THE UNIVERSITY OF WARWICK
Department of Chemistry

INTERACTION OF DNA WITH GROOVE BINDING LIGANDS

Thesis Submitted for the Degree of Doctor of Philosophy

By

Mark John Broome

May 2008

Acknowledgements

I am indebted to a number of people for their assistance and support during this research. First and foremost I would like to thank Prof. Mark Rodger for his tuition and support throughout my time at Warwick. I am also very grateful to Prof. Alison Rodger, Prof. Michael Hannon , and Prof. Tiffany Walsh for sharing their excitement for science and the interaction and support with their research groups.

My thanks also goes to the members of the Computer Modelling Group past and present for their regular support and friendship. Also, a special thank you to Dr. Syma Khalid for introducing me to DNA, a fascinating puzzle that continues to "attract" me and keep my interest "bound" as we try to "uncoil" its secrets.

A special thanks to my family for their continued support and to Nicky for her patience friendship and shared excitement for learning.

I would also like to thank and acknowledge the EPSRC for the funding of this project.

Abstract

Synthetic molecules that target the major groove in a sequence-selective way are a major goal in molecular medicine. Recently a major step has been taken toward achieving this goal: a novel cylinder has been developed that binds strongly into the major groove of DNA. Experimental techniques have provided some information regarding the binding strength and preferred binding sites of the cylinder on DNA. From all the experimental data it is clear that the parent cylinder binds in the major groove and is able to induce dramatic conformational changes in the DNA; these are unprecedented effects with synthetic DNA binders. However, gaining molecular level information in such a macromolecular system is challenging. Molecular dynamics (MD) simulations can provide information at the molecular level that is complementary to experiment and therefore are an ideal way to get a better understanding of this system. In this work we present the results of various MD simulations designed to probe the DNA-cylinder system.

We have studied the effect of using CHARMM22 and CHARMM27 as the force-field for the simulations. Results showed that uncomplexed DNA simulated with CHARMM22 was less stable in the B-form than the comparable strand of DNA simulated with CHARMM27.

Investigations into the effects of the cylinders charge and shape are also reported. Multi-nanosecond simulations were performed using two related synthetic cylinders, one with two Fe(II) metal centers and the other with two Cu(I) centers, and DNA.

Finally the role of DNA within the system was investigated by performing a series of simulations of the cylinders with $d(ATATATATATAT)_2$, $d(CGCGCGCGCGCG)_2$ and d(CGCGCATATACG). Simulations with these DNA strands has only produced one system (C_{Cu}^{2+} with d(ATATATATATAT)) where the cylinder causes a conformational change in the DNA.

Declaration

The observations and recommendations in this thesis are those of the author except where acknowledgement has been made to results and ideas previously published. The work was undertaken at the Department of Chemistry, University of Warwick between October 2002 and September 2008 and has not been previously submitted for a degree at any institution.

1 DNA structure, interaction and solvation environment

"We wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A). This structure has novel features which are of considerable biological interest.[1]" With this famous statement James Watson and Francis Crick announced their discovery of the structure of DNA in Nature on 25th April 1953. Their model, based on Rosalind Franklin's X-ray diffraction patterns from DNA crystals, finally explained how genetic material could be copied and passed on to future generations. It seeded a revolution which we are a part of today.

The complete set of instructions for making an organism is called its genome. It contains the master blueprint for all cellular structures and activities for the lifetime of the cell or organism. Found in every nucleus of a persons many trillions of cells, the human genome consists of tightly coiled threads of DNA and associated protein molecules, organised into structures called chromosomes. If unwound and tied together, the strands of DNA would stretch more than 5 feet in length. For each organism, the components of these slender threads encode all the information necessary for building and maintaining life, from simple bacteria to remarkably complex human beings.

Many biological functions are activated and deactivated depending on the conformation of the DNA. During this research computational methods have been used to investigate the interaction of DNA with potential groove binding synthetic cylinders. The following chapter is an overview from the literature documenting the many techniques that are being used to advance our understanding of the complex and interesting systems related to this field of research. All of the techniques and results summarised in this section have been included to aid the development of meaningful hypotheses built on substantiated scientific research with input from many perspectives capturing both practical and theoretical learning. Chapters two and

three outline the theory behind the methods being used in this research while Chapters four and five present the results in the context of the original hypotheses.

1.1 DNA

In humans, as in other higher organisms, a DNA dimer consists of two strands that wrap around each other to resemble a twisted ladder with sides made of sugar and phosphate molecules connected by "rungs" of nitrogen-containing chemicals, called nucleic acid bases. Each strand is a linear arrangement of similar repeating units called nucleotides which are each composed of one sugar, one phosphate, and a nitrogenous base.

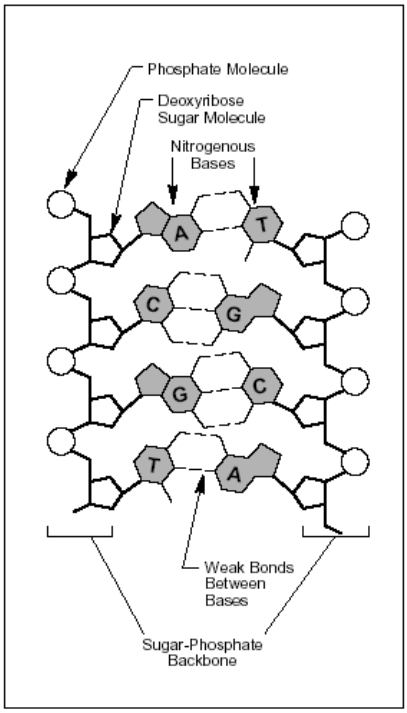


Figure 1-1 DNA Structure. The four nitrogenous bases of DNA are arranged along the sugar phosphate backbone in a particular order (the DNA sequence), encoding all genetic instructions for an organism. Adenine (A) pairs with thymine (T), while cytosine (C) pairs with guanine (G). The two DNA strands are held together by H-bonds between the bases.

1.1.1 Nucleic acid bases

Four different bases are present in DNA – adenine (A), thymine (T), cytosine (C), and guanine (G). The particular order of the bases arranged along the sugar-phosphate backbone is called

the DNA sequence; the sequence specifies the exact genetic instructions required to create a particular organism with its own unique traits.

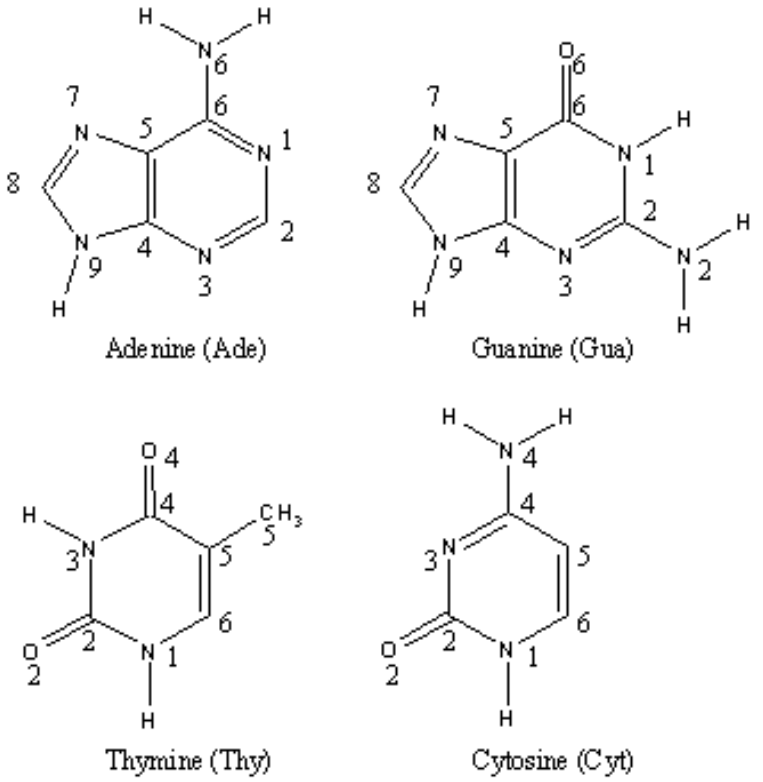


Figure 1-2 Adenine (A), Guanine (G), Thymine (T) and Cytosine (C). This figure includes the numbering system used throughout this work.

The two DNA strands are held together by hydrogen bonds between specific pairs of bases. Two H-bonds form readily between A and T, and the hydrogen bonded pair is then denoted A·T or T·A, depending on which is associated with the first strand. Similarly, three H-bonds will form between G and C to give a G·C or C·G base pair (Figure 1-3). All four of these base pairs readily fit within the phosphate backbone (Figure 1-4) of the DNA double helix.

Adenine and guanine are the larger of the nucleic acids and are members of a class of doubly ringed structures call purines; the smaller nucleic acids cytosine and thymine are members of a class of singly-ringed chemical structures called pyrimidines. When forming Watson Crick

base pairs (Figure 1-3) Purines are only complementary with pyrimidines: pyrimidine-pyrimidine pairings are energetically unfavourable because the molecules are too far apart for hydrogen bonding to be established; purine-purine pairings are energetically unfavourable because the molecules are too close, leading to electrostatic repulsion. The only other possible pairings are GT and AC; these pairings are mismatches because the pattern of hydrogen donors and acceptors do not correspond. In addition to the usual Watson-Crick pairs a number of non-canonical base pairs can be found. The majority of these alternative base pairs are aintiparallel with the bases in the *anti* configuration relative to the ribose. The most common are the GU wobble, the Sheared GA pair, the Reverse Hoogsteen pair and the GA imino pair.[2]

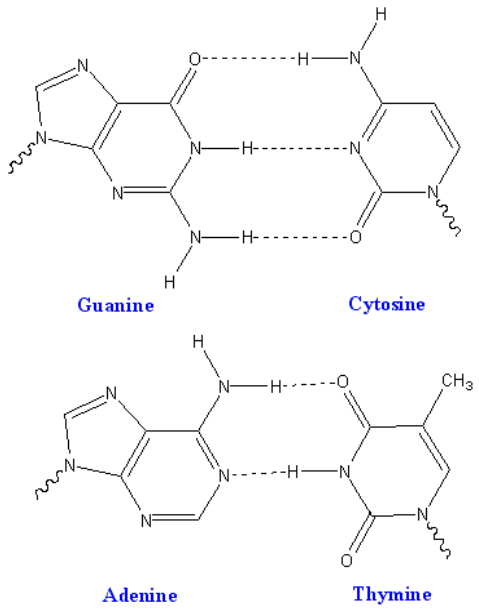


Figure 1-3 Watson Crick base pairs. A GC base pair demonstrating three intermolecular hydrogen bonds and an AT base pair demonstrating two intermolecular hydrogen bonds.

1.1.2 Phosphate backbone

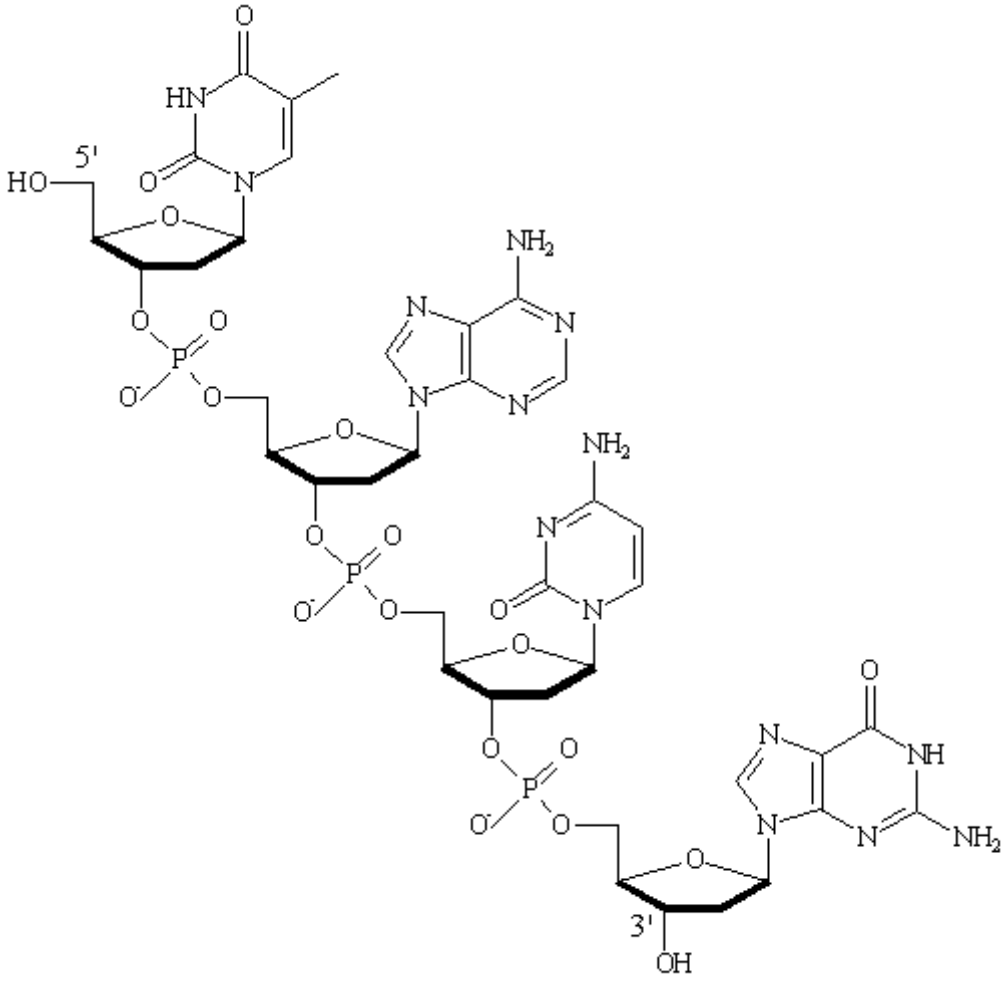


Figure 1-4 Phosphate backbone linked by phosphodiester linkages through sugar units containing individual bases.

The basic structure of DNA can be divided into two portions: the external sugar-phosphate backbone, and the internal bases. The sugar-phosphate backbone, as its name implies, is the major structural component of the DNA molecule. The backbone is constructed from alternating ribose sugar and phosphate molecules which are highly polar. With a polar hydrophilic backbone surrounding a core of hydrophobic bases the surrounding environment is important to the stability and structure of DNA.

The phosphate groups have a negative charge, giving the DNA a concentration of negative charge on the backbone. In vivo metallic ions[3] such as Na^{+}, Ca^{2+} Mg^{2+} and polyamides such as spermine and spermidine[4] bind along the phosphate backbone and help neutralise some of the charge. The charge is also neutralised by DNA-binding proteins that contain the positively charged amino acids lysine and arginine, which are attracted to the negatively charged phosphate backbone.

1.1.3 Directionality

The chemical convention of naming carbon atoms in the nucleotide sugar-ring numerically gives rise to a 3′ end and a 5′ end (Figure 1-4). The relative positions of structures along a strand of nucleic acid are usually noted as being upstream (towards the 5′ end) or downstream (towards the 3′ end).

To form a chemically stable structure, the two DNA strands in a duplex are antiparallel. That is, one strand running from the 5′-phosphate to 3′-OH is paired with the other strand arranged with its 3′-OH opposite the 5′-phosphate of the first strand, and its 5′-phosphate opposite the 3′-OH of the first strand.

1.1.3.1 3′ end

The 3′ end of a strand is so named due to it terminating at the hydroxyl group of the third carbon in the sugar-ring (Figure 1-5) and is known as the tail end. The 3′-hydroxyl is necessary in the synthesis of new nucleic acid molecules as it is ligated to the 5′-phosphate of a separate nucleotide, allowing the formation of strands of linked nucleotides.

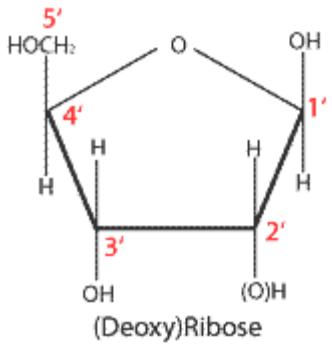


Figure 1-5 A furanose (sugar-ring) molecule with carbons labelled numerically.

Molecular biologists can use nucleotides that lack a 3′-hydroxyl (dideoxyribonucleotides) to interrupt the replication of DNA. This technique is known as the Dideoxy termination method and was used to work out the order of nucleotides in DNA.

1.1.3.2 5′ end

The 5′ end is so named as the strand terminates at the chemical group attached to the fifth carbon in the sugar-ring. If a phosphate group is attached to the 5′ end, ligation of two nucleotides can occur, via a phosphodiester bond from the 5′-phosphate to the 3′-hydroxyl group of another nucleotide. If it is removed no ligation can occur. Molecular biologists can use this phenomenon to their advantage by removing the 5′-phosphate with a phosphatase to prevent any unwanted nucleic acid ligation.

1.1.4 Sugar Pucker

The five membered deoxyribose sugar ring in DNA is inherently non-planar. This non-planarity is termed puckering. The precise conformation of a deoxyribose ring can be completely specified by the five endocyclic torsion angles within it, Figure 1-7 **Backbone torsion angles in nucleic acid structures and the atom labels for sugar puckering configurations.** The ring puckering arises from the effect of non-bonded interactions between

substituents at the four ring carbon atoms, the energetically most stable conformation for the ring has all substituents as far apart as possible. The puckering can be described by either a simple qualitative description of the conformation in terms of atoms deviating from ring coplanarity, or precise descriptions in terms of the ring internal torsion angles.

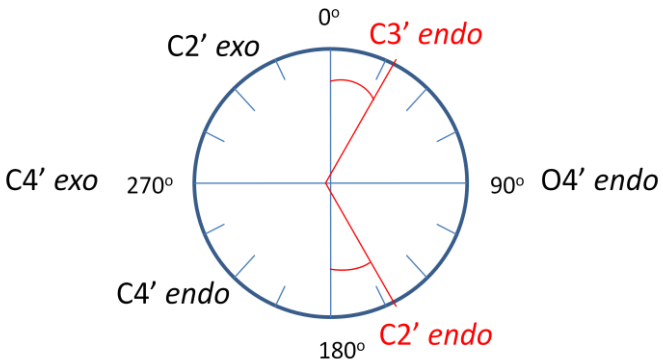
In principle, there is a continuum of interconvertible puckers, separated by energy barriers. These various puckers are produced by systematic changes in the ring torsion angles. The puckers can be defined by the parameters P and v_m. The value of P is defined in terms of the five torsion angles v_0-v_4.

$$\tan P = \frac{(v_4 + v_1) - (v_3 + v_0)}{2 \times v_2 \times (\sin 36^o + \sin 72^o)}$$

And the maximum degree of pucker, v_m, by

$$v_m = \frac{v_2}{\cos P}$$

The pseudorotation phase angle can take any value between 0^o and 360^o. If v_2 has a negative value, then 180^o is added to the value of P. The pseudorotation phase angle can be represented by the pseudorotation wheel, which indicates the continuum of ring puckers.



