HEPES sodium pH 7.0	37.	20% v/v Polyethylene glycol 200
38.	0.2 M Ammonium acetate, 0.15 M Magnesium acetate tetrahydrate	38.	0.05 M HEPES sodium pH 7.0	38.	5% w/v Polyethylene glycol 4,000
39.	0.1 M Ammonium acetate, 0.02 M Magnesium chloride hexahydrate	39.	0.05 M HEPES sodium pH 7.0	39.	5% w/v Polyethylene glycol 8,000
40.	0.01 M Magnesium chloride hexahydrate	40.	0.05 M TRIS hydrochloride pH 7.5	40.	1.6 M Ammonium sulfate
41.	0.1 M Potassium chloride, 0.015 M Magnesium chloride hexahydrate	41.	0.05 M TRIS hydrochloride pH 7.5	41.	10% v/v Polyethylene glycol monomethyl ether 550
42.	0.01 M Magnesium chloride hexahydrate	42.	0.05 M TRIS hydrochloride pH 7.5	42.	5% v/v 2-Propanol
43.	0.05 M Ammonium acetate, 0.01 M Magnesium chloride hexahydrate	43.	0.05 M TRIS hydrochloride pH 7.5	43.	10% v/v (+/-)-2-Methyl-2,4-pentanediol
44.	0.2 M Potassium chloride, 0.05 M Magnesium chloride hexahydrate	44.	0.05 M TRIS hydrochloride pH 7.5	44.	10% w/v Polyethylene glycol 4,000
45.	0.025 M Magnesium sulfate hydrate	45.	0.05 M TRIS hydrochloride pH 8.5	45.	1.8 M Ammonium sulfate
46.	0.005 M Magnesium sulfate hydrate	46.	0.05 M TRIS hydrochloride pH 8.5	46.	2.9 M 1,6-Hexanediol
47.	0.1 M Potassium chloride, 0.01 M Magnesium chloride hexahydrate	47.	0.05 M TRIS hydrochloride pH 8.5	47.	30% v/v Polyethylene glycol 400
48.	0.2 M Ammonium chloride, 0.01 M Calcium chloride dihydrate	48.	0.05 M TRIS hydrochloride pH 8.5	48.	30% w/v Polyethylene glycol 4,000

◊ Buffer pH is that of a 1.0 M stock prior to dilution with other reagent components: pH with HCl or NaOH.

Natrix contains forty-eight unique reagents. To determine the formulation of each reagent, simply read across the page.

HAMPTON RESEARCH
Solutions for Crystal Growth

34 Journey
Aliso Viejo, CA 92656-3317 U.S.A.
Tel: (949) 425-1321 • Fax: (949) 425-1611
E-mail: tech@hrmail.com
Website: www.hamptonresearch.com

NATRIX Screen formulation from Hampton Research

References

Abu-Daya, A., Brown, P. M., and Fox, K. R. (1995) DNA sequence preferences of several AT-selective minor groove binding ligands, *Nucleic Acids Research 23*, 3385-3392.

Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution, *Acta Crystallographica Section D 66*, 213-221.

Arnott, S., Chandrasekaran, R., Birdsall, D. L., Leslie, A. G. W., and Ratliff, R. L. (1980) Left-handed DNA helices, *Nature 283*, 743-745.

Baeuerle, P. A., and Henkel, T. (1994) Function and Activation of NF-kappaB in the Immune System, *Annual Review of Immunology 12*, 141-179.

Bakshi, R. P., and Shapiro, T. A. (2003) DNA topoisomerases as targets for antiprotozoal therapy, *Mini Reviews in Medicinal Chemistry 3*, 597-608.

Baldwin, A. S. (1996) The NF-κB and IκB proteins: New Discoveries and Insights, *Annual Review of Immunology 14*, 649-681.

Beerman, T. A., and Lebowitz, J. (1973) Further analysis of the altered secondary structure of superhelical DNA. Sensitivity to methylmercuric hydroxide a chemical probe for unpaired bases, *Journal of Molecular Biology 79*, 451-470.

Berg, J. M., Tymoczko, J. L., and Stryer, L. (2002) *Biochemistry*, 5 ed., W H Freeman.

BIACORE (1997) *BIACORE AB*, BiaEvaluation 3.0.

BIACORE (1998) *BIACORE AB*, Biacore Technology Handbook.

Blakeley, M. P., Langan, P., Niimura, N., and Podjarny, A. (2008) Neutron crystallography: opportunities, challenges, and limitations, *Current Opinion in Structural Biology 18*, 593-600.

Blakeley, M. P., Teixeira, S. C. M., Petit-Haertlein, I., Hazemann, I., Mitschler, A., Haertlein, M., Howard, E., and Podjarny, A. D. (2010) Neutron macromolecular crystallography with LADI-III, *Acta Crystallographica Section D 66*, 1198-1205.

Bouteille, B., Oukem, O., Bisser, S., and Dumas, M. (2003) Treatment perspectives for human African trypanosomiasis, *Fundamental & Clinical Pharmacology 17*, 171-181.

Bragg, W. H., and Bragg, W. L. (1913) The Reflection of X-rays by Crystals, *Proceedings of the Royal Society of London. Series A 88*, 428-438.

Brünger, A. T. (1992) Free R value: a novel statistical quantity for assessing the accuracy of crystal structures, *Nature 355*, 472-475.

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR System: A New Software Suite for Macromolecular Structure Determination, *Acta Crystallographica Section D 54*, 905-921.

Brünger, A. T., Kuriyan, J., and Karplus, M. (1987) Crystallographic R factor refinement by molecular dynamics, *Science 235*, 458-460.

Cao, S., Zhang, X., Edwards, J. P., and Mosser, D. M. (2006) NF-κB1 (p50) Homodimers Differentially Regulate Pro- and Anti-inflammatory Cytokines in Macrophages, *Journal of Biological Chemistry 281*, 26041-26050.

Chargaff, E. (1950) Chemical specificity of nucleic acids and mechanism of their enzymatic degradation, *Experientia 6*, 201-209.

Chen, A. Y., Yu, C., Gatto, B., and Liu, L. F. (1993) DNA minor groove-binding ligands: a different class of mammalian DNA topoisomerase I inhibitors, *Proceedings of the National Academy of Sciences USA 90*, 8131-8135.

Chen, Y.-Q., Ghosh, S., and Ghosh, G. (1998) A novel DNA recognition mode by the NF-κB p65 homodimer, *Nature Structural & Molecular Biology 5*, 67-73.

Chidambaram, R., and Sikka, S. K. (2003) Neutron crystallography-Then and Now, *Current Science 85*, 871-877.

Clark, G. R., Gray, E. J., Neidle, S., Li, Y.-H., and Leupin, W. (1996) Isohelicity and Phasing in Drug - DNA Sequence Recognition: Crystal Structure of a Tris(benzimidazole) - Oligonucleotide Complex, *Biochemistry 35*, 13745-13752.

Collaborative Computational Project Number 4. (1994) The CCP4 suite: programs for protein crystallography, *Acta Crystallographica Section D 50*, 760-763.

Cramer, P., Larson, C. J., Verdine, G. L., and Müller, C. W. (1997) Structure of the human NF-κB p52 homodimer-DNA complex at 2.1 Å resolution, *European Molecular Biology Organisation Journal 16*, 7078-7090.

Dauter, Z. (1999) Data-collection strategies, *Acta Crystallographica Section D55*, 1703-1717.

Davies, D. R., and Baldwin, R. L. (1963) X-ray studies of 2 synthetic DNA copolymers, *Journal of Molecular Biology 6*, 251-255.

Delespaux, V., and de Koning, H. P. (2007) Drugs and drug resistance in African trypanosomiasis, *Drug resistance updates : reviews and commentaries in antimicrobial and anticancer chemotherapy 10*, 30-50.