1-6 The pseudorotation wheel for a deoxyribose sugar. The red angles indicate the preferred ranges of the pseudorotation angle for the two principle sugar conformations.

Values of v_m indicate the degree of puckering of the ring: typical experimental values from crystallographic studies on mononucleosides are in the range 25-45°. The five internal torsion

angles are not independent of each other, and so to a good approximation any one angle, v_j, can be represented in terms of just two variables.

$$v_j = v_m \cos[P + 0.8\pi\ (j - 1)]$$

A large number of distinct deoxyribose ring pucker geometries have been observed experimentally, by X-ray crystallography and NMR techniques. When one ring atom is out of the plane of the other four, the pucker type is an 'envelope' one. More commonly, two atoms deviate from the plane of the other three, with these two either side of the plane. It is usual for one of the two atoms to have a larger deviation from the plane than the other in this, the 'twist' conformation. The direction of atomic displacement from the plane is important. If the major deviation is on the same side as the base and C4'-C5' bond, then the atom involved is termed *endo*. If it is on the opposite side, it is called *exo*. The most commonly observed puckers in crystal structures of isolated nucleosides and nucleotides are close to either C2'*endo* or C3'*endo* types. In practice, these pure envelope forms are rarely observed, largely because of the differing substituents on the ring. Consequently the puckers are then best described in terms of twist conformations. When the major out of plane deviation is on the *endo* side, there is a minor deviation on the opposite, *exo* side. The convention used for describing a twist deoxyribose conformation is that the major out of plane deviation is followed by the minor one, for example C2' *endo*, C3' *exo*. The C2'*endo* family of puckers have P values in the range 140-185^o in view of their position on the pseudorotation wheel, they are sometimes termed S (south) conformations. The C3'*endo* domain has P values in the range –10 to +40^o, and can be termed N (north).

The pseudorotation wheel implies that deoxyribose puckers are free to inter-convert. In practise, there are energy barriers between major forms that is dependent on the route around the pseudorotation wheel. For interconversion of C2'*endo* to C3'*endo* the preferred pathway is via

the 04'*endo* state, with a barrier of 8-21kJ/mol representing the total free energy for interconversion.

Relative population of puckers can be monitored directly by NMR measurements of the ratio of coupling constants between H1'-H2' and H3'-H4' protons. These show that in contrast to the 'frozen out' puckers found in the solid state structures of nucleosides and nucleotides, there is a rapid interconversion in solution. Nonetheless, the relative populations of the major puckers are dependent on the type of base attached. Purines show a preference for the C2'*endo* pucker conformational type whereas pyrimidines favour C3'*endo* form.

Correlations have been found, from crystallographic and NMR studies, between pucker and several backbone conformational variables, both in isolated nucleosides/nucleotides and in oligonucleotide structures. These are discussed later in this chapter. Changes in sugar pucker are important in polynucleotides because they can alter the orientation of C1', C3', and C4' substituents, resulting in major changes in backbone conformation and overall structure. Sugar pucker is thus an important determinant of polynucleotide conformation.

The A- and B-DNA forms differ mainly in their sugar pucker (see Figure 5-8). In the A form, the C3′ configuration is above the sugar ring, whilst C2′ is below it making the correct description C3′-endo. Likewise, in the B form, the C2′ configuration is above the sugar ring, whilst C3′ is below; this is called C2′-endo. The effect of the sugar puckering in A-DNA results in a shortening of the distance between adjacent phosphates by around one angstrom. This gives 11 to 12 base pairs to each helical turn in the DNA strand, instead of 10.5 in B-DNA. These parameters and those defined in Figure 1-7 are used extensively in the results section as indicators of the DNA response to the presence of the cylinders and to investigate the hypothesis.

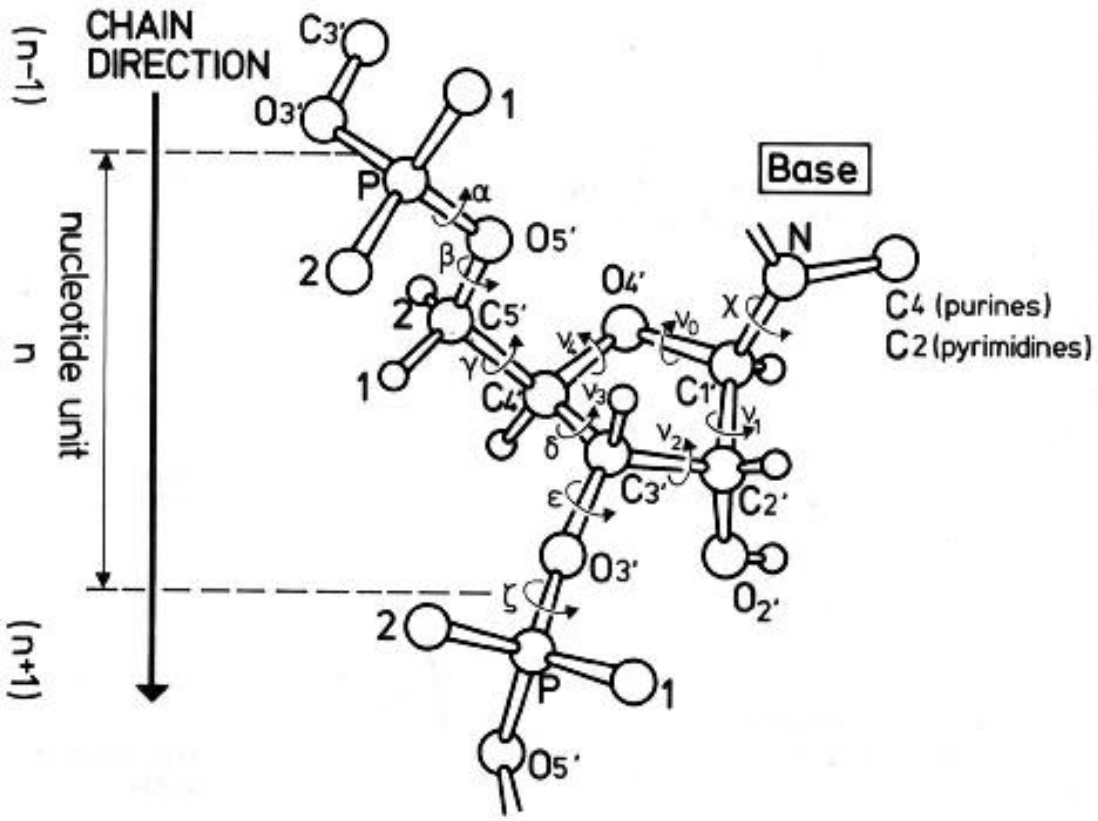


Figure 1-7 Backbone torsion angles in nucleic acid structures[5] and the atom labels for sugar puckering configurations.

1.1.5 Major and minor grooves

The average B-DNA structure of DNA is described as having a "major groove" and "minor groove". In DNA the helix is offset so the two grooves that form between the twisted backbone are not the same; one is bigger (major) than the other (minor). Although the dimensions of the major and minor grooves are different for each of the three helix types, A, B and Z, relative to the bases, the major groove is always on the same side for a given base pair, see Figure 1-8. In B-DNA the major groove is wider (12 versus 6Å) and deeper (8.5 versus 7.5Å) than the minor groove, making it more accessible to interacting molecules[6].

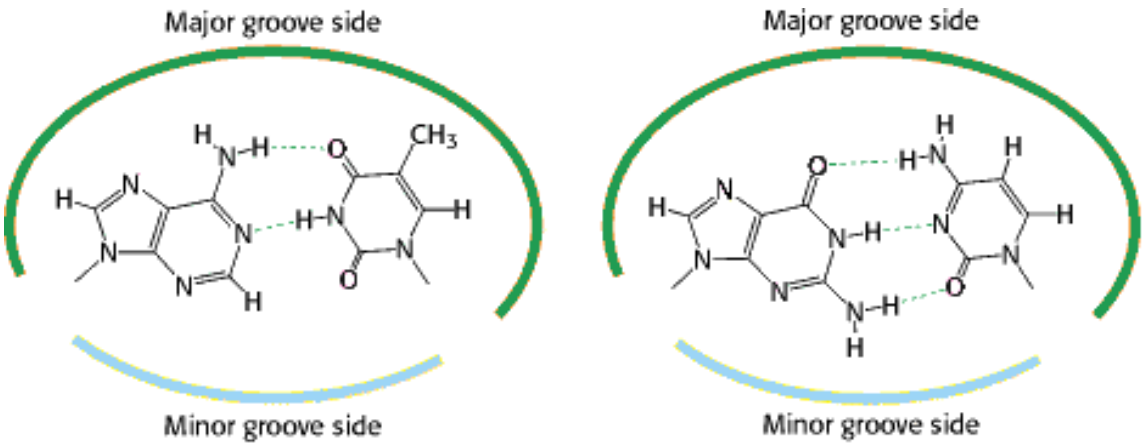


Figure 1-8 Base pairs and their positions relative to the major and minor groove.

The sugars are closer to one side of the base pair than the other. There is less space on the side between the sugars. The convention is that the side closest to the sugars is called the minor groove side.

For B-DNA helices, proteins binding in the major groove usually bind to specific sequences, often through the insertion of a protein α helix into the major groove. In addition, the major groove of B-DNA is approximately the correct width to accommodate a third base pair, usually a pyrimidine, and so triplex DNA structures are sometimes found to form.

Proteins that bind DNA non-specifically, such as chromatin proteins, will often bind to DNA in the minor groove. In addition, water molecules and small ions are able to bind and stabilise the minor groove. In A-DNA the minor groove is almost the same size as the major groove, while in Z-DNA, the minor groove is deep and narrow, and the major groove is almost nonexistent. The various modes of binding and their experimental derivation are explained in more detail later on.

1.2 DNA Secondary Structure

Double stranded DNA can assume one of three types of secondary structure, type A, B or Z.

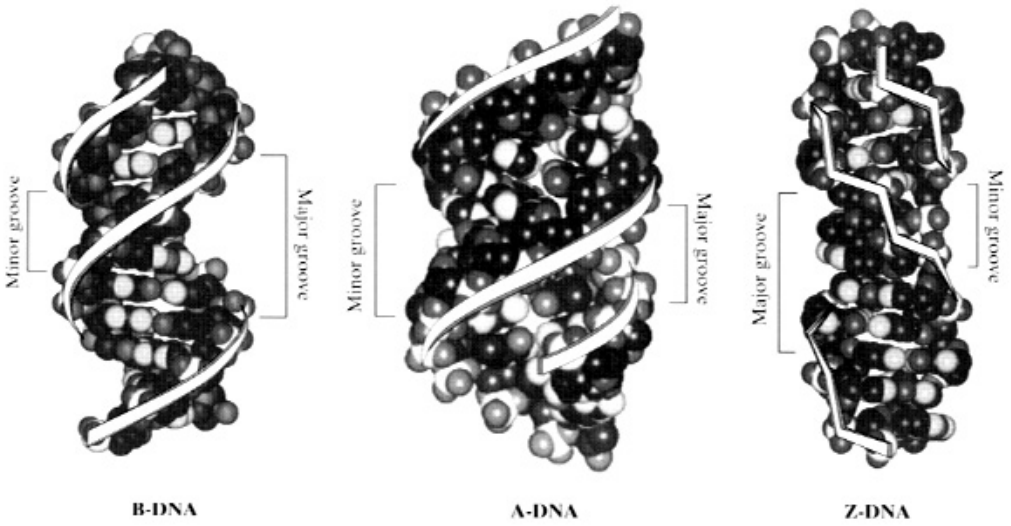


Figure 1-9 The three forms of DNA, A, B and Z.

Crystallographic studies [7,8,9] have been used to investigate DNA conformation for many years showing that DNA can have several conformations. More recently the techniques of Atomic force [10] (AFM) and scanning tunneling microscopy [11] (STM) have been used to measure features of the DNA duplex[12] and other related structures.[13],[14] Nuclear magnetic resonance spectroscopy [15] (NMR) is another powerful technique that enables geometrical details of the structure of DNA on an atomic level to be elucidated, binding modes of a variety of DNA binding complexes[16] have been investigated using variations on this technique.[17] Information regarding the orientation of DNA binders can also be accessed using linear dichroism (LD). During flow linear dichroism the orientation of the binding molecule relative to the linearly arranged DNA can be calculated[18]. Circular dichroism (CD) also complements the suite of analytical techniques available for analysis of DNA with its chiral nature.

1.2.1 B-DNA

In summary B-DNA is the most common form of DNA in solution and exists as an antiparallel helix. It is a right-handed helix with a wide major groove and narrow minor groove. Each base

pair extends the helical length by about 3.3Å making the length of a complete turn of the helix, ten base pairs, to be approximately 33Å. The structural properties for B-DNA, A-DNA and Z-DNA can be seen summarized in Table 1-1.

Property	A-Form	B-Form	Z-Form
Helical sense	Right-handed	Right-handed	Left-handed
Rotation/base pair (degrees)	33.6	35.9 +/- 4.2	-60/2
Mean base pairs/turn	10.7	10.0 +/- 1.2	12
Inclination of base normals to helix axis (degrees)	+19	-1.2 +/- 4.1	-9
Rise/base pair along helix axis (Å)	2.3	3.32 +/- 0.19	3.8
Pitch/turn of helix (Å)	24.6	33.2	45.6
Mean propeller twist (degrees)	+18	+16 +/- 7	~0
Sugar pucker conformation	C3'-endo	O1'-endo to C2'-endo	C2'-endo at C, C2'-exo to C1'-exo at G

Table 1-1 Parameters of A-, B-, and Z-DNA Double Helices.[19]

1.2.2 A-DNA

A-DNA is a wider right-handed helix with a shallow, wide minor groove and a narrower, deeper major groove. A-DNA is often seen to occur under non-physiological conditions in dehydrated samples of DNA, while in the cell it may be produced in hybrid pairings of DNA and RNA strands, as well as in enzyme-DNA complexes. The A form can be seen where consecutive base pairs are of the same type. In cases of higher hydration levels adoption of the A-conformation by purine-pyrimidine stretches is less unfavourable than for other possible sequences. It appears that when at least four purines, or four pyrimidines, are present in a row, formation of local A-DNA helix is less unfavourable[20], although certain purine stretches are more likely to form A-DNA than others[21]. It is thus possible to have a DNA sequence that contains some regions in the A-form within the context of a mainly B-conformation. These effects on the stability of the DNA form will be investigated during the MD simulations in this research.

The A-DNA helix is a wider structure than B-DNA, and Z-DNA. This is mainly due to the fact that the base pairs stack nearly on top of each other in B-DNA, but stack a little off-centre in the A-conformation. For this reason, a view down the centre of the A-conformation sees a hole that is absent in the two other helical conformations (Figure 1-10). As might be expected A-DNA is also more rigid than B-DNA, again because of the off-centre stacking of the bases makes them less flexible.

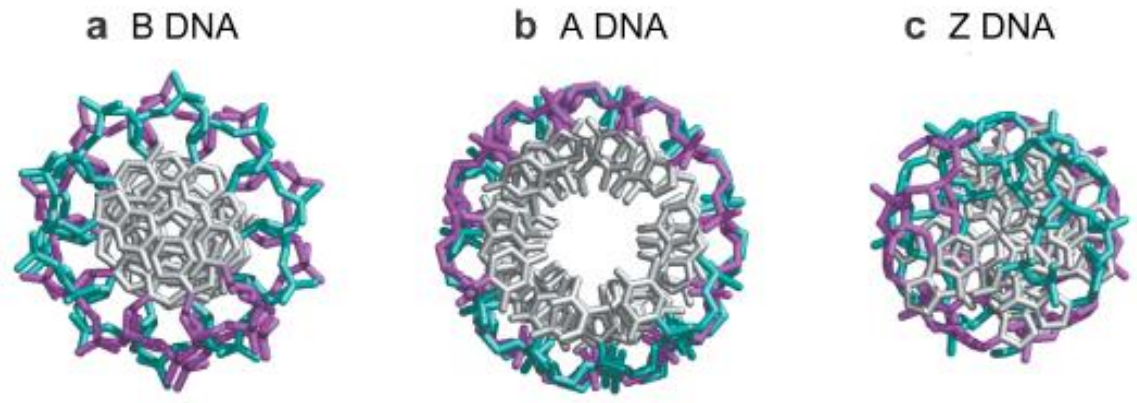


Figure 1-10 A view down the centre of A- B- and Z-DNA

Another contributing factor to the different conformations concerns the stacking of the base pairs in a non parallel manner. This base-pair tilt is higher in A-DNA than in B-DNA. An A-helix is the common form for DNA-RNA hybrids, as well as double stranded RNA; this is due to the extra OH group on the ribose sugar, which cannot fit easily into the tight space allotted to it in B-DNA.

1.2.3 Z-DNA

A more unusual form of DNA where the bases have been modified by methylation is the left-handed Z-DNA. This unique type of DNA forms under sequence-dependant conditions that require an alternating purine-pyrimidine sequence. Other chemical environmental factors favour the formation of Z-DNA such as high salt concentration, the presence of some cations,

and DNA supercoiling. Z-DNA forms excellent crystals a property that made it the first crystal structure of a DNA molecule to be solved[22].

The A-and B-DNA families are right handed helices, while the Z-DNA family has a left-handed orientation of the helix. These different forms for the DNA helix have important biological functions with the B-form being the most common under the conditions found in cells. The backbone of Z-DNA is not a smooth helix, but is irregular and zigzag in shape; hence its name. The Z helix is narrower than the A-and B-conformations, and it has 12 bp per turn. The nucleotide bases are flipped upside down, relative to the phosphate backbone, in Z-DNA when compared with A-DNA and B-DNA. While the structures of Z-DNA are uncommon and won't be considered during this work an awareness of their existence as an example of the varied nature of DNA is valuable. Biologically this unusual structure can be recognised by specific Z-DNA binding proteins and may be involved in the regulation of transcription.[23]

1.2.4 Base-stacking interactions

From an energetic point of view, the most important contribution to the DNA helix is the stacking of the bases on top of each other. The stacking energy is a measure of how much energy is required to destack or melt a region of double-stranded DNA. Table 1-2 lists the stacking energies for all 16 different dinucleotide combinations. There is a strong sequence dependence on the amount of stabilizing energy from base stacking. As a general trend, alternating pyrimidine-purine steps have less energy, and in particular T·A steps have the lowest (-3.82 kcal mol^{-1}). G·C steps have the largest value (-15kcal mol^{-1}), and require the most energy to melt.

Dinucleotide step	**Stacking energy kcal mol^{-1}**	**Twist angle (^{o})**	**Propeller twist (^{o})**
AA	-5.37	35.6	-18.66

AC	-10.51	34.4	-13.10
AG	-6.78	27.9	-14.00
AT	-6.57	32.1	-15.01
CA	-6.57	34.5	-9.45
CC	-8.26	33.7	-8.11
CG	-9.61	29.8	-10.03
CT	-6.78	27.9	-14.00
GA	-9.81	36.9	-13.48
GC	-14.59	40.00	-11.08
GG	-8.26	33.7	-8.11
GT	-10.51	34.4	13.10
TA	-3.82	36.0	-11.85
TC	-9.81	36.9	-13.48
TG	-6.57	34.5	-9.45
TT	-5.37	35.6	-18.66
Average	-7.92 +/- 2.57	35.7 +/- 8.0	-12.60 +/- 3.2

Table 1-2 Dinucleotide stacking energies[24] and helical parameters[25] for B-DNA

Another important dinucleotide parameter related to base-stacking interactions is the twist angle of the two bases (Figure 3-5). Depending on the base pair sequence being observed the values for B-DNA in solution range from 28^{o}, corresponding to a helix with 12.9 base pairs per turn, through to 40^{o}, corresponding to 9 base pairs per turn. It is therefore apparent that the pitch of the helix, the length of a full turn, can range considerably depending on the sequence. This has important biological consequences. For example if there are two protein-binding sites, each facing the same side of the DNA helix, roughly two turns of the helix apart, it is possible to have different intervening sequences resulting in quite different orientations of the two sites, ranging over about 45^{o} of orientation relative to each other.

In addition to the contribution of base stacking to the DNA structure hydrophobic and hydrophilic interactions are also involved. Both the sugar and phosphate which constitute the DNA backbone are quite soluble in water. However, the DNA bases which are in the middle of the helix are relatively hydrophobic, and hence insoluble. Since the bases are flat, they stack on top of each other in order to form a more hydrophobic 'mini-environment'. The bases twist slightly in order to maximise their hydrophobic interactions with each other, and it is this twisting of the stacked bases that gives rise to a helix. Free adenine molecules, on their own, will spontaneously stack on top of themselves to form single-stranded helices in solution. Thus hydrophobicity is a contributing factor of DNA forming a helix.

If the hydrophobic effects stated above were the only stabilising force to the DNA helix then there is no reason why the structure couldn't be columnar in nature. However, contributing to the driving forces that make the DNA structure helical are substantial electrostatic interactions. The planar components of the nucleic bases have positive edges (the hydrogens) and negative middles (the rings) and so as a result stacking is offset so that the positive edge of one base aligns with the negative middle of the next.

1.3 Water and DNA

Nucleic acid macromolecules are highly charged made up of many polar atoms on the sugar-phosphate backbone and heterocyclic bases. The structure of nucleic acid results from three main contributing factors. Electrostatic forces have a major role due to the negatively charged phosphates. Stacking interactions between base pairs are as a result of hydrophobic and dispersion forces as well as hydrogen bonding interactions between the polar atoms of the bases and water molecules. The conformational energy of the sugar-phosphate backbone also contributes to the nucleic acid structures. In its most stable conformation the polynucleotide

backbone exposes the negatively charged phosphates to the solvent and promotes the stacked helical arrangement of adjacent bases. The hydrophobic core created promotes hydrogen bond formation between the nucleic acid bases along with additional sugar-base[26] interactions.

The equilibrium conformation displays a sugar-phosphate backbone which interacts favourably with the molecules of the solvent. These two very different regions within DNA, and the interaction between solvent and nucleic acid structure, is the physiochemical basis for DNA polymorphism. The influence of water on the helical structure is experienced by all atoms in contact with the water, while the internal atoms participating in hydrogen bonding between base pairs are protected from the solvent. Considering these interactions the role of the water molecule in contributing to the overall stability of the helical conformation of nucleic acid is by screening the charges of the phosphates, bonding to and bridging between the polar exocyclic atoms of the bases, and influencing the conformations of residues with methyl groups via hydrophobic interactions. When considering the periodicity and helical structure of nucleic acids, water sites and water bridges involving polar base atoms or phosphate oxygens lead to structured arrangements of water molecules.

Nucleic acid hydration is seen to display water binding sites and water bridges repeatedly throughout the structure. The anionic phosphate oxygen atoms (see Figure 1-4 for atom labels) are the most hydrated, the sugar ring oxygen atom O4' is intermediate, and the esterified O3' and O5' backbone atoms are the least hydrated. The hydrophilic atoms of the bases are about equally well hydrated, at half as many waters in the solvation shell of the phosphate oxygen atoms in DNA helices. The relative order of hydration affinities is[27]:

anionic phosphate oxygens > polar base atoms > sugar oxygen atoms.

The most frequent water bridges appear mainly in the minor groove of DNA. The hydration around phosphate groups of helical structures is characterized by "cones of hydration" centred

on each anionic phosphate oxygen, by water bridges between anionic phosphate oxygen atoms of successive residues on the same strand and by 5'-phosphate-water-base bridges. Water bridges are observed also in non-standard conformations and vary systematically around non-canonical base pairs (e.g. A-G, A-A, U-G pairs).

Water molecules participating in or mediating structural bridges between atoms of the nucleic acid should be regarded as an integral constituent of nucleic acids in aqueous solution. As a result, in helical nucleic acids, water molecules might strongly influence fine structural parameters like stacking geometries, twist, and roll angles between base pairs as well as some propeller-twist angles of base pairs. In non-helical elements, like loops and bends, water molecules participate in the stabilization of non-canonical base pairs, which often close hairpins, and bridge approaching phosphate groups. While detailed analysis of water within the simulations is beyond the scope of this work chapter 6 shows some techniques that could be developed to probe the structure properties of the solvent, DNA and cylinder.

1.4 DNA binding

DNA is a biomolecule in which the genetic data of most organisms is encoded. Its sequence defines many features ranging from organism type through physical traits to disease susceptibility. The information encoded in the DNA sequence is put into practice principally through the action of proteins. The DNA sequence is copied onto RNA molecules, which are then used in protein synthesis to encode a specific protein sequence. Inherent in this process is a dramatic information amplification of the genetic information: the single master copy of information in the DNA is used to create multiple proteins. The fact that expression of the DNA information is regulated by proteins, which bind to DNA, allows the information expression to respond to the environment. Protein binding to DNA is normally reversible and non-covalent in nature.

With the genetic sequence of many organisms now known (and particularly the human genome) attention is turning to establishing ways to control specific gene expression. This could be achieved by creating agents which can bind selectively to specific genes and turn their expression either on or off. The ability to turn a gene on or off is important in trying to elucidate the complex and intertwined biological pathways in the cell and also in medical treatment. Many diseases (most notably cancers) are expressed through protein action but actually originate at the DNA level, in the code or in its processing. The information amplification from DNA to proteins, makes DNA a particularly attractive medicinal target, since in principle a single drug molecule per cell could change the expression of a gene on DNA; multiple drug molecules per cell are required to act on the multiple copies of proteins. Both small synthetic molecules and larger modified biomolecules have potential as such drugs, although small synthetic agents are more readily transported inside cells. This field is one which supramolecular chemists are ideally placed to exploit. While the DNA groove binding ligands modeled in this research are examples of such synthetic supramolecular molecules the review that follows describes the various binding modes used by a variety of DNA binding agents. As will be seen, the principles and forces involved in DNA recognition are very similar to those seen elsewhere in the wider field of supramolecular chemistry, although the size, surface dimensions and nature of the DNA biomacromolecules introduce new possibilities and challenges.

While the following section describes a variety of interactions it is important to continually review the research described in the context of a complex genome. Before trying to consider the synthetic cylinder, described in section 1.5.5, as a singular drug molecule able to identify sequence specific stretches of DNA it is worth considering the cylinder relative to the human genome. The complex genome found in humans is three-billion base pairs in size of which the

synthetic cylinders modeled for this research are able to span five base pairs. Statistically, assuming random base distribution, a unique 16- or 17- base pair sequence will occur only once in 4.3 billion or 17 billion nucleotides, roughly the same or a bigger size than the human genome ($3.0x10^9$ base pairs)[28]. Considering the scale of the genome, for the synthetic metallo-supramolecular cylinders modeled in this work to be effective at being sequence selective three or four molecules, depending on the choice of linking structure and the DNA sequence being targeted, would need to be joined in series to enable targeting of unique DNA sequences.

In the context of DNA binding and the potential medicinal uses outlined above the following section presents an overview of ligand interactions with double-stranded DNA a relationship that is fundamental for many intracellular processes. Especially proteins that bind to specific DNA target sequences controlling a variety of processes such as regulation, transcription, and translation. Small binding ligands with reduced or no sequence specificity are often able to interfere with those processes because they are capable of changing mechanical properties of the DNA strands and are, therefore frequently used in cancer therapy.[29]

Because of the complex double-helical structure of DNA, different binding modes are possible. Besides covalent binding there are several classes of specific or nonspecific binding modes: External binding, where nonspecific interactions that are primarily electrostatic in origin allow binding along the exterior of the DNA helix. Binding interactions that involve direct interaction of the bound ligand with the edge of base pairs in either the major or minor groove, these are known as groove binding interactions. And, insertion of planar (or approximately) planar aromatic ring systems between stacked base pairs, this is termed intercalation. Selective binding to the narrow minor groove of AT-rich sequences by van der Waals interactions, formation of hydrogen bonds, and electrostatic interaction is characteristic for minor groove

binders.[30] Electrostatic interaction is characteristic for major groove binders, especially helical peptide ligands.[31]

Detailed information about the structural aspects of binding are given by x-ray diffraction[32] and NMR spectroscopy.[33] With an improvement in molecular dynamics simulations[34,35,36] having happened dramatically over the last decade simulations are able to contribute greatly to the understanding of important biomolecular processes concerned with the control of DNA structure and function.

A search through the literature finds a number of examples of when simulations have been successfully used to provide a greater understanding[37] about systems such as the "Interaction of DNA with groove binding ligands" that will be reported in this work. To summarize the potential methods of binding a variety of examples will be presented of related DNA binding ligands which in the presence of DNA have been observed to; bind either with an external binding mode, as a base pair intercalator, within the minor groove, or as a major groove binding ligand.

1.4.1 External binding

DNA is a highly charged anion. The charge is localised on the phosphate groups and as a result these influence its structure and interactions strongly. This high charge density can lead to the binding of molecules to the phosphate backbone primarily by electrostatic interactions, this is known as external binding. Electrostatic interactions are influenced by the ligand charge, hydrophobicity and size. External binding may also be due to either covalent or non-covalent interactions. This mode of binding could potentially be sampled during simulations where the charge and shape of helical molecules are both changed.

1.4.2 Intercalation

The term intercalation was first introduced by Lerman in the early 1960's to describe a binding mode whereby planar aromatic molecules could be accommodated between DNA base pairs.[38]

Intercalating drugs have been used for many years as anti-cancer treatments. Intercalation causes a lengthening of the DNA helix and perturbation of the phosphate backbone by rotation about the torsional bonds in the DNA backbone as observed by Waring[39]. During this movement in backbone the DNA chain must lengthen and unwind slightly to accommodate the intercalating molecule. This adjustment can in turn lead to long-range deformation of the DNA helix altering the structure and functionality of the molecule. This alteration of the DNA backbone by the binding of an intercalating molecule may inactivate the DNA preventing RNA synthesis and replication of DNA, leading to cell death.

An intercalating drug must be able to fit in between adjacent base pairs within the DNA structure. Most intercalating drugs involve one or more planar aromatic rings, for example the anthrcycline structure. Doxorubicin, also known as adriamycin, is a very effective inhibitor of DNA replication and transcription, intercalating at a purine-pyrimidine step with a preference for the CG sequence.[40] When the drug intercalates with DNA, the cyclohexane ring resides in the minor groove acting as an anchor, hydrogen bonding to base-pairs above and below. The drug is held in place by the formation of favorable hydrogen bonds to the bases within DNA; for example the hydroxyl group in the 9 position forms two hydrogen bonds to N2 and N3 of an adjacent guanine.[41] However, these drugs are not specific to the base-pair sequence and show a high level of toxicity as they will also interact with the DNA in many other tissues.

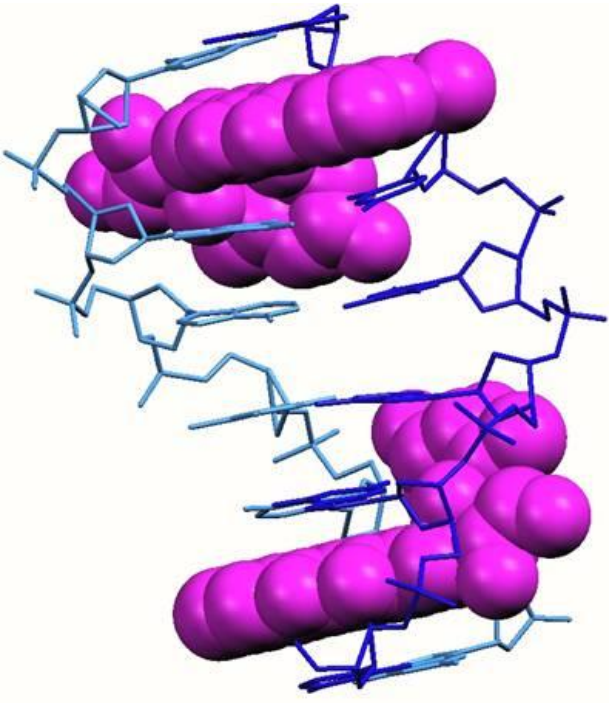


Figure 1-11 Structure of a DNA oligonucleotide with two doxorubicin drugs (shown in pink) intercalated between the base pairs.

A lot of research has gone into reducing the toxicity of intercalating drugs, using approaches such as encasing the drug in a phospholipid shell or combining treatment with another drug that reduces the toxicity. Another problem with these drugs is that the body may develop resistance to them, for example an increase in the level of anti-oxidants such as glutathione greatly reduces the drugs effectiveness. Doxorubicin has an additional effect; it also inhibits the ability for enzymes such as topoisomerase II to interact with DNA. This enzyme cleaves double stranded DNA to reduce the strain that comes from local unwinding. Interference with this enzyme leads to breakage of the DNA backbone and formation of a drug-enzyme complex again preventing replication of the target DNA.[42]

It is also possible for a molecule containing two planar aromatic systems to intercalate twice with the DNA strand. This is called bis-intercalation and brings about twice the helix extension and unwinding than for a singly intercalating drug. An example of a drug that can achieve this is echinomycin, an extremely potent naturally occurring antibiotic which is administered as an

anti-cancer drug. As a general guideline, bis-intercalators consist of two aromatic groups which will fit between the stacked bases linked by a rigid (and often cyclic) peptide chain. Due to the shape and bulk of the DNA binding molecules being modeled we do not expect to see this mode of binding during the simulations. While low energy configurations with the cylinder placed in the major groove will be selected as starting points for the simulation it is not expected that intercalation will be witnessed, although there is a possibility that part of the larger cylinder molecules may be able to bind in this mode.

1.4.3 Minor groove binding ligand

Small molecules that interact with DNA in a non-covalent way generally do so either by intercalating or binding in the minor groove. DNA can interact with a wide range of small molecules including metal complexes, drugs and complex antibiotics [43 44 45]. Many of these small molecules are clinically proven therapeutic agents (e.g. Hoechst 33258) although, in most cases their exact binding mechanisms are still not fully understood. Experimental techniques such as X-ray crystallography and NMR are useful in providing a static view of the final DNA-ligand complex, but are unable to provide information regarding the dynamics of the system. Computational methods such as MD simulations, docking calculations and energy minimizations are invaluable tools able to examine the physical and chemical properties that drive these molecules to bind DNA. Various small molecule DNA complexes have been studied by MD simulations. In the following Netropsin, Hoechst 33258 and metal complexes are described in detail as examples of how MD can be used to characterize DNA response at the molecular level.

The major and minor grooves have significant differences in their hydrogen bonding characteristics, steric effects and hydration. As a result the types of molecules that bind to the major and minor grooves have very different characteristics[46]. Many proteins exhibit specificity

for major groove interactions whilst smaller groove binders generally exhibit preference for the minor groove[47].

Minor groove binding molecules usually have several simple aromatic rings (benzene, pyrrole etc) connected by bonds with torsional freedom that allow appropriate twisting to enable a good shape fit into the minor groove, Figure 1-12. Close van der Waals contacts with edges of the minor groove provide favorable interactions, additional favorable components to the free energy of binding are provided by electrostatic interaction of cationic groups with the large negative electrostatic potential in the minor groove. Minor groove binders interact mainly with A-T rich regions for a number of reasons. In particular, the minor groove is not as wide in A-T rich regions compared to G-C rich regions, and so aromatic molecules can generally fit better into A-T rich regions[48]. Further, the carbonyl oxygen of T and the N-3 nitrogen of A are both capable of accepting hydrogen bonds from bound molecules, whilst G-C base pairs have similar functional groups, the hydrogen bound between the amino group of G and the carbonyl oxygen of C lies in the minor groove, sterically hindering the entry of molecules into the minor groove in this region.

Minor groove binding drugs have the effect of inhibiting the DNA and preventing the replication and growth of cells. They can also prevent molecules such as enzymes and proteins from being able to bind with and act upon DNA.