De Scantis, D., and Leonard, G. (2011) *ID29 - Multiple-wavelength Anomalous Diffraction (MAD).* (Last accessed December 2011) Available: http://www.esrf.eu/UsersAndScience/Experiments/MX/About_our_beamlines/ID29

Ducruix, A., and Giege, R. (1992) *Crystallization of Nucleic acids and proteins: A practical approach*, Oxford University Press.

Dykstra, C. C., McClernon, D. R., Elwell, L. P., and Tidwell, R. R. (1994) Selective inhibition of topoisomerases from Pneumocystis carinii compared with that of topoisomerases from mammalian cells, *Antimicrobial Agents and Chemotherapy 38*, 1890-1898.

Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics, *Acta Crystallographica Section D 60*, 2126-2132.

Evans, P. (2006) Scaling and assessment of data quality, *Acta Crystallographica Section D 62*, 72-82.

Felsenfeld, G., Davies, D. R., and Rich, A. (1957) Formation of a three-stranded polynucleotide molecule, *Journal of the American Chemical Society 79*, 2023-2024.

Flot, D., Mairs, T., Giraud, T., Guijarro, M., Lesourd, M., Rey, V., van Brussel, D., Morawe, C., Borel, C., Hignette, O., Chavanne, J., Nurizzo, D., McSweeney, S., and Mitchell, E. (2010) The ID23-2 structural biology microfocus beamline at the ESRF, *Journal of Synchrotron Radiation 17*, 107-118.

Franke, D., and Svergun, D. I. (2009) DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering, *Journal of Applied Crystallography 42*, 342-346.

Fuller, W., Forsyth, T., and Mahendrasingam, A. (2004) Water-DNA interactions as studied by X-ray and neutron fibre diffraction, *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 359*, 1237-1248.

Fuller, W., Wilkins, M. H., Wilson, H. R., and Hamilton, L. D. (1965) The molecular configuration of deoxyribonucleic acid .IV. X-ray diffraction study of the A form., *Journal of Molecular Biology 12*, 60-76.

Ghosh, G., Duyne, G. V., Ghosh, S., and Sigler, P. B. (1995) Structure of NF-κB p50 homodimer bound to a κB site, *Nature 373*, 303-310.

Glusker, J. P., and Trueblood, K. N. (1985) *Crystal Structure Analysis: A primer*, 2 ed., Oxford University Press.

Hanh, T. (1987) *International Tables for Crystallography Volume A: Space-group symmetry*, Springer.

Hart, D. J., Speight, R. E., Blackburn, M. J.., Cooper, M. A., and Sutherland, J. D. (1999) The salt dependence of DNA recognition by NF-κB p50: A detailed kinetic analysis of the effects on affinity and specificity, *Nucleic Acids Research 27*, 1063-1069.

Hauptman, H. A., and Langs, D. A. (2003) The phase problem in neutron crystallography, *Acta Crystallographica Section A59*, 250-254.

Huang, D.-B., Phelps, C. B., Fusco, A. J., and Ghosh, G. (2005) Crystal Structure of a Free κB DNA: Insights into DNA Recognition by Transcription Factor NF-κB, *Journal of Molecular Biology 346*, 147-160.

Jacrot, B., and Zaccai, G. (1981) *Biopolymers 20,* 2413-2426.

Jean-Moreno, V., Rojas, R., Goyeneche, D., Coombs, G. H., and Walker, J. (2006) Leishmania donovani: Differential activities of classical topoisomerase inhibitors and antileishmanials against parasite and host cells at the level of DNA topoisomerase I and in cytotoxicity assays, *Experimental Parasitology 112*, 21-30.

Johnson, C. K. (1967) ORNL-4168, p 115.

Jonsson, U., Fagerstam, L., Ivarsson, B., Johnsson, B., Karlsson, R., Lundh, K., Lofas, S., Persson, B., Roos, H., and Ronnberg, I. (1991) Real-time biospecific interaction analysis using surface plasmon resonance and a sensor chip technology, *BioTechniques 11*, 620-627.

Kabsch, W. (2010) XDS, *Acta Crystallographica Section D 66*, 125-132.