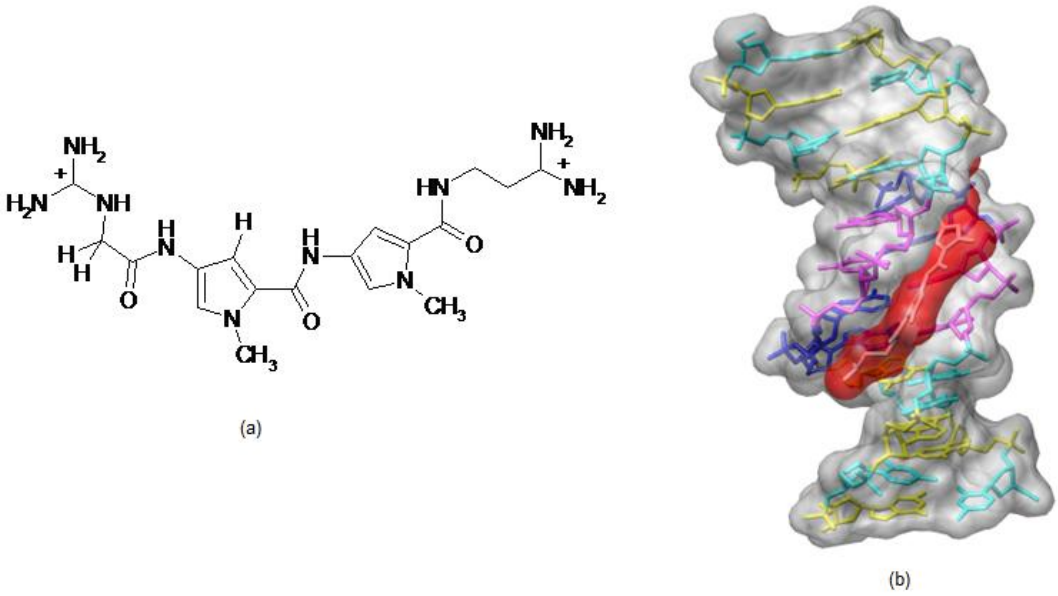


Figure 1-12 (a) Netropsin (b) DNA Helix with bound Netropsin.

Netropsin has a charge of +2 *e* and binds to the minor groove by displacing water molecules that are hydrating the DNA strand. The NH groups of netropsin form hydrogen bonds to the N3 of adenine and O2 of thymine found on the adjacent base pairs.[49] After binding, the molecule sits lengthwise down the minor groove, Figure 1-12.

Binding in the minor groove does not distort the DNA backbone in the same way as intercalation, although it does move the AT residues slightly further apart at the site of binding. Netropsin is specific to sequences of four or more AT base pairs, as in addition to the hydrogen bonding there are also stabilizing van der Waals forces between the pyrrole rings and C2 of adenine.

Netropsin has been extensively modeled with computational models [50 51 52]. Simulations have contributed to understanding the binding mode of this naturally occurring antibiotic with antitumor and antiviral activity. One of the first computational studies of B DNA-netropsin complexes was that carried out by Gago and Richards[53] in 1989. Docking calculations of netropsin binding to dodecamers $d(ATATATATATAT)_2$ and $d(CGCGCGCGCGCG)_2$ was

modeled by gas phase docking calculations. These docking calculations had two obvious limitations. Firstly the DNA and netropsin were minimized separately and secondly, explicit counterions were used but a distance-dependent dielectric function was used to mimic the screening of electrostatics by water. Despite these limitations these calculations did give results that qualitatively agreed with experiment. As expected from experimental results, the binding energy of the $d(ATATATATATAT)_2$-netropsin complex was found to be more favorable than that of the $d(CGCGCGCGCGCG)_2$-netropsin complex, also in agreement with experiment was the observation that the width of the minor groove in the region covered by netropsin was narrower in the $d(ATATATATATAT)_2$-netropsin complex than the $d(CGCGCGCGCGCG)_2$-netropsin complex. In this research the choice of DNA sequences being simulated, $d(ATATATATATAT)_2$, $d(CGCGCGCGCGCG)_2$ and $d(CGCGCATATACG)_2$ have been chosen to enable investigation into the sequence specific characteristics of the groove. However, as all of the initial low energy starting configurations have been selected with the cylinder being docked in the major groove the minor groove will not be initially exposed to the cylinder but may sample this region if there is significant migration of the cylinder from the initial starting configuration.

Since the crystal structure of the $d(CGCGAATTCGCG)_2$ DNA dodecamer was solved by Dickerson *et al,* (often referred to as the Dickerson dodecamer), as seen in the context of netropsin binding it has been know that the width of the minor groove in B-DNA varies in a sequence-dependent way[54,55]. It is wider in G-C rich regions compared with the narrower A-T tracts. Two models are used to explain the heterogeneity in the minor groove structure. In one model the variation of the groove width is a consequence of the short-range interactions between the DNA bases [56,57] In this model the width of the minor groove is unaffected by the localization of water and ions in the groove. In an alternative model, narrowing of the minor

groove is attributed to the sequence dependent localization of water and cations in the groove which then screen the negative charges of the phosphate groups, thus this model predicts that the groove will widen locally as cations move away from a minor groove site and will narrow as they move closer[58,59,60,61].

In this context of minor groove binding it is interesting to discover what role the charge of the netropsin plays in determining its selectivity. Wellenzohn *et al.* studied this using different charge models for the netropsin[62] Three MD simulations were performed in which the charges on the guanidinium and amidinium ends of the netropsin were varied, giving net charges on the netropsin of +2 *e* (normal), 0 *e* and +4 *e*. These simulations were performed with an explicit representation of solvent and a complete treatment of electrostatics using the particle mesh Ewald (PME) method[63]. Results clearly showed that the positive charges of the netropsin affect the minor groove by reducing its width. This is in agreement with an earlier MD study by Hamelberg *et al* which concluded that ions have a profound effect on the minor groove and a narrow minor groove cannot exist for any significant length of time without interacting with cations[64]. These results are also in agreement with both experimental and theoretical studies in which cations have been observed the minor groove of DNA, supporting the model where charge neutralization is responsible for the width of the groove.

The association of DNA and netropsin is also dependent upon salt concentration and decreases in strength with increasing salt concentration. Thus, there is a significant electrostatic contribution to the association of these two molecules. Singh and Kollman calculated free energy of association of netropsin to the Dickerson dodecamer and were able to quantify the effect of the electrostatics on the free energy of association of the positively charged netropsin and the negatively charged DNA[65]. A similar study had been performed earlier for two protein-ligand interactions by Miyamoto and Kollman, but this was the first attempt to calculate the

absolute free energy of binding of a drug to DNA from a simulation performed with explicit representation of the solvent[66]. They concluded that the energy of association is dominated by the van der Waals component, the electrostatic component, whilst being large in magnitude, is almost identical for the DNA-netropsin complex and separated DNA and netropsin. This conclusion is in contrast to conventional ideas of electrostatic forces dominating the binding of cations to DNA but is in agreement with continuum solvent electrostatics calculations by Misra and Honig[67].

While being quite different in structure to netropsin the *bis*-benzimidazole-based ligands[68,69,70], of which Hoechst 33258 is the best studied member, also bind preferentially to the AT rich minor groove regions. Hoechst 33258 is known to be active against certain tumors and is also used as a chromosome stain[71]. Several analogues of Hoechst 33258 have also shown activity against a number of micro-organisms[72]. Two important studies are summarized in this section the first is a MD study of cooperativity in such systems, followed by a consideration of the importance of water molecules in the DNA binding of small ligands.

In biochemistry, a macromolecule exhibits cooperative binding if its affinity for its ligand changes with the amount of ligand already bound[73]. Large molecules such as DNA, with a variety of binding sites, exhibit cooperative binding displaying interactions between binding sites. If the binding of ligand at one site increases the affinity for ligand at another site, the macromolecule exhibits positive cooperativity. Conversely, if the binding of ligand at one site lowers the affinity for ligand at another site, the protein exhibits negative cooperativity. If the ligand binds at each site independently, the binding is non-cooperative. When reading the literature and designing experiments or simulations it is important to consider if the system being inspected allows for the phenomenon of cooperativity. Without an awareness of cooperativity there is a danger of drawing incorrect conclusions from results of relatively small

systems that if expanded to be more representative of reality would show very different behavior due to cooperativity. For this reason cooperativity could sometimes help explain differences between practical experiments and simulations where extended DNA strands are present along with varying concentrations of binding agents.

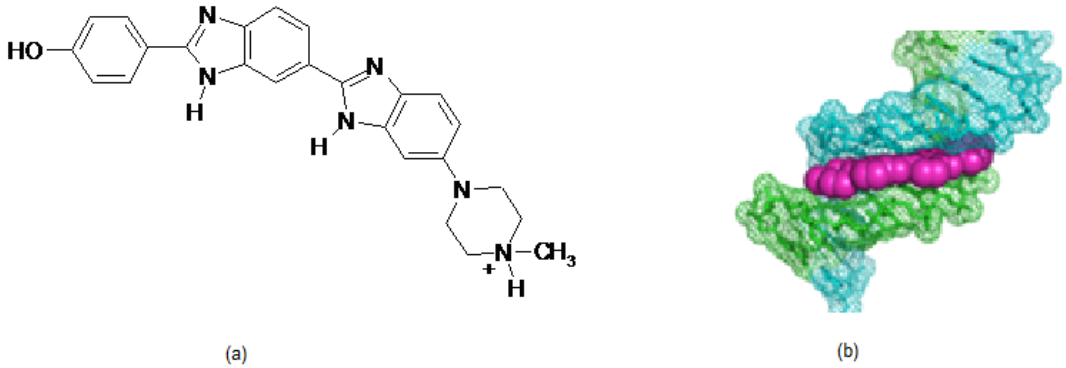


Figure 1-13 (a) Hoechst 3358 (b) DNA Helix with bound Hoechst 33258.

Studies of protein-DNA and drug-DNA interactions have shown that cooperativity can play an important role in sequence-selectivity[74]. There is strong evidence to suggest that positive cooperative binding to homeodomain DNA binding proteins is due to conformational changes in the DNA that are induced by the binding of the first ligand[75]. In the case of drug-DNA recognition, there is evidence to suggest that cooperativity can be mediated by either the conformational changes that are induced by the initial drug binding event, or by direct contact between drug molecules[76,77]. NMR studies have shown Hoechst 33258 to bind in a cooperative way to the minor groove of DNA, however these studies were unable to provide any insight into the origins of the cooperativity [70]. The first MD investigations into the cooperative binding of such systems was reported by Harris *et al* [78] Three systems were simulated, the free DNA, the 2:1 drug-DNA complex and the 1:1 theoretical drug-DNA complex. They were able to show that for this system cooperativity is not related to a significant conformational change. Configurational entropies were calculated by diagonalisation of the Cartesian coordinate

covariance matrix followed by the method described by Shlitter[79]. Surprisingly, analysis of hydration and enthalpic factors predicted the binding process to be slightly anti-cooperative. However further analysis showed that the enthalpic penalty is outweighed by an entropic term and when the both enthalpic and entropic terms are considered the overall free energy differences predict cooperative binding in good agreement with experiment. These results demonstrate the power of MD simulations in providing insights into the molecular origins of phenomena such as cooperativity that could never be probed by static models alone.

A comparison of the DNA binding of netropsin and Hoechst 33258 was reported by Wellenzohn *et. al.* in 2001[80]. They investigated the significance of the ligand tails for interaction with the minor groove. Two 5ns simulations of $d(CGCGAATTCGCG)_2$ in complex with Netropsin and Hoechst 33258 were performed. These simulations enabled the detection of both common effects and differences on the DNA induced by the binding of the two drugs.

Analysis of the simulations showed that both Hoechst 33258 and Netropsin ligand tails undergo conformational changes that are correlated with the time dependence of the minor groove width. However, the two drugs were shown to affect the width of the minor groove in different ways. Complexation of Netropsin led to an overall reduction in the width of the minor groove, except where this was prevented by steric hindrance. This effect extended to the whole DNA and was not just localized at the binding site. On the other hand, complexation of Hoechst 33258 led to a widening of the minor groove at the region where the piperazine tail was bound. These simulations identified the ability of the ligand tails to influence the width of the minor groove and the flexibility exhibited by the groove as it changes structure to allow the ligands to fit exactly.

Various computational studies of Hoechst analogues have been reported[81,82]. Recently a study into the role of water molecules in the sequence selectivity of *bis*-benzimidazoles was reported by Bailly *et. al.*[83] MD simulations and energy minimizations were performed alongside DNA foot printing and surface Plasmon resonance (SPR) experiments to probe the DNA binding of the analogues, 1 and 2 shown in Figure 1-14.

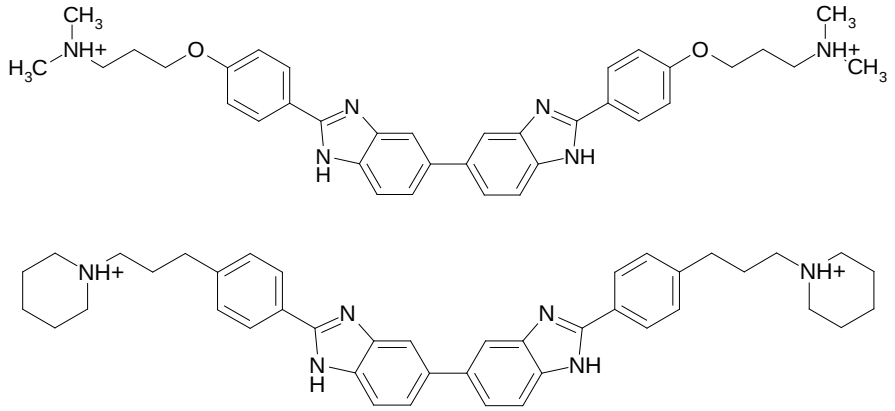


Figure 1-14 Hoechst analogues **1** (top) and **2** (bottom).

These compounds are symmetrical molecules that show preference for AT rich minor groove regions. The compounds differ structurally in that they have different tails, compound **1** has dimethylaminopropyl groups at the tails whereas compound **2** has piperidinpropylphenyl tails.

It was thought that the piperidinpropylphenyl side chains of **2** would enhance its binding affinity for AT rich regions. However, SPR experiments showed **1** to have the higher binding affinity. The molecular basis for this initially puzzling result was probed by MD simulations. 1ns simulations of **1** and **2** complexes with $d(CGCGAATTCGCG)_2$ DNA, were performed using Amber 6.0 in explicit water. The simulations revealed that discrete water molecules located in the minor groove played a key role in determining the DNA binding of these compounds. While the two benzimidazole subunits formed stable hydrogen bond interactions with the four AT base pairs, there were significant differences in the interactions of the tails of the two compounds with the GC base pairs. Three hydrogen bonds mediated by a water

molecule were formed between the phenoxy oxygen of **1** and the DNA. The phenoxy oxygen is replaced by a carbon in **2** removing the possibility of forming hydrogen bonds with the DNA in this region. A water molecule was observed to form two hydrogen bonds with the DNA in this region, but did not interact with the ligand. It was suggested that this tight, water mediated, hydrogen bond network observed in the DNA-**1** complex, and absent in the DNA-**2** complex was a major cause of the higher binding affinity of **1**.

There has been much interest in metal complexes as ligands for binding to DNA[84,85,86]. Chiral, octahedral ruthenium complexes containing planar aromatic ligands are known to bind to DNA and have been studied extensively using a variety of techniques. Tris bidentate complexes containing the 1,10-phenanthroline, (phen) ligand are amongst the most widely studied [87]. Experimental studies have, however, resulted in some controversy as to their exact binding modes. Barton *et al*. have proposed two binding modes for each enantiomer, one in which partial intercalation occurs from the major groove in an enantioselective way, and the other in which the ligands are either weakly surface bound or bind in the minor groove [88]. Nordén and co workers however have proposed a facial binding mode for the Δ-enantiomer (right handed) and intercalation or partial insertion from the minor groove for the Λ-enantiomer (left handed) [89].

Molecular modeling studies by Haworth *et al*. have shown that for both enantiomers the partially inserted geometry with insertion from the major groove is the most favorable[90]. However for the Δ-enantiomer facial binding to the surface of the minor groove is similar in energy to partial insertion and so it is conceivable that both sites could be occupied. The difference in energies is greater for the Λ-enantiomer and therefore it is unlikely to bind facially.

Molecular modeling methods have also been used to study the DNA binding modes of chiral cobalt intercalators. Experimental results show that these complexes bind as classic intercalators in an enantioselective way. However, at the time of the initial modeling study there had not been a detailed three-dimensional structure solved for these complexes bound to DNA and hence molecular modeling approaches were employed by Yang *et al*. to test the Barton and Nordén interpretations[91]. Gas phase manual docking followed by minimization was used to identify optimum binding sites. Results showed that the complexes bind via intercalation and with the probability of intercalating from the minor groove being considerably greater than intercalation from the major groove. This also provided further support for the Nordén model. However they also concluded that the observed enantioselectivity in the binding provides support for the Barton model too. These results contradicted the earlier study by Haworth *et al.* who concluded partial insertion from the major groove was the most favourable binding geometry. It should be noted, however, that the docking calculations (*Yang et al.*) were performed in the gas phase and therefore do not include solvent effects which could play a major role in the binding.

Studies of DNA binding by metal complexes through the formation of metal-ligand bonds are extensive and as a result much is known about the binding to N7 of G and A residues[92]. By contrast, noncovalent binding of metal complexes to DNA is a less well-developed area with a common area of study being the approximately spherical ruthenium polypyridyl complexes, complexes with planar intercalating units, or combinations thereof.[93,94,95] The size of these compounds means they often bind in the minor, rather than the major, groove. Moreover their small size means that they cannot target more than 2-3 base pairs and consequently are not suitable scaffolds for sequence-specific recognition.

1.4.4 Major groove binding ligands

The major groove of DNA most often provides the best site for recognition and binding of proteins. This is achieved through its larger size which allows interaction with a greater number of base pairs and therefore access to greater functionality. In comparison with the minor groove, there is an increased number of hydrogen bonding sites and greater scope for van der Waal interactions in the major groove. Upon complex formation, molecules binding in the major groove can often induce significant structural changes in the DNA. In comparison, the minor groove is more rigid and molecules binding in it rarely induce structural changes.

The helix turn helix and zinc finger are two of the motifs found in nature to bind selectively to DNA. Transcription of DNA is regulated by proteins that recognise specific DNA sequences through discrete DNA-binding domains in their polypeptide chains. The helix-turn-helix (HTH) motif is a common feature of most prokaryotic DNA-binding domains. An extended turn chain of amino acids connects the two α-helices which are held at a fixed angle. The HTH motif is typically about 20 amino acids long, the first 7 amino acids for the first helix, the turn is formed by amino acids 8 to 11 and the second (recognition) helix is formed by the remaining 9 amino acids. While these proteins are too big to fit into the minor groove, the recognition helix fits into the major groove of B-DNA. The sequence specificity of the DNA binding is determined by direct interactions with the nucleotides of the groove. These interactions involve hydrogen bonding between amino acid side chains and the edges of the base pairs.[96]

Transcription of DNA in eukaryotic cells is controlled by site-specific DNA-binding proteins called transcription factors. The DNA-binding domains of transcription factors are built up of a very limited number of structural motifs. These include the leucine zipper and zinc finger.

The zinc finger motif first described in 1985 by Klug and Rhodes contains nine repeated amino acid sequences of about 30 residues each, with a total of 344 residues in the whole transcription factor, the proteins also contain intrinsic zinc atoms which are essential for transcriptional activity[97]. They suggested that each zinc finger binds into the major groove of DNA. Klug and Rhodes proposed a model for DNA binding of zinc fingers whereby the protein lies on one face of the DNA helix with successive fingers pointing into the major groove alternately from opposite directions. Subsequent experimental studies have shown this model to be quite accurate. A zinc finger DNA complex from a crystal structure is represented schematically below.

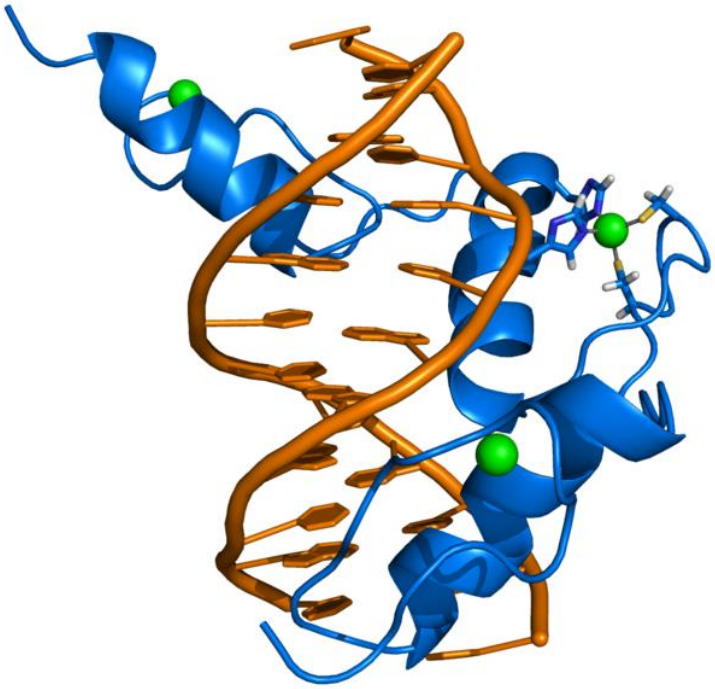


Figure 1-15 Cartoon representation of a complex between DNA and the ZIF268 protein, containing 3 zinc finger motifs. The coordinating residues of the middle zinc finger are highlighted. Based on the x-ray structure of PDB 1A1L[98].

The major groove, while providing binding sites for protein molecules, can also accommodate smaller molecules such as the well know anti-cancer drug, cisplatin. Cisplatin was established as a potent anti-leukamic agent in 1969 after a series of tests on neutral complexes by Rosenberg *et al* [99]. In 1970, DNA was identified as its biological target with UV data revealing

it to interact with the base pairs[100]. Competition binding studies with Ethidium bromide showed it to bind covalently to the bases rather than intercalating.

Cisplatins mechanism of action has been the focus of much research over the past few decades. It has been shown that cisplatin binds to purine bases with high selectivity. Furthermore, it binds preferentially at the N7 position in guanine inducing unstacking of the bases. Unstacking occurs at the Pt-binding site, causing the DNA to kink. In 1988, Sherman *et al.* obtained a crystal structure for the major cisplatin-DNA adduct, cis-[$Pt(NH_3)_2${d(pGpG)}], showing the platinum metal centre coordinated in a square planar mode to two cis-Ammine ligands and two guanine N7 atoms in a manner that would lead to unstacking of the bases in duplex DNA[101].

While cisplatin has been used successfully to cure a large number of patients since its introduction in 1979, it does produce toxic side effects and certain tumours can develop resistance to the drug. Second generation platinum drugs such as carboplatin and oxaliplatin have been developed. Carboplatin is the most widely used second generation drug, and whilst it offers greatly reduced side effects it is not as potent an anti-tumour agent as cisplatin. Over the past few years third generation, polynuclear platinum complexes have been developed and currently the possibility of developing anti-cancer drugs using transition metals other than platinum is being investigated[102].

Despite the success of cisplatin and subsequent transition metal complexes in the treatment of cancer there is still a need to develop drugs that will bind in a sequence specific way to DNA, are efficient at low doses, produce few side effects, bind strongly enough to prevent DNA replication and will resist displacement by proteins[103].

1.5 Computer simulation of DNA

1.5.1 DNA structure

How to treat long-range electrostatic interactions in simulations has been a major problem for modeling DNA in the past. With the charged phosphate backbone and the charge-neutralizing complement of counter-ions in the simulation cell it is important to have a method that can accurately and efficiently deal with the electrostatics. Early simulators treated electrostatics by essentially ignoring them or truncating the interactions at short range. 10 Å cutoffs were common due to the computational cost that scales as the cube of the cut-off. As a result when long simulations began to be possible due to computer advances, substantial distortions were seen in the DNA being simulated,[104] even when counter-ions and explicit solvent were present. Thus, nucleic acid simulations were seen to be particularly sensitive to the treatment of electrostatics. Improvements in DNA stability resulted with the introduction of the Ewald summation. As a result it was a significant step forward for DNA simulation work when fast algorithms for Ewald summation such as the particle mesh Ewald (PME) were implemented. This method and the more recent SPME are now commonly used and will have been used to ensure the electrostatics were correctly accounted for in the majority of the more recent systems in the remaining review.

MD simulation has frequently been used to look at the factors underlying DNA structure and flexibility and also the sequence dependence of these properties. At a local level, the free energy of both base pairing[105] and base stacking[106] have been calculated employing molecular dynamics techniques and using isolated bases or dinucleotide monophosphate fragments, while other studies have probed these interactions within the double helix via induced base pair opening.[107]

Base sequence effects on stacking have been demonstrated with short MD runs, which confirm that Adenine-Adenine, ApA, steps seem to have well-defined geometries, while CpA steps can show bimodal behaviour.[108] Other studies have generalised the malleability of CpA to other pyrimidine-purine steps.[109,110] In agreement with NMR data, runs of GC base pairs have been shown to alter the stacking pattern to resemble that found in A-DNA, even if the backbone geometry remains close to that of B-DNA.[111,112,113]

Studies of backbone fluctuations have shown that phosphate movements associated with the B-DNA occur for most nucleotides. These transitions are linked to base unstacking and water rearrangement and are also favoured by low water activity.[114] These B-DNA states can last from a few picoseconds to several nanoseconds.

Considering the molecule as a whole, a number of simulations have looked at the nature of the A → B transition. This transition, which is directly coupled to changes in sugar puckering,[115] has been shown to occur spontaneously in water, although the reverse transition does not occur in an ethanol/water mixture, due to a significant energy barrier.[116] Born solvation energy calculations using MD snapshots confirm the experimental observation that GC-rich sequences are more A-philic.[117] MD simulations also show that the transition can be linked to a shift in the balance of interphosphate repulsion, desolvation, and ion-DNA interaction terms.[118]

In addition to the specific DNA binding complexes reviewed earlier an extensive project has been begun by the Ascona B-DNA Consortium (ABC) in recent years. In summary the group carried out calculations on 136 unique tetranucleotides embedded in 39 DNA oligomers. All simulations were carried out with a consensus protocol using the AMBER suite of programs[119] and the parm94 force field of Cornell et al.[120] This force field, was verified to produce overall agreement between calculated and observed DNA structures in crystals and in solution[121,122]

MD trajectories of 15 nano-seconds were obtained for each of the 39 oligomers. Initial analysis from this work presented results for properties of the d(CpG) step extracted from its 10 unique sequence contexts.[123] In the subsequent paper the structural analysis of all 136 unique tetranucleotides was reported.[124] As well as being an informative and extensive study this work is an indication of the rapid advancement in the length, scale and reliability of DNA MD simulations. While no DNA binding is simulated in these studies they provide a good reference for comparison of force-field and DNA sequence effects.

1.5.2 Water and ion distribution

Structure and function of nucleic acids are determined, to a large extent, by interactions with the surrounding aqueous solvent in the biological environment. Water activity is a major factor in the stability of nucleic acid conformation and complexes while localised structural waters effectively extend specific nucleic acid structures around the groove regions of multi-stranded nucleic acids.[125,126] By mediating sequence and conformation specific recognition of nucleic acids, water plays a significant role in interactions with proteins and other ligands. Evidence of interfacial water molecules exchanging with bulk water on the sub-nanosecond timescale up to 10-15ns for A-tracts, challenging the view that T-A steps widen the minor groove and disrupt the hydration structure. A 'spine of hydration' in the minor groove, as observed by X-ray diffraction in DNA crystals, is present also in solution, with residence times significantly longer than 1 ns. The water environment surrounding DNA has been investigated both by physical experiment[127,128,129,130] and by theoretical computer simulations [131,132] that describe a variety of contexts in which water can become stabilised with residence times greater than the pico second timescale.

The correlation of the two major DNA conformations in the biological context, A and B, with water activity has been known for a long time. Reduced water activity due to low humidity

environments or the addition of organic solvents to an aqueous solution creates an environment in which the more compact A form becomes more favourable over the B conformation. High concentrations of salt, reducing the water activity, also cause transitions from the B to A form.[133] Furthermore, the tendency for DNA molecules to form A- or B-DNA conformations depends on the base sequence. As a general rule, A-DNA is more favourable for G·C base pairs than for A·T base pairs under reduced water activities[134] or increased salt concentration.[135] While B-DNA is often considered to be the predominant conformation in the biological environment, A-DNA structures may occur in DNA-protein interactions.

Although there is much data on hydration around A- and B-DNA from a wide variety of experiments, the detailed mechanism of how DNA conformations quantitatively correlate with their water environment is still elusive. Theoretical studies of nucleic acid hydration have become a useful addition to the experimental data by exploring new viewpoints beyond experimental reach within the approximations of the theoretical models. Studies of this kind, such as molecular dynamics (MD) computer simulations, are particularly attractive for studying nucleic acid-solvent interactions. In principle, MD simulations can provide a complete picture of structural and dynamical aspects of nucleic acid solvation on an atomic scale for a given model.[136] In experiments, a similar level of detail is available, in principle, from x-ray and neutron diffraction as well as NMR measurements, but their sensitivity is limited to the study of only the most localised solvent molecules.

The reliability of MD simulation results depends on both the accuracy of the model for the physical interactions and using sufficiently long simulation times to obtain meaningful averages of equilibrium populations. Following recent advances in methodology and computational power, it is has been possible to simulate nucleic acid fragments in explicit solvent over many nanoseconds [137,138,139] for some time with more recent force field developments enabling stable

simulations of DNA in the region of the microsecond[157] timescale. The nucleic acid and solvent structures remain close to the experimentally expected conformations, giving support to the validity of the theoretical models. However, closer inspection of the influences of different force fields has revealed biases towards A- or B-form conformations in different force-fields. DNA structures simulated with the CHARMm22 force field[140] favour A-type structures, while B-DNA is favoured by the AMBER force field and is largely independent of sequence and the environment. In order to alleviate these problems, subtle modifications in the backbone parameterisation have been applied in the next generation force fields that shift the equilibrium from B- towards A-DNA for the AMBER96 force field and generate balanced equilibria of A- and B-DNA structures with CHARMM27[141]. However, these biases can be used to gain an insight into the correlation of solvent and DNA conformations if the force field is considered as applying weak constraints that keep the DNA either in the A or B form.

The explicit simulation of water molecules over many nanoseconds allows the calculation of high-resolution water density distributions around the whole DNA complex and accurate estimates to what extent hydration is based on geometric criteria. The ability to see all water molecules around DNA in atomic detail with MD simulations overcomes the restrictions of experimental techniques that limit atomic resolution to highly localised water molecules.

Water molecules associated strongly with DNA have much reduced mobilities and occupy well defined hydration sites in the minor and major groove as well as around the phosphate oxygens. They are sufficiently localised to be observed in crystals by X-ray or neutron diffraction.[142] The most ordered water molecules are also visible by NMR techniques.[143,144] In particular, a "spine of hydration" of single water molecules along the unusually narrow minor groove of the sequence $d(CGCGAATTCGCG)_2$ has been known for a long time. Recent studies of this hydration pattern at very high resolution have confirmed the early findings but also raised the

possibility that some of the observed water molecules may, in fact, be sodium ions that exchange with water for some or most of the time. Similarly well ordered minor groove hydration sites were also reported around other sequences. Other recurring hydration patterns that have been identified are pentagonal water arrangements in A-DNA major grooves and "cones of hydration" with three water molecules oriented in tetrahedral form around each phosphate oxygen in B-DNA.[145]

In many cases the reported water molecules from crystal structures give only an incomplete picture of ordered hydration sites around a given DNA fragment. This may be due to resolution limits and the less pronounced density differences in much of the first hydration layer, but often sites are also not accessible to water due to crystal packing interactions or other solvent molecules and ions that are present in the crystal blocking the hydration sites. As a consequence, the average number of water molecules per base pair in crystal structures is only between 9 and 10, which is much lower than the total number of 20 bound water molecules expected from other experimental evidence and molecular dynamics simulations.[146]

Because of the limitations in observing water molecules with crystallographic methods, a statistical analysis of hydration sites from many data sets is particularly valuable. This has been done for the groove regions and around the phosphate groups.[147] In these studies pseudo-electron densities were calculated from water positions after aligning nucleic acid bases for the groove regions and phosphate groups for backbone hydration.[148] The distribution of water molecules relative to a nucleotide type or phosphate group was found to consist of well-defined hydration sites that suggest a predominantly local nature of hydration largely independent of sequence context since the water sites were averaged from different DNA sequences.

In more detail, around B-DNA prominent sites were found in the minor groove at the purine N3 and pyrmidine 02 atoms and in the major groove around purine N6/O6 and N7 as well as pyrimidine N4/O4 atoms. Less well hydrated sites were found in the major groove near pyrimidine C6 and near guanine N2. A-DNA hydration sites were found to be similar but more complex in the major groove.[149,150] Additional high densities of crystallographic water molecules occur at a second site near the guanine O_6 atom and near cytosine C_5 atoms. Using these local hydration densities as building blocks for DNA hydration, research was able to predict the particular hydration of $d(CGCGAATTCGCG)_2$ close to the actual observed patterns.

Statistical analysis of hydration sites around the phosphate group reveals the previously suggested tetragonal water arrangement at each phosphate oxygen while well defined hydration sites around the ester and oxygens are missing to a large extent.[151] The formation of water bridges between subsequent (O2'-P) phosphate oxygens is observed around both A- and B-DNA. In B-DNA minor differences were found between purine and pyrimidine bases reflecting stronger water bridges between O2'-P and pyrimidine C6' than purine C8' atoms in the major groove. Both water bridges were also observed in A-DNA.

Computer simulations of explicit solvent allow calculation of water density distributions around DNA,[152 153 154] complementing the crystallographic results. The locations with the highest water densities can be easily compared with experimental hydration sites to establish confidence in the theoretical methods while additional information about the distribution of less ordered water molecules is available from the simulation beyond experimental resolution. Early simulation studies were focused mainly on highly localised waters in the spine of hydration along the minor groove. This was due to a total simulation time of 28ps and 491 water molecules that did not provide sufficient statistics.[155] Calculations including all-atom solvated DNA were reaching 10ns[156] by the start of the century with the most recent simulations now

extending to microsecond timescales[157] providing a more detailed view of the properties of DNA than previously achieved. While the simulation results generally compare well with experimental data, slight discrepancies are attributed to the averaging procedure and resolution limits in the experimental data. This ability, not only to match but also to evaluate experimental results, suggests that MD simulations can provide an accurate representation of water distributions beyond the currently available experimental resolution.

Using computational methods[158] Cheatham and Kollman simulated a DNA sequence composed of the ten A·T base-pairs. Their findings, Figure 1-16, show the average structure of this sequence overlapped with two related snapshots from the final nanosecond of a two nanosecond simulation. Clearly visible, twisting from lower left to upper right, is DNA's "spine of hydration" in the minor groove that runs between the double-helical ridges. Density contours for water show the most probable (red) and slightly less probable (yellow) positions for water molecules.

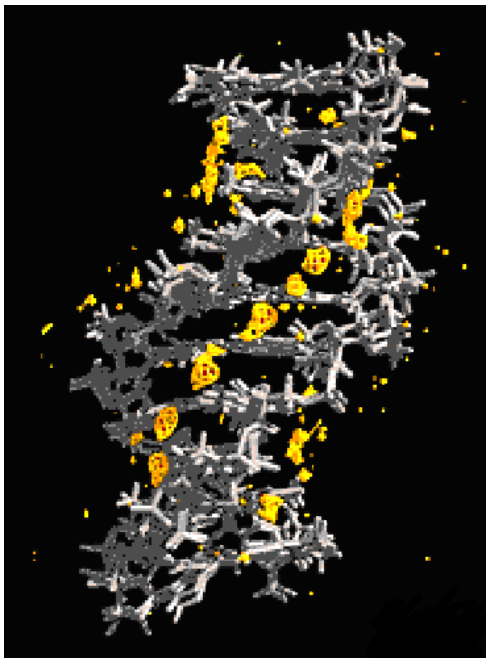


Figure 1-16 "spine of hydration" in the minor groove that runs between the double-helical ridges.

More recent studies by Auffinger and Westhof[159] reported that no long-lived water molecule was detected in the minor groove of the $d(TpA)_{12}$ duplex. This finding was inline with current NMR data which indicates that although long-lived water molecules could be detected in the minor groove of A2T2 to A4T4 sequences, no long-lived water molecules were detected in the spine of $d(CGATATATCG)_2$ oligomer.[160] Identical sequence-dependent effects relating to the detection of long-lived water molecules by NMR have been reported.[161] Thus, the simulations by Auffinger and Westhof support the assumption that d(TpA) steps are able to break the minor groove spine of hydration.