Karlsson, R., and Fält, A. (1997) Experimental design for kinetic analysis of protein-protein interactions with surface plasmon resonance biosensors, *Journal of Immunological Methods 200*, 121-133.

Klingbeil, M. M., Drew, M. E., Liu, Y., Morris, J. C., Motyka, S. A., Saxowsky, T. T., Wang, Z., and Englund, P. T. (2001) Unlocking the secrets of trypanosome kinetoplast DNA network replication, *Protist 152*, 255-262.

Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J., and Svergun, D. I. (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis, *Journal of Applied Crystallography 36*, 1277-1282.

Kopka, M. L., Yoon, C., Goodsell, D., Pjura, P., and Dickerson, R. E. (1985) Binding of an antitumor drug to DNA: Netropsin and C-G-C-G-A-A-T-T-BrC-G-C-G, *Journal of Molecular Biology 183*, 553-563.

Krężel, A., and Bal, W. (2004) A formula for correlating pK_a values determined in D_2O and H_2O, *Journal of Inorganic Biochemistry 98*, 161-166.

Langridge, R., Wilson, H. R., Hooper, C. W., Wilkins, M. H. F., and Hamilton, L. D. (1960) The molecular configuration of deoxyribonucelic acid I. X-ray diffraction study of a crytsalline form of lithium salt, *Journal of Molecular Biology 2*, 19-37.

Lansiaux, A. l., Dassonneville, L., Facompré, M. l., Kumar, A., Stephens, C. E., Bajic, M., Tanious, F., Wilson, W. D., Boykin, D. W., and Bailly, C. (2002) Distribution of Furamidine Analogues in Tumor Cells: Influence of the Number of Positive Charges, *Journal of Medicinal Chemistry 45*, 1994-2002.

Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures, *Journal of Applied Crystallography 26*, 283-291.

Laughton, C. A., Tanious, F., Nunn, C. M., Boykin, D. W., Wilson, W. D., and Neidle, S. (1996) A Crystallographic and Spectroscopic Study of the Complex between d(CGCGAATTCGCG)$_2$ and 2,5-Bis(4-guanylphenyl)furan, an Analogue of Berenil. Structural Origins of Enhanced DNA-Binding Affinity, *Biochemistry 35*, 5655-5661.

Leal, R. M. F., Teixeira, S. C. M., Blakeley, M. P., Mitchell, E. P., and Forsyth, V. T. (2009) A preliminary neutron crystallographic study of an A-DNA crystal, *Acta Crystallographica Section F 65*, 232-235.

Leslie, A. (1999) Integration of macromolecular diffraction data, *Acta Crystallographica Section D 55*, 1696-1702.

Lu, X.-J., and Olson, W. K. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures, *Nucleic Acids Research 31*, 5108-5121.

Mallena, S., Lee, M. P. H., Bailly, C., Neidle, S., Kumar, A., Boykin, D. W., and Wilson, W. D. (2004) Thiophene-Based Diamidine Forms a "Super" AT Binding Minor Groove Agent, *Journal of the American Chemical Society 126*, 13659-13669.

Marvin, D. A., Spencer, M., Wilkins, M. H. F., and Hamilton, L. D. (1958) A new configuration of deoxyribonucleic acid., *Nature 182*, 387-388.

Mathis, A. M., Holman, J. L., Sturk, L. M., Ismail, M. A., Boykin, D. W., Tidwell, R. R., and Hall, J. E. (2006) Accumulation and Intracellular Distribution of Antitrypanosomal Diamidine Compounds DB75 and DB820 in African Trypanosomes, *Antimicrobial Agents and Chemotherapy 50*, 2185-2191.

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

McCarthy, A. A., Brockhauser, S., Nurizzo, D., Theveneau, P., Mairs, T., Spruce, D., Guijarro, M., Lesourd, M., Ravelli, R. B. G., and McSweeney, S. (2009) A decade of user operation on the macromolecular crystallography MAD beamline ID14-4 at the ESRF, *Journal of Synchrotron Radiation 16*, 803-812.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software, *Journal of Applied Crystallography 40*, 658-674.