In their review of nucleic acids and their complexes Giudice and Lavery[162] report how improvements in force fields, simulation techniques and protocols, and increasing computer power have all contributed to making nanosecond scale simulations of DNA commonplace. The results of these longer simulations are helping to explain how nucleic acids respond to their environment and to their base sequence. Concentrating on their findings with regards to water and ion distribution they conclude that the longer and more stable simulations have sparked a deeper analysis of the role of water and ions in stabilizing both DNA and RNA conformations. First shell water residence times can extend to the nanosecond range, and this is also true for monovalent ions.[163,164,165] Waters are found around the backbones and in both grooves of the helix, although long-lived hydration patterns are mainly confined to the narrower groove of both DNA and RNA and appear more sequence dependent in the case of DNA.[166,167] Ions occupy both major and minor groove sites and are strongly coupled to DNA bending[168] via major groove interactions with CpG steps and possibly to minor groove width in AT rich tracts. First-shell hydration and ion binding have also been studied with respect to the B → A transition, where the electrostatics dominate, although many factors contribute to a subtle sequence dependent free energy balance.[169,170]

With the stable multi-nanosecond trajectories of both helical and folded nucleic acid a reality, and showing good agreement with experimental data, the effects of base sequence and of environmental changes are also becoming accessible. Such computational studies provide an atomic-scale view of thermally induced fluctuations and allow a detailed structural and energetic analysis which goes beyond current experimental possibilities. Given the polyelectrolyte nature of nucleic acids, particularly interesting developments involve a better understanding of how solvent and counterions influence the behaviour of these environmentally sensitive macromolecules.

1.5.3 DNA-protein interactions

DNA-binding proteins play a central role in all aspects of genetic activity within an organism. MD simulations have been used to probe the nature of DNA-protein complexes thereby giving valuable insights into the genetic processes that they control. The majority of DNA-protein simulations employ either the CHARMM or AMBER force-fields (see Chapter 2). Currently, simulations of DNA-protein complexes are performed either by solvating the complex within a water sphere and use a switch/shift potential for electrostatic interactions or implement periodic boundary conditions with PME treatment of electrostatics. [171]

Roxtrom *et al*. reported the first MD simulation of a fully charged protein-DNA complex in the nanosecond timescale.[172] They performed 1ns simulation of a zinc finger-DNA complex. Zinc fingers are important DNA-binding domains found in many transcription factors. They bind in the major groove of DNA where the larger size of the groove allows access to a greater number of base pairs, thereby increasing the scope for sequence specific binding. Zif268-DNA was modelled with the GROMOS87 force-field. Explicit water molecules were included using the SPC water model with the modifications of Daura *et. al.*[173] Results showed the DNA to be quite stable throughout the 1 ns simulation with an average rmsd of 1.35Å, however a change of

conformation was observed for the Zif268 after 850ps. Although these results agreed well with experiment, a 1ns trajectory may well be too short to pick up significant dynamical changes in the protein-DNA complex.

As the computer power available has increased, it has become possible to perform longer simulations of these complex systems. Tsui *et. al*. studied the hydration of a zinc finger-DNA complex using both NMR and MD simulations.[174] They performed two 2 ns simulations of the same system in explicit water. This was the first time multiple simulations of nanosecond timescales had been performed to compare the accuracy of MD for predicting hydration patterns. The standard AMBER force field parameters and the TIP3P model of water were used for all simulations [175] [176]. Patterns of hydration and the trajectories of water molecules intimately associated with the zinc finger-DNA complex were analysed. It was not possible to calculate accurate residence times from these nanosecond timescale simulations as long residence water molecules can remain bound for up to 10^3 ps. Nevertheless the results gave excellent agreement with experimental intermolecular NOEs. Furthermore the MD results were instrumental in interpreting the otherwise ambiguous protein-water NOEs in terms of water residence times.

These simulations provide an excellent example of a system in which MD simulations with explicit water have provided atomic level detail of protein-water-DNA interactions which could not otherwise have been identified from experimental data alone.

MD simulations have also been used to study several other DNA-protein complexes. One of the best studied DNA-protein complexes is the estrogen receptor DNA-binding domain (ERDBD). Eriksson and Nilsson reported an MD study on the comparative DNA-binding of ERDBD both as a dimer and a monomer.[177] They used the CHARMM22 force-field modified for the zinc ions

and the side chains of the coordinating cysteins. The protein monomer-DNA simulation was run for 1.3 ns whereas the protein dimer-DNA was run for 0.5 ns. Results showed a considerable difference in the structure and dynamics of ERDBD complexed as a monomer to DNA compared with the dimer in complex with DNA. The monomer-DNA complex showed significant rmsd from the starting structure whereas the dimer was very well behaved. The MD simulations suggested that dimerisation facilitates the DNA binding of ERDBD by ordering the ZnII region of the protein.

Few simulation studies of protein-DNA complexes have considered both the bound and free DNA. One such study was reported by Tang and Nilsson.[178] They conducted an MD simulation study of the human sex-determining region Y (hSRY) protein interacting with DNA. Analysis of their results demonstrated the hydrophobic nature of the DNA-protein interaction. They were also able to show that both hSRY and DNA undergo significant conformational changes during binding which enable an almost perfect fit in the dimer.

Crystal structures of protein-DNA complexes often reveal ordered water molecules at the protein-DNA interface. For example the structure of the trp repressor-DNA complex has three ordered water molecules at the protein-DNA interface that can hydrogen bond with the base pairs and the protein side chains. The presence of these ordered water molecules begs a molecular explanation of their presence and role. Reddy *et al*. reported a study in which they address this issue.[179] They analysed X-ray/NMR structures of 109 protein-DNA complexes that contained interfacial water molecules. Hydrogen atoms were added to the protein-DNA complex and water oxygen atoms and the systems were energy–minimised using the Amber 6 suite of programs. To test the validity of this method, MD simulations were performed on 35 of these complexes with explicit solvent and counterions. The simulations were run for 100 ps

with 5 of them extended to 200 ps as a check of convergence. The results of the MD simulations were essentially identical to the minimisations and the conclusions unchanged.

The water molecules were split into four broad classes; class I contained water molecules that made hydrogen bonds with the protein as well as the DNA, class II contained water molecules that made contact with either the protein or the DNA, but not both, water molecules proximal to hydrophobic atoms of either protein or DNA were placed in group III and water molecules hydrogen-bonding to other water molecules as in bulk solvent made up group IV. About 6% of the water molecules belonged to class I and were responsible for water mediated protein-DNA interactions. They concluded that majority of the water molecules belonged to class II. The role of these waters was to buffer electrostatic repulsions between the protein and DNA. There were very few water molecules in class III.

The simulations of protein-DNA complexes reported in this section illustrate the quality of current force-fields. The importance of accurate representation of the solvent in MD simulations of nucleic acids is also highlighted.

1.5.4 Lipid-DNA interactions

Binary mixtures of suitable cationic and neutral lipids can form stable complexes with DNA and have the potential to act as synthetic carriers of DNA. The study of lipid-DNA complexes has received considerable attention over the last few years, however due to the complexity of the problem, few atomistic molecular dynamics simulations have been attempted for these systems.

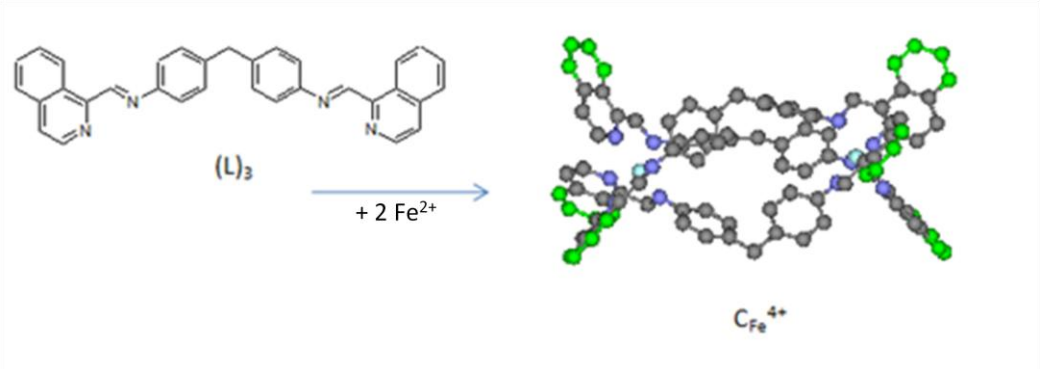
Klein et al reported the first MD simulation of a DNA-lipid complex in 1999.[180] The results of a 5.5ns simulation of a binary mixture of dimyristoylphosphatidylcholine, (DMPC) and dimyristoyltrimethylammonium propane, (DMTAP) intercalated with B-DNA decamer were presented. The simulation predicted the existence of the polar head groups of DMPC and DMTAP with almost equal probability around the DNA phosphates. The 5.5ns simulation was performed with the CHARMM force-field, TIP3P model for water and PME treatment of electrostatics.[181,176,63] The DNA structure remained stable throughout the simulation. These results have shown that the current generation of force fields implemented with state-of-the-art simulation methods do have the ability to offer insights into these complex biological systems.

1.5.5 Synthetic metallosupramolecular cylinders

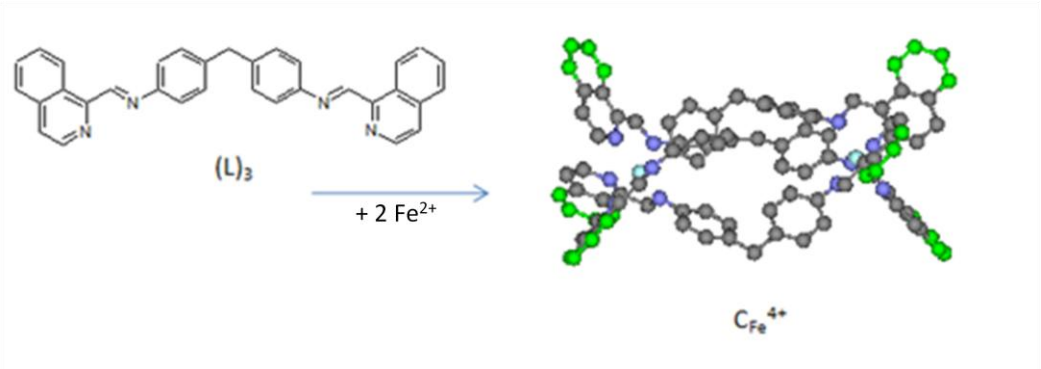
Gene expression is one of the most fundamental processes in biology. It involves the transfer of information encoded within the gene to produce a biologically active protein. However, not all genes are expressed in every cell all the time. Genes are expressed only when they are needed for a cell to function. Regulation of gene expression is controlled by proteins that activate or repress transcription by binding to short, specific DNA sequences. Control of gene expression is one of the key areas of interest in molecular medicine. The ability to turn genes on or off artificially by the action of synthetic analogues of DNA binding proteins is an important goal that would open up new possibilities for disease control and prevention as well as cure.

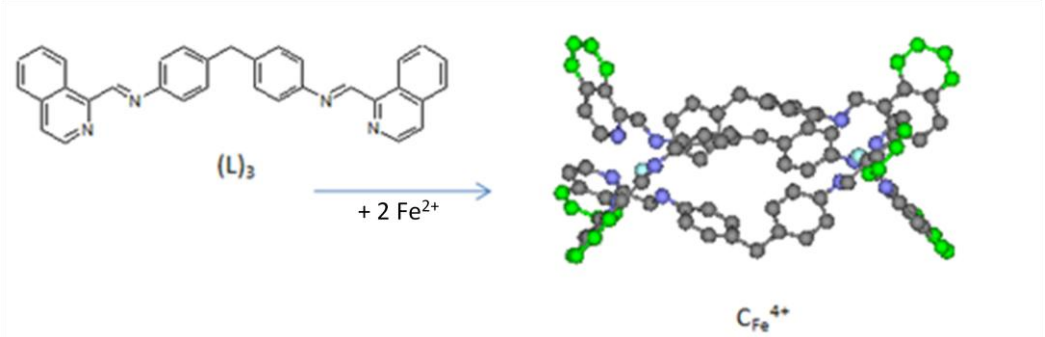
Proteins that bind DNA frequently achieve sequence-specific code recognition by binding in a non-covalent way in the major groove of DNA[182,183,184]. In biological systems, the major groove is the preferred binding site for sequence recognition as it shows the greatest variation in size and shape with base sequence, it would therefore be the ideal target for synthetic molecules designed to recognise and bind to specific DNA sequences. However, until recently, relatively little progress had been made in producing synthetic major groove binders. Most of the

compounds synthesised are either minor groove binders,[185] intercalators[186,187] or metal complexes that span only 2-3 base pairs[188,189,190]. In either case the interaction with DNA is too limited for sequence selectivity. Synthetic molecules that target the major groove in a sequence-selective way remain a major goal in molecular medicine. Recently a step has been taken toward achieving this goal. Hannon *et al.*, developed a novel parent cylinder, ($[Fe_2(C_{25}H_{20}N_4)_3]Cl_4$), ${}^{P}C_{Fe}{}^{4+}$ (Figure 1-18) that binds strongly into the major groove of DNA and is large enough to span more than four base pairs[191]. This molecule is called the Parent Cylinder as derivatives of the ligand ${}^{P}L$, for example ligand, L in this



1-17 (left) Ligand, (L) = ($C_{33}H_{28}N_4$), (right) (${C_{Fe}}^{4+}$) = $[Fe_2(C_{33}H_{28}N_4)_3]Cl_4$ additional structure highlighted in green





study (Figure 1-21), form the basis for a whole series of potential DNA binders with various potential binding modes and each with varying ability to induce conformational change to DNA . The compound is a metallo-supramolecular tetracation with a triple helical framework that can be tuned with respect to shape, adding functionality to the parent ligand ^{P}L, or charge, by experimenting with different metal centres. The parent cylinder and each derivative have the potential to be valuable components of a non-covalent molecular tool kit for DNA. It remains for us to better understand the relationships between cylinder properties and DNA properties to guide future developments and maximise the potential of the "tool kit".

Supramolecular chemistry[192,193] is the methodology that enables the design of large synthetic arrays and as a result has a role to play in bridging the size gap between traditional small-molecules and larger-biomolecular DNA recognition motifs. Within this field, metallo-supramolecular assembly is particularly attractive for the design of noncovalent DNA recognition agents[194] because of the cationic charge that the metallo-centers impart, along with a substantial energetic contribution to the noncovalent binding to anionic DNA. This has been confirmed in studies with the racemic mixture of $^{P}C_{Fe}^{4+}$. The effect of the racemic mixture is to induce a dramatic intramolecular bending that results in coils of DNA. The NMR data collected for the racemate confirms major groove binding, but was intriguing in that only the minus helix, the ***M*** enantiomer, was present in the refined structure.[195] In this work the ***M*** enantiomer of $^{P}C_{Fe}^{4+}$ is used due to the supporting data that is available relating to its binding mode, the ***P*** or

plus helix will not be modeled in this research. The naming structure for the enantiomers comes from the ability of enantiomers to rotate plane polarized light. If the enantiomer rotates the light clockwise (as seen by a viewer towards whom the light is traveling), that enantiomer is labeled,+, or Plus (***P***) in this work. Its mirror-image that has the opposite effect on the plane polarized light is labeled, -, or Minus (***M***). Enantiomers are identical in their chemical formula and with respect to ordinary chemical reactions, but differences arise when they are in the presence of other chiral molecules such as DNA. The interaction of circularly polarized light with two enantiomers of a chiral compound shows the different absorbance properties clearly utilized in Circular Dichroism, CD[196], as shown in Equation 1-1.

$$\mathrm{CD}(\lambda) = A_{\mathrm{L}}(\lambda) - A_{\mathrm{R}}(\lambda)$$

Equation 1-1

Where $A_{\mathrm{L}}(\lambda)$ and $A_{\mathrm{R}}(\lambda)$ are the absorption spectra measured with left and right circularly polarized light, respectively. An achiral molecule binding to a chiral polynucleotide acquires an induced CD. For this reason CD spectra of ligand-polynucleotide adducts provide information about both the polynucleotide and the bound ligand[197 198] which can then be associated with the relevant absolute configuration determined using crystallographic methods.

As has already been established $^{P}C_{Fe}{}^{4+}$ is an example of the DNA binding of a synthetic tetracationic cylinder with a triple-helical architecture formed by three ligand strands wrapped around two metal dications, two iron(II) centres.[199,200]

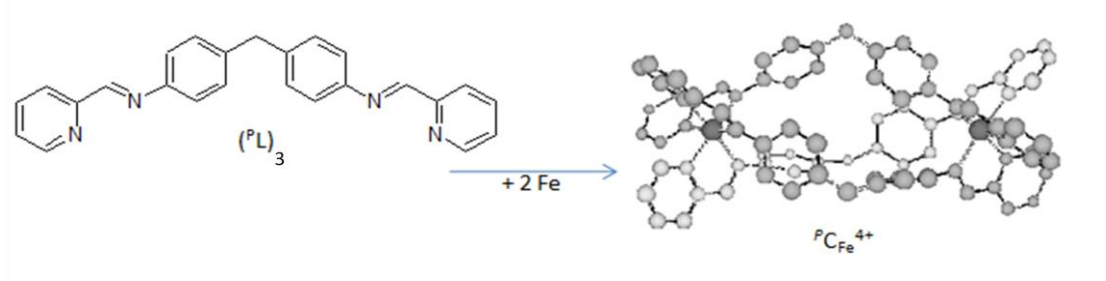


Figure 1-18 (left) Parent Ligand, (^{P}L) = ($C_{25}H_{20}N_4$), (right) ($^{P}C_{Fe}^{4+}$) = $[Fe_2(C_{25}H_{20}N_4)_3]Cl_4$

The tetracation is approximately cylindrical in shape with length 19 Å and diameter 11 Å, hereafter it will be referred to only as $^{P}C_{Fe}^{4+}$. This cylinder is too big to bind in the minor groove of DNA, but appears to have just the right shape and size to lie along the major groove[201] Circular dichroism (CD) and linear dichroism (LD) experiments showed that upon binding to the cylinder, DNA undergoes dramatic conformational change unprecedented in synthetic DNA binders,

LD is the difference in absorption of light polarised parallel and perpendicular to an orientation direction. In the case of DNA-ligand systems it is usually a good indicator of the orientation of the ligand on the DNA. However, for the racemic metallo helicate, ***R***, the changes in the DNA LD spectrum upon addition of the ligand were dominated by the loss of DNA orientation associated with the bending or coiling of the DNA by the ligand. Flow LD data for the two enantiomers with ct-DNA, a randomly sequenced DNA originating from calf thymus, was collected to determine the average bending effect of the enantiomers compared with ***R***. The LD signal of ct-DNA in the presence of the ***M*** enantiomer is very small in the Metal Ligand Charge Transfer (MLCT) band and the DNA signal itself decreases dramatically even at very small mixing ratios.