McHugh, M. M., Woynarowski, J. M., Sigmund, R. D., and Beerman, T. A. (1989) Effect of minor groove binding drugs on mammalian topoisomerase I activity, *Biochemical Pharmacology 38*, 2323-2328.

McRee, D. E. (1999) *Practical protein crystallography*, 2 ed., Academic Press Publication.

Miao, Y., Cui, T., Leng, F., and Wilson, W. D. (2008) Inhibition of high-mobility-group A2 protein binding to DNA by netropsin: A biosensor-surface plasmon resonance assay, *Analytical Biochemistry 374*, 7-15.

Miao, Y., Lee, M. P. H., Parkinson, G. N., Batista-Parra, A., Ismail, M. A., Neidle, S., Boykin, D. W., and Wilson, W. D. (2005) Out-of-Shape DNA Minor Groove Binders: Induced Fit Interactions of Heterocyclic Dications with the DNA Minor Groove, *Biochemistry 44*, 14701-14708.

Morgan, A. R. (1970) Model for DNA Replication by Kornberg's DNA Polymerase, *Nature 227*, 1310-1313.

Müller, C. W., Rey, F. A., Sodeoka, M., Verdine, G. L., and Harrison, S. C. (1995) Structure of the NF-κB p50 homodimer bound to DNA, *Nature 373*, 311-317.

Munde, M., Lee, M., Neidle, S., Arafa, R., Boykin, D. W., Liu, Y., Bailly, C., and Wilson, W. D. (2007) Induced Fit Conformational Changes of a "Reversed Amidine" Heterocycle: Optimized Interactions in a DNA Minor Groove Complex, *Journal of the American Chemical Society 129*, 5688-5698.

Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of Macromolecular Structures by the Maximum-Likelihood Method, *Acta Crystallographica Section D 53*, 240-255.

Neidle, S. (1997) Crystallographic insights into DNA minor groove recognition by drugs, *Biopolymers 44*, 105-121.

Neidle, S. (2001) DNA minor-groove recognition by small molecules, *Natural Product Reports 18*, 291-309.

Neidle, S. (2007) *Principles of nucleic acid structure*, Academic Press.

Neidle, S., and M. Nunn, C. (1998) Crystal structures of nucleic acids and their drug complexes, *Natural Product Reports 15*, 1-15.

Neidle, S., Puvvada, M. S., and Thurston, D. E. (1994) The relevance of drug-DNA sequence specificity to antitumor activity, *European Journal of Cancer 30A*, 567-568.

Neidle, S., and Thurston, D. E. (1994) *DNA sequences as target for new anticancer agents*, CRC Press.

Nguyen, B., Hamelberg, D., Bailly, C., Colson, P., Stanek, J., Brun, R., Neidle, S., and Wilson, D. W. (2004) Characterization of a novel DNA minor-groove complex, *Biophysical Journal 86*, 1028-1041.

Nguyen, B., Neidle, S., and Wilson, W. D. (2008) A Role for Water Molecules in DNA-Ligand Minor Groove Recognition, *Accounts of Chemical Research 42*, 11-21.

Nguyen, B., Tardy, C., Bailly, C., Colson, P., Houssier, C., Kumar, A., Boykin, D. W., and Wilson, W. D. (2002) Influence of compound structure on affinity, sequence selectivity, and mode of binding to DNA for unfused aromatic dications related to furamidine, *Biopolymers 63*, 281-297.

Nunn, C. M., Garman, E., and Neidle, S. (1997) Crystal Structure of the DNA Decamer d(CGCAATTGCG) Complexed with the Minor Groove Binding Drug Netropsin *Biochemistry 36*, 4792-4799.

Otwinowski, Z., Minor, W., and Charles W. Carter, Jr. (1997) Processing of X-ray diffraction data collected in oscillation mode, in *Methods in Enzymology*, pp 307-326, Academic Press.