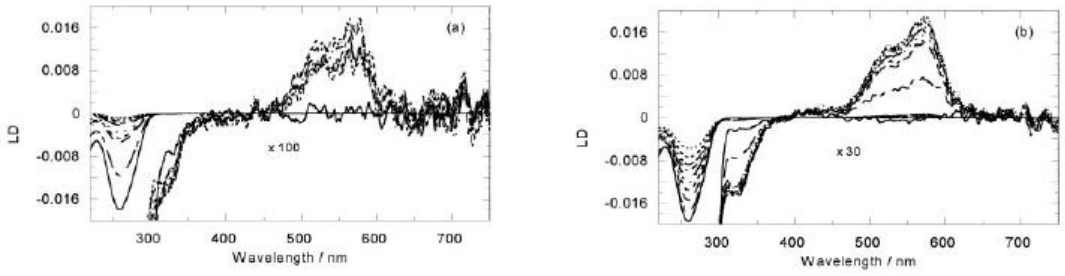


Figure 1-19 LD spectra of free and (a) M or (b) P-bound ct-DNA (200μM). Solid line is DNA alone, decreasing mixing ratios correspond to decreases in the 260nm LD signal. Ratios are (a) 100:1, 50:1, 40:1, 35:1, 30:1, 27:1, 25:1, 22:1, 20:1, and (b) 100:1, 50:1, 40:1, 35:1, 30:1; 25:1, 20:1, 16:1, 13:1. Long wavelength region is shown to scale and x 100 (a) or x 30 (b) as indicated.

The DNA solution was also observed to become significantly more viscous on the addition of ***M***. At 100:1 DNA base/***M***, a 40% orientation loss was observed, all signs of dramatic coiling

of the ct-DNA. At 50:1 two-thirds of the LD DNA intensity signal is lost, and at 20:1 ratio less than 5% of the signal remains. The ***P*** enantiomer behaves differently, although still giving an increase in solution viscosity. The changes induced in the DNA LD region are less. For a 100:1 ratio there is only 8% loss in the LD signal intensity and for a 15:1 ratio a 43% loss. A change in the DNA LD signal shape was observed for ratios below 20:1, with the signal becoming broader suggesting the existence of two different DNA-binding modes.

To further understand the effect of $^{P}C_{Fe}^{4+}$ on DNA, some Atomic Force Microscopy (AFM) images of the cylinder with a 'random' sequence linearised plasmid DNA were acquired by Professor V. Moreno and Dr M.J. Prieto, University of Barcelona, Spain [201].

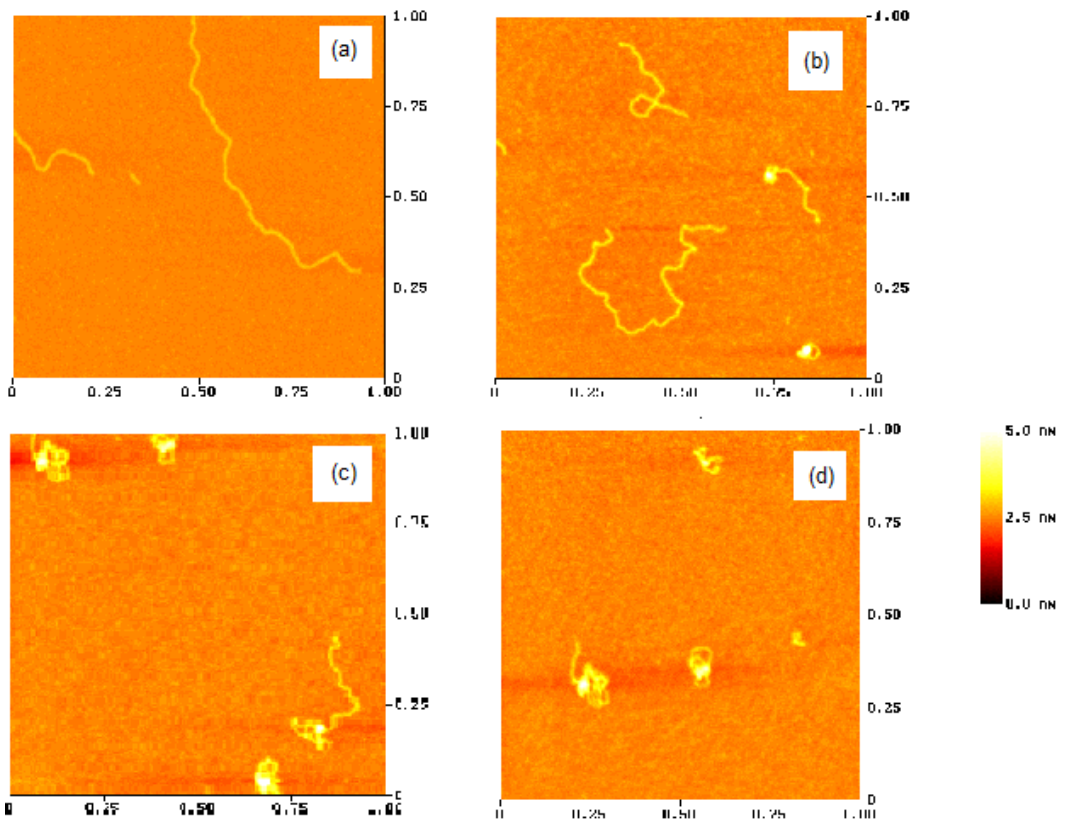


Figure 1-20 AFM images of plasmids with $^{P}C_{Fe}^{4+}$ at different ratios, (a) DNA without metal complex (b) Complex : DNA ratio = 0.1 (c) Complex : DNA ratio = 0.3 (d) Complex : DNA ratio = 0.5. White patches are areas of high DNA concentration.

At moderate helicate ligand concentration (10:1), the presence of $^{P}C_{Fe}^{4+}$ seems to affect the plasmid structure by homogeneously increasing the size and the number of kinks observed. At

reasonably high loading (10:3), cooperative intramolecular coiling is observed, with some fully coiled strands observed in the presence of uncoiled but kinked plasmids and other strands being half coiled and half kinked. Furthermore, high $^{P}C_{Fe}^{4+}$ loading (10:5) results in aggregation of the plasmids into balls.

Previous simulations have shown $^{P}C_{Fe}^{4+}$ to bind in the major groove of DNA and consequently affect the DNA conformation resulting in coiling around the bound cylinder[202]. With the knowledge that $^{P}C_{Fe}^{4+}$ successfully coils DNA two different cylinders are going to be simulated during this study following docking with DNA. The simulations undertaken in this study have been designed to investigate the effect of shape on the DNA binding properties. For the derivative of $^{P}C_{Fe}^{4+}$, named C_{Fe}^{4+}, see **Figure 1-211**, the increased diameter of the cylinder will further investigate the need for a cylinder to be a good fit in the major groove in order for it to induce DNA bending. The active parent cylinder $^{P}C_{Fe}^{4+}$ was a snug fit within the DNA major groove.

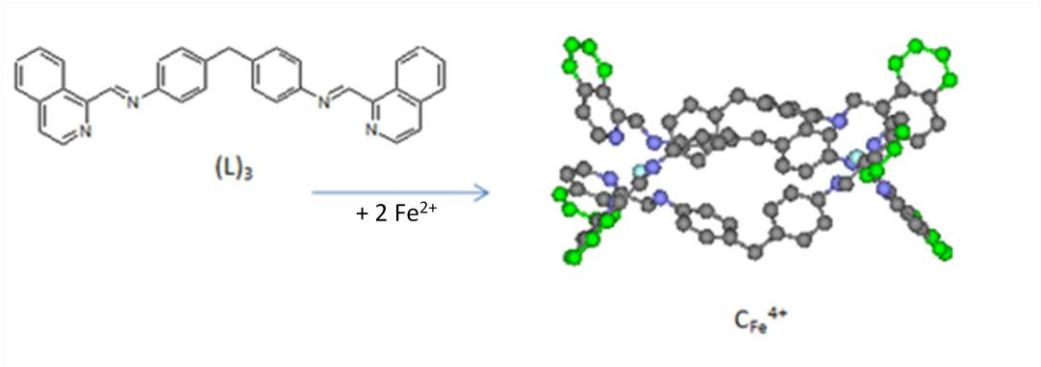


Figure 1-21 (left) Ligand, (L) = ($C_{33}H_{28}N_4$), (right) (C_{Fe}^{4+}) = $[Fe_2(C_{33}H_{28}N_4)_3]Cl_4$, additional structure not present in the parent cylinder highlighted in green

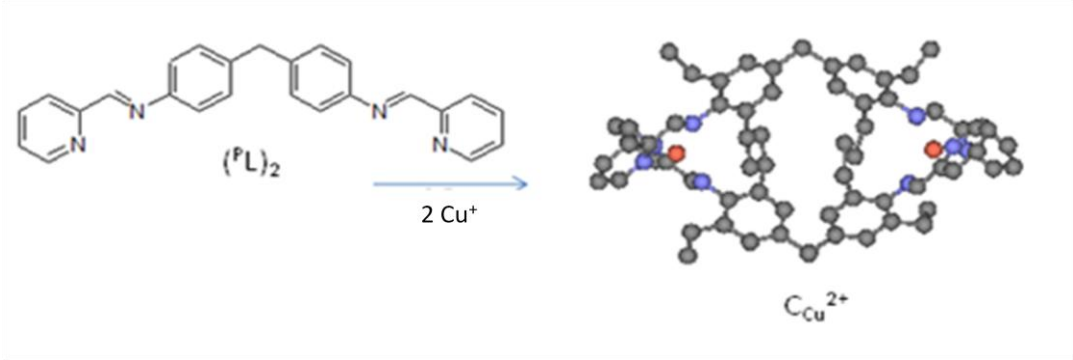


Figure 1-22

(left) Parent Ligand, (^{P}L) = ($C_{25}H_{20}N_4$), (right) (C_{Cu}^{2+}) = $[Cu_2(C_{25}H_{20}N_4)_3]Cl_2$

To investigate DNA-binding interactions with a cylinder smaller than $^{P}C_{Fe}^{4+}$, the dicopper (I) metallo-supramolecular cylinder $[Cu_2(C_{25}H_{20}N_4)_3]Cl_2$, C_{Cu}^{2+} will also be simulated. Simulation of these two cylinders with different charge as well as shape will also offer insight into which characteristics are most effective at inducing an effect on DNA.

Reaction of the parent ligand, ^{P}L, with monocations, copper (I), leads to a dinuclear double-helical metallo-supramolecular cylinder assembled by copper(I) centers[203] with 2:2 stoichiometry, $[M_2L_2]^{2+}$.[204] In solution, these complexes are an equilibrium mixture of two dimeric isomers: a helicate (rac-isomer) and a metallocyclophane (meso-isomer), Figure 1-24 Ligand with ethyl groups, L^a, exclusively present as the helical isomer in solution. To obtain exclusively the helical isomer for practical experiment ethyl groups were introduced onto the spacer, this is the ligand used in this work, L^a.

Figure 1-23 Ligand, L, forms dinuclear double-stranded Complex $[M_2L_2]^{2+}$ a component of both the rac-isomer and meso-isomers

.

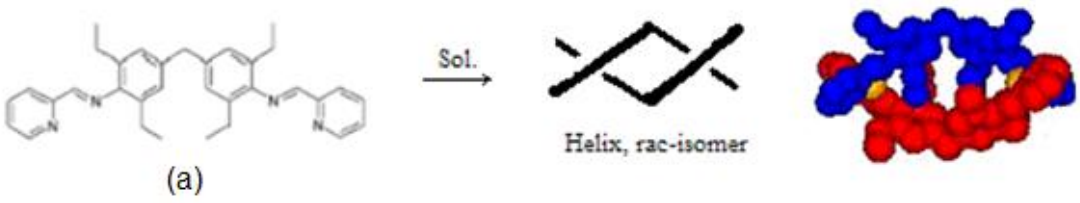


Figure 1-24 Ligand with ethyl groups, L^a, exclusively present as the helical isomer in solution.

Addition of ethyl groups to the central spacer destabilizes the cyclophane configuration so that only $[M_2(L^a)_2]^{2+}$ helicate species are present in solution. This double-helical $[Cu_2(L^a)_2]^{2+}$ cation represents a dicationic cylinder. By including C_{Cu}^{2+} in this study it has made it possible to consider a number of variables relating to the properties of the different cylinders.

To understand the differences and similarities between the C_{Fe}^{4+} and C_{Cu}^{2+} a brief comparison follows. Lengths along the cylinder axis between the extreme carbon atoms or the extreme hydrogen atoms[205] are: for C_{Cu}^{2+} 16.6 and 18.5 Å, respectively, and for C_{Fe}^{4+} 18.7 and 20.6 Å, respectively. From these measurements it is clear to see that C_{Cu}^{2+} is nearly 2 Å shorter than C_{Fe}^{4+}. However the biggest difference is in the length of the radius: which for C_{Cu}^{2+}, radius-to-carbon is 4.5 Å and radius-to-hydrogen is 5.5A, and for C_{Fe}^{4+}, radius-to-carbon is 7.1 Å and radius-to-hydrogen is 8.1 Å. In summary the two cylinders differ in external dimensions with the increased radius of C_{Fe}^{4+} at either end of the molecule being used to test the importance of a snug fit in the major groove as part of the mechanism by which DNA bending occurs. This increased size is expected to inhibit the cylinder from copying the binding mode of ${}^{P}C_{Fe}^{4+}$, which was able to bind deep in the major groove. In addition to the differences between the dimensions between extreme atoms within the molecules the two cylinders have other differences in their shape and charge.

The C_{Cu}^{2+} cylinder viewed down the metal-metal axis is seen to be squarer in shape with a more consistent radius along the length of the cylinder. The shape C_{Fe}^{4+} is similar to that of a propeller when viewed down the metal-metal axis, Figure 1-25 With the extended structure at either end of the molecule the smooth sides of the cylinder have been interrupted with the overall shape from the side being more concave. The difference in shape and charge will contribute to the differences shown in the following results.

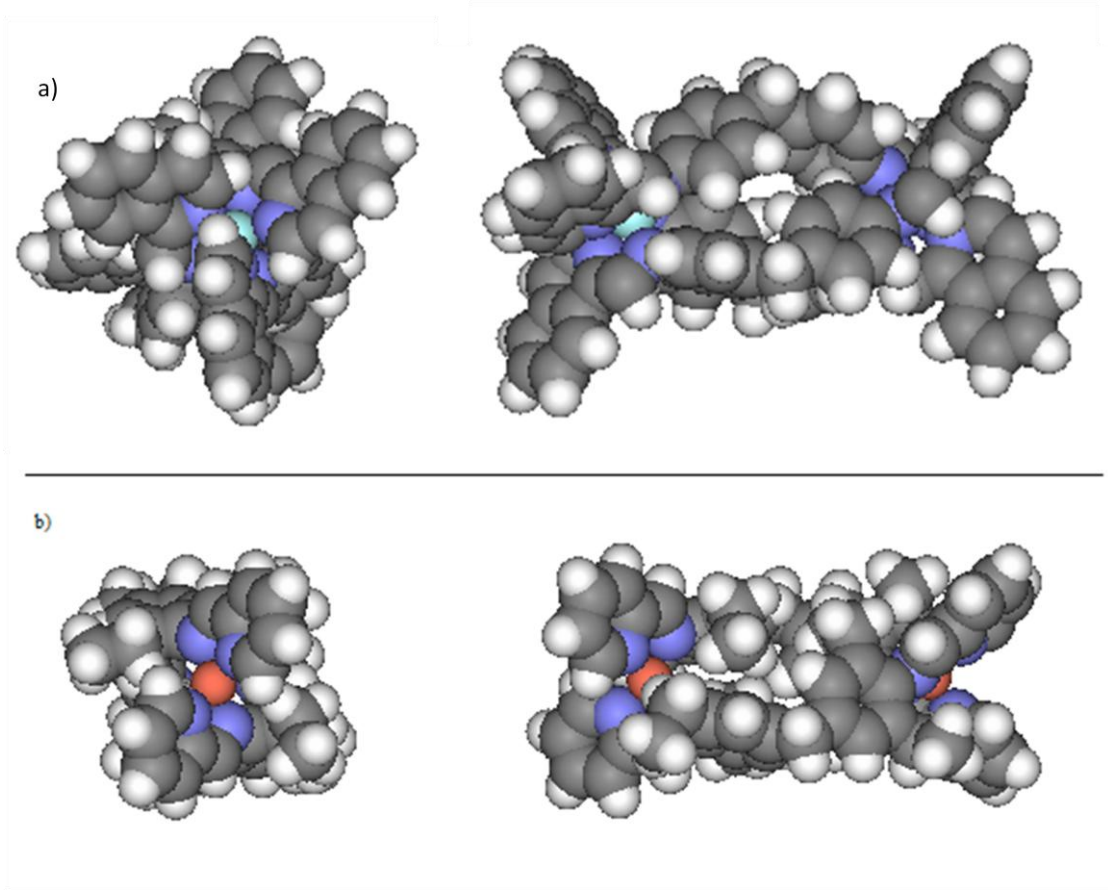


Figure 1-25 Space-filling images showing the end (left) and side views (right) of (a) $[Fe_2(L)_3]^{4+}$ and (b) $Cu_2(L^a)_2]^{2+}$. (not to scale relative to each other)

C_{Cu}^{2+} has a lower charge compared to C_{Fe}^{4+} allowing us to investigate the effect of charge on strength of binding and extent of DNA coiling.