Pereira, S. G., and Oakley, F. (2008) Nuclear factor-κB1: Regulation and function, *The International Journal of Biochemistry & Cell Biology 40*, 1425-1430.

Perkins, N. D. (2006) Intergrating cell-signaling pathways with NF-κB and IKK function, *Nature Reviews 8*, 49-62.

Pernot, P., Round, A., and Zerrad, L. (2011) *BM29: bioSAXS beamline.* Available http://www.esrf.eu/UsersAndScience/Experiments/MX/About_our_beamlines/BM29 (Last accessed December 2011).

Pohl, F. M., and Jovin, T. M. (1972) Salt-induced co-operative conformational change of a synthetic DNA: Equilibrium and kinetic studies with poly(dG-dC), *Journal of Molecular Biology 67*, 375-396.

Popov, A. (2011) *ID23-1: Gemini - Macromolecular Crystallography.* Available: http://www.esrf.eu/UsersAndScience/Experiments/MX/About_our_beamlines/ID23-1 (Last accessed December 2011).

Rhodes, G. (2006) *Crystallography made crystal clear*, 3 ed., Academic Press Publications.

Seitz, O. (2003) DNA and RNA Binders. From Small Molecules to Drugs. Vols. 1 & 2. Edited by Martine Demeunynck, Christian Bailly and W. David Wilson, *Angewandte Chemie International Edition 42*, 4994-4994.

Semenyuk, A. V., and Svergun, D. I. (1991) GNOM - a program package for small-angle scattering data processing, *Journal of Applied Crystallography 24*, 537-540.

Sen, R., and Baltimore, D. (1986) Inducibility of κ immunoglobulin enhancer-binding protein NF-κB by a posttranslational mechanism, *Cell 47*, 921-928.

Shapiro, T. A., and Englund, P. T. (1995) The Structure and Replication of Kinetoplast DNA, *Annual Review of Microbiology 49*, 117-143.

Siebenlist, U., Franzoso, G., and Brown, K. (1994) Structure, regulation and function of NF-Kappa-B, *Annual review of Cell Biology 10*, 405-455.

Simpson, I. J., Lee, M., Kumar, A., Boykin, D. W., and Neidle, S. (2000) DNA minor groove interactions and the biological activity of 2,5-bis-[4-(N-alkylamidino)phenyl] furans, *Bioorganic & Medicinal Chemistry Letters 10*, 2593-2597.

Sines, C. C., McFail-Isom, L., Howerton, S. B., VanDerveer, D., and Williams, L. D. (2000) Cations Mediate B-DNA Conformational Heterogeneity, *Journal of the American Chemical Society 122*, 11048-11056.

Speight, R. E., Hart, D. J., and Blackburn, J. M. (2002) Distamycin A affects the stability of NF-κB p50-DNA complexes in a sequence-dependent manner, *Journal of Molecular Recognition 15*, 19-26.

Stein, N., and Ballard, C. (2009) Intensity to amplitude conversion using Ctruncate, *Acta Crystallographica Section A65*, s161.

Stenberg, E., Persson, B., Roos, H., and Urbaniczky, C. (1991) Quantitative determination of surface concentration of protein with surface plasmon resonance using radiolabeled proteins, *Journal of Colloid and Interface Science 143*, 513-526.

Svergun, D., Barberato, C., and Koch, M. H. J. (1995) CRYSOL - a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates, *Journal of Applied Crystallography 28*, 768-773.

Svergun, D. I., Sokolova, A. V., Konarev, P. V., and Volkov, V. V. (1994-2004) OLIGOMER - Computation of volume fractions of mixtures with known scattering intensities from the components.

Thanos, D., and Maniatis, T. (1992) The High Mobility Group protein HMG I(Y) is required for NF-κB-dependent virus induction of the human IFN-β gene, *Cell 71*, 777-789.

Thurston, D. E. (2006) *Chemistry and Pharmacology of Anticancer Drugs*, 1 ed., CRC Press.

Trent, J. O., Clark, G. R., Kumar, A., Wilson, W. D., Boykin, D. W., Hall, J. E., Tidwell, R. R., Blagburn, B. L., and Neidle, S. (1996) Targeting the Minor Groove of DNA: Crystal Structures of Two Complexes between Furan Derivatives of Berenil and the DNA Dodecamer $d(CGCGAATTCGCG)_2$, *Journal of Medicinal Chemistry 39*, 4554-4562.

van Aalten, D. M., Bywater, R., Findlay, J. B., Hendlich, M., Hooft, R. W., and Vriend, G. (1996) PRODRG, a program for generating molecular topologies and unique molecular descriptors from coordinates of small molecules, *Journal of Computer Aided Molecular Design 10*, 255-262.

van de Sande, J. H., Ramsing, N. B., Germann, M. W., Elhorst, W., Kalisch, B. W., von Kitzing, E., Pon, R. T., Clegg, R. C., and Jovin, T. M. (1988) Parallel stranded DNA, *Science 241*, 551-557.

Vega, M. C., Garcia Saez, I., Aymami, J., Eritja, R., Gijs, A., Marel, V. D., Van Boom, J. H., Rich, A., and Coll, M. (1994) Three dimensional crytsal structure of the A-tract DNA dodecamer d(CGCAAATTTGCG) complexed with the minor groove binding drug Hoechst 33258, *European Journal of Biochemistry 222*, 721-726.

Vriend, G. (1990) WHAT IF: A molecular modeling and drug design program, *Journal of Molecular Graphics 8*, 52-56.

Wang, A. H.-J., Quigley, G. J., Kolpak, F. J., Crawford, J. L., van Boom, J. H., van der Marel, G., and Rich, A. (1979) Molecular structure of a left handed double helix DNA fragment at atomic resolution, *Nature 282*, 680-686.

Wang, J. C. (1985) DNA Topoisomerases, *Annual Review of Biochemistry 54*, 665-697.

Watson, J. D., and Crick, F. H. C. (1953) Molecular structure of nucelic acid - A structure of deoxyribose nucleic acid, *Nature 171*, 737-738.

Wilson, W. D., Nguyen, B., Tanious, F. A., Mathis, A., Hall, J. E., Stephens, C. E., and Boykin, D. W. (2005) Dications That Target the DNA Minor Groove: Compound Design and Preparation, DNA Interactions, Cellular Distribution and Biological Activity, *Current Medicinal Chemistry - Anti-Cancer Agents 5*, 389-408.

Wilson, W. D., Tanious, F. A., Mathis, A., Tevis, D., Hall, J. E., and Boykin, D. W. (2008) Antiparasitic compounds that target DNA, *Biochimie 90*, 999-1014.

Wing, R., Drew, H., Takano, T., Broka, C., Tanaka, S., Itakura, K., and Dickerson, R. E. (1980) Crystal structure analysis of a complete turn of B-DNA, *Nature 287*, 755-758.

Winn, M. D., Isupov, M. N., and Murshudov, G. N. (2001) Use of TLS parameters to model anisotropic displacements in macromolecular refinement, *Acta Crystallographica Section D 57*, 122-133.

Wood, A. A., Nunn, C. M., Czarny, A., Boykin, D. W., and Neidle, S. (1995) Variability in DNA minor groove width recognised by ligand binding: the crystal structure of a bis-benzimidazole compound bound to the DNA duplex d(CGCGAATTCGCG)$_2$, *Nucleic Acids Research 23*, 3678-3684.

Woynarowski, J. M., Sigmund, R. D., and Beerman, T. A. (1989) DNA minor groove binding agents interfere with topoisomerase II mediated lesions induced by epipodophyllotoxin derivative VM-26 and acridine derivative m-AMSA in nuclei from L1210 cells, *Biochemistry 28*, 3850-3855.

Yu, Z., Zhou, D., Cheng, G., and Mattson, M. (2000) Neuroprotective role for the p50 subunit of NF-κB in an experimental model of Huntingtons disease, *Journal of Molecular Neuroscience 15*, 31-44.