The aim of the work outlined in this thesis was to use molecular simulation techniques to gain insights at the molecular level, into the DNA binding of the tetracationic bis iron (II)

supramolecular derivative cylinder and a related dicationic pyridylimine based dicopper (I) supramolecular cylinder and the subsequent DNA conformational changes. We were intrigued to explore the effects of using dicopper(I) cylinders for two reasons. Firstly, the dicopper(I) cylinders would have a relatively low charge, which would allow us to probe the effect of charge on strength of binding and extent of DNA coiling. Secondly, copper complexes of diamine ligands, such as phenantholine, are known to exhibit oxidative DNA-cleavage activity that is thought to proceed by means of Fenton-generated hydroxyl-radical or copper-bout oxidants, such as $[CuO]^{+}$ or $[Cu(OH)]^{2+}$ [206,207] (and there has been some recent interest[208] in polynuclear analogues). This opens up the recent exciting possibility that dicopper(I) cylinders might act as artificial nucleases.

There has been a single report of DNA binding of copper(I) supramolecular double-helicates based on the oligobipyridine unit.[209] Binding to DNA was confirmed, although the precise binding mode was not unambiguously established. Laboratory experiment of the dicationic copper(I) cylinders modelled herein are known to bind strongly to DNA and do indeed cleave DNA in the presence of an oxidising agent[210]. In contrast to mononuclear copper-containing artificial nucleases, such as the Clip-Phen agents of Meunier,[211] which usually cleave a single DNA strand.

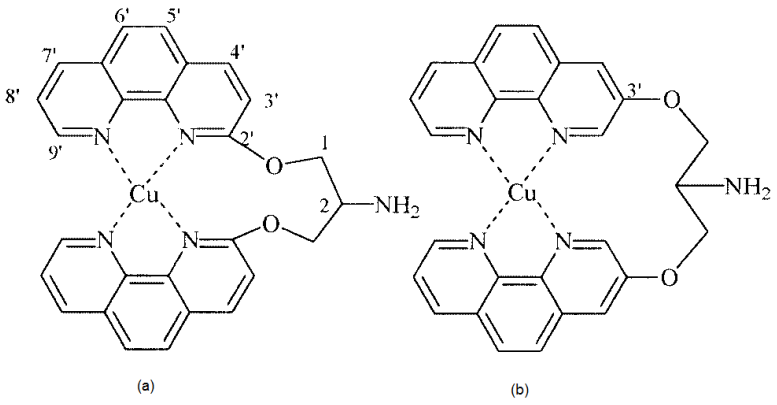


Figure 1-26 (a) Structure of 2-Clip-Phen (b) 3-Clip-Phen. Numbering corresponds to NMR assignments; the type of coordination with Cu attempted from the design of the ligands is shown.

The dinuclear double-helical metallo-supramolecular cylinders, C_{Cu}^{2+} show a tendency to cleave both DNA strands at the same site.[212] As expected the CD signal for the racemic mixture of *P* and *M* copper helices has no intrinsic CD signal. Therefore the signal present above 300 nm, a range above which no DNA signal is observed, indicates interaction of the metal complex with the chiral DNA. With ICD signals present in the DNA region there are several possible explanations. Either there is a competing ICD signal (from the cylinder at the same wavelength, there are ligand-induced DNA ICD, or small structural changes have taken place in the DNA as a result of the helicate binding.

In the absence of DNA the small size of the cylinder prevents it from being oriented by the LD flow method with the consequence of no signal being generated. However, on addition of the cylinder to ct-DNA, of a constant concentration, in a titration series signals were generated in addition to the negative signal in the 220-300nm range characteristic of DNA base pairs lying approximately perpendicular to the DNA axis.

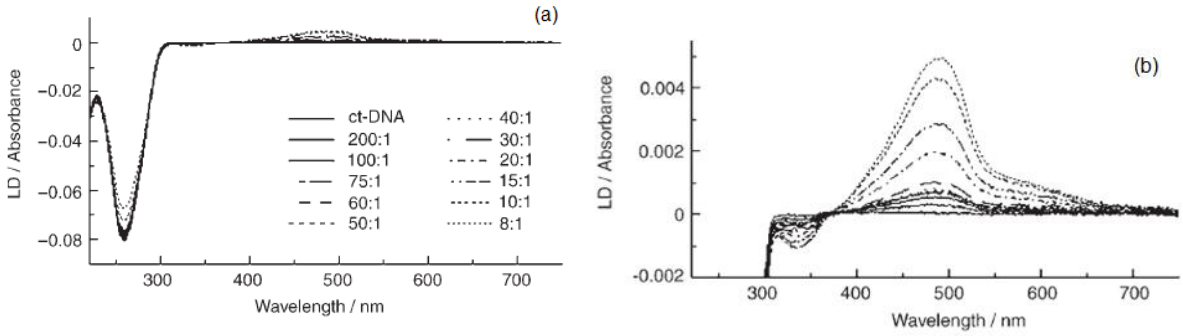


Figure 1-27 LD spectra of free ct-DNA (500 μM; 10mM Na cacodylate, 20 nM NaCl) and in the presence of $[Cu_2(L^a)_2]^{2+}$. Mixing ratios ([DNA base]:$[Cu_2(L^a)_2]^{2+}$]) are indicated, (a) full scale (b) expanded scale.

The positive signal observed from 400-700 nm upon the addition of C_{Cu}^{2+} to the DNA solution and a smaller negative signal is apparent for the in-ligand band at ~340 nm. The presence of these signals confirm that the cylinder is binding to the DNA in a specific orientation. The increase of the metal-to-ligand charge-transfer (MLCT) band increases as more C_{Cu}^{2+} is added,

indicating that more helicate is binding to DNA upon each addition, in agreement with the CD data. In contrast to the effect previously observed with the iron helicate, the copper cylinder has little bending effect on the DNA (little or no loss in the DNA signal at ~260nm confirms DNA still orientated). With the DNA not coiled Hannon *et al* were able to use the DNA LD to determine the orientation of the DNA and, hence, estimate the orientations of the cylinder transition moments on the DNA from calculation of the reduced LD[213]. It follows that the long axis of the copper helicate lies at approximately 70^{o} to the axis of the DNA helix. One of the aims for this research is to contribute to this understanding with the use of molecular dynamics simulations of this and related systems.

Fluorescence competition-binding assays have been used to estimate the strength of binding. The method monitors the displacement of ethidium bromide (EB) from DNA by following the decrease in its fluorescence intensity as it is displaced from DNA into an aqueous environment. The C_{Cu}^{2+} displaces EB and the EB fluorescence is quenched. This study shows that the copper cylinder binds more strongly to DNA than does EB or $[Ru(phen)_3]^{2+}$ and less strongly than the iron cylinder. This confirms that electrostatic charge is an important factor in determining the strength of binding to the anionic DNA (the tetracations bind much more strongly than the dications). Nevertheless, the shape is also a significant factor, with the cylindrical dication exhibiting a higher binding constant than the spherical $[Ru(phen)_3]^{2+}$, despite the fact that the latter will have a higher charge density. Thus, the shape (and the fit on the DNA) of the supramolecular cylinder is indeed important.

Attempts to study the interaction between the C_{Cu}^{2+} and a decamer $d(TATGGCCATA)_2$ were hindered by the precipitation of a red-brown solid when trying to achieve a cylinder:duplex ratio of 1:1. Consistent with the results for high loading ratios in the more dilute CD and LD

studies it seems likely that the cylinder precipitates the DNA at the concentrations required for NMR studies.

To assess the DNA-cleavage activity the interaction of the cylinder with pUC19 plasmid DNA was studied. Gel electrophoresis was used to monitor a variety of conditions. Interestingly addition of both cylinder and peroxide led to a reduction in supercoiled DNA and a corresponding increase in relaxed DNA. Relaxation of the supercoiling requires nicking of a single strand and, thus, implies that cylinder-induced DNA-strand scission occurs under these conditions. Increasing the cylinder concentration leads to a further increase in relaxed DNA and a decrease in supercoiled DNA. Intriguingly, a band corresponding to linearised DNA (arising from cleavage of both strands at the same site) is apparent even well before the supercoiled DNA has disappeared. This is unusual and could be evidence that the two-fold symmetry of the copper cylinder and the approximate two-fold symmetry of the DNA helix make it possible to propose that the complex has more than one reactive centre and cuts both DNA strands simultaneously. A similar gel electrophoresis experiment with ct-DNA (a linear, polymeric DNA, as used in the spectroscopic experiments) confirmed that the effect is not merely restricted to circular DNAs. After addition of cylinder and peroxide to the ct-DNA, the bands corresponding to long DNA disappeared and only fast-running, short DNA fragments were observed.

On increasing the concentration of the cylinder further evidence of DNA being cleaved into short fragments with high mobility within the gel was seen, confirming that C_{Cu}^{2+} is essential for DNA cleavage.

Molecular-level images of the pBR322 plasmid DNA and C_{Cu}^{2+}, obtained by using tapping-mode atomic force microscopy (AFM) provide confirmation of the DNA-cleavage activity,

Figure 1-28. Plasmid samples incubated for one hour with cylinder alone show no evidence of strand scission (Figure 1-28a, b). Whereas the samples incubated with 6μM C_{Cu}^{2+} and hydrogen peroxide shows clear evidence of DNA-strand scission to give linear fragments, (c). This effect is even more dramatic upon addition of a higher concentration of C_{Cu}^{2+} cylinder (20 μM) to the DNA, (d). In contrast to the AFM images of ${}^{P}C_{FE}^{4+}$ where coiling of the DNA around the cylinder is seen.

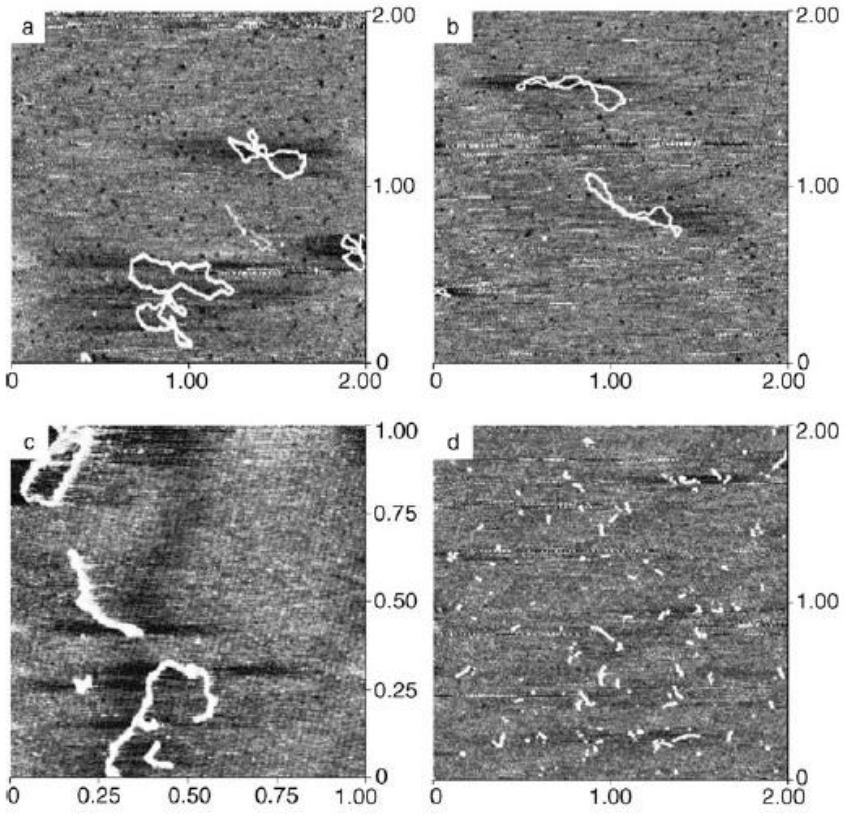


Figure 1-28 pBR322 DNA with (a) 6μM copper(I) cylinder, (b) 20μM copper(I) cylinder, (c) 6μM copper(I) cylinder with H_2O_2 and (d) 20μM copper(I) cylinder with H_2O_2. All samples were incubated for 1 hr at 20°C. The dimensions are in μm. [210]

This work is a continuation of the research undertaken in the group by Syma Khalid[214] for which a summary is provided below. All simulations were performed with the CHARMM22 force field and with variations of the ${}^{P}C_{Fe}^{4+}$ cylinder (Figure **1-18**). The symbols defined in Table 3 apply to this review section and are all derivatives of the ${}^{P}C_{Fe}^{4+}$ cylinder.

The aim of the research reported to date was to achieve a greater understanding of the interaction between DNA and the macromolecular ligand, $[Fe_2(C_{25}H_{20}N_4)_3]Cl_4$, (referred to as the *cylinder*). Molecular dynamics simulations were used to investigate the DNA binding mode of the M and P-enantiomers of the cylinder and the effect of the cylinder charge and shape on its DNA binding properties. A summary of the range of simulations, and the notation used to refer to them, is given in Table 3.

Symbol	**Description of system**
m-C_1^{4+}	M–enantiomer with an overall charge of +4. Major groove binding site from docking calculation.
m-C_2^{4+}	Similar to m-C_1^{4+}. Slightly different binding site, also from docking calculation
p-C_3^{4+}	P–enantiomer with an overall charge of +4. Major groove binding site from docking calculation
p-C_4^{4+}	P–enantiomer with an overall charge of +4. Manually docked into the minor groove.
m-C_5^{0}	M–enantiomer, overall neutral. Similar binding site to m-C_1^{4+}
m-C_6^{0}	Similar to m-C_5^{0}. Slightly different binding site.
m-C_7^{2+}	M–enantiomer with an overall charge of +2. Same binding site as m-C_5^{0}
m-C_8^{2+}	M–enantiomer with an overall charge of +2. Same binding site as m-C_6^{0}
m-D_9^{4+}	M–enantiomer with an overall charge of +4. Major groove binding. Derivatised by addition of six methyl groups.

Table 3 A summary of the simulations described in previous work[214].

M–enantiomer

Docking calculations were used to confirm the major groove binding of the M–cylinder. For the two most energetically favoured DNA-cylinder conformations, docking calculations indicated the most favourable binding site to be in the major groove with the cylinder lying along the groove.

Solvated simulations showed that the DNA wrapped closely around the cylinder, which with multiple binding cylinders could lead to the experimentally observed coiling. There is some potential disruption of the DNA double helix in the vicinity of the cylinder binding site. Mismatched base pairs formed in both simulations.

Energetic analysis indicated the m-C_1^{4+} to be the more favoured binding geometry. However the comparison of the average DNA backbone after 2 ns was very similar in both cases.

P–enantiomer

The DNA in the p-C_3^{4+} system, in which the cylinder was located in the major groove showed a greater response to the cylinder than the p-C_4^{4+} system in which the cylinder was located in the minor groove. However, energetically p-C_4^{4+} is the more favoured binding geometry. So on the basis of the simulations discussed here, although the DNA response to the cylinder would be greater if it bound in the major groove, it in fact binds in the minor groove. This is in agreement with experimental results which showed less response of the DNA to the P–enantiomer, compared with the M–enantiomer.

Even if the P–enantiomer were to target the major grove, substantial differences between the two enantiomers would still be evident. Most importantly, whereas the m-C fit closely into the major grove, and remained there while the DNA coiled, the P–enantiomer tended to move out of the major groove, so that in the end, only one end of the cylinder remained in contact with the DNA.

An important point to note from this study was that the docking calculation failed to pick up the p-C_4^{4+} system as a favourable binding geometry. There could be two reasons for this. Firstly, it

is possible that a 2 ns docking calculation is not long enough to adequately sample binding sites, however visual inspection of the docking calculation trajectory showed the cylinder to move into the vicinity of the minor groove and hence it can be concluded that the docking calculation was long enough.

Effect of the cylinder charges

The DNA in the m-C_5^{0} system showed a greater response to the cylinder than the DNA in the m-C_6^{2+} system, *i.e.* the m-C_5^{0} DNA wrapped more tightly around the DNA. There was a lack of kinks and mismatched base pairs in C_5^{0} and C_6^{2+} systems compared to the C_2^{4+} system over the same timescale. The DNA in the C_7^{0} system was wrapped around the cylinder. In contrast the C_8^{2+} cylinder was no longer lying in the major groove by the end of the simulation. The base pairs in the C_7^{0} DNA were generally stable whereas in the C_8^{2+} simulation, disruption of the base pairs had resulted in the two strands of the DNA double helix coming apart at one end of the molecule. The variation of the DNA groove-widths did not show a charge dependence.

These results suggested that the lower charges on the cylinder induced fewer disruptions of the DNA base pairs. The disruption of the base pairs could restrict the wrapping of the DNA around the cylinder. On the basis of these results, not only is the DNA-cylinder system largely insensitive to the cylinder charges, but high charges may reduce the ability of the DNA to wrap effectively around the cylinder.

Effect of the cylinder shape

The results obtained from the docking calculation and simulation indicate a major groove-binding mode for m-D_9^{4+}. The DNA shows a response to the cylinder on the timescale of this simulation that is similar to that seen in the m-C_1^{4+} cylinder. In other words, the addition of six methyl groups to the m-C^{4+} cylinder made little difference to the wrapping of the DNA on the 5

ns timescale. However, the cylinder did move away from the DNA, whereas in the m-C^{4+} simulations it remained near to its starting position. Energetic analysis suggests binding of the derivatised cylinder is less favourable than the m-C^{4+}.

Competitive fluorescence binding studies

A series of six fluorescence competitive binding experiments designed to compare the binding modes of both enantiomers of the cylinder to DNA solutions to which the minor groove binding ligand Hoechst 33258 had already been added, were undertaken. One random sequence and two specific DNA sequences (poly[$(AT)_2$] and poly[$(GC)_2$]) were investigated in this experiment. The binding experiments showed that both enantiomers of the cylinder bind to the Hoechst-DNA complex in a manner that results in stoichiometric disassociation of the Hoechst-DNA complex. This was true of all three DNA sequences with both M and P–enantiomers of the cylinder. The fluorescence spectra obtained from both sets of experiments were similar and therefore it was not possible to gain further insights into any potential differences between the binding modes of the m and p-cylinders.

Current Research

The research documented within this thesis will extend the investigation to include the CHARMM27 force field. Solvated uncomplexed DNA systems have been simulated for reference against which to compare the bound systems. While reference systems have been performed for earlier studies it has been necessary to redefine the reference systems due to the introduction of new variables such as force-field and sequence. Each cylinder will then be simulated with the following three DNA sequences d$(CGCGCATATACG)_2$, d$(ATATATATATAT)_2$, and d$(CGCGCGCGCGCG)_2$.

These reviews from the current literature have been considered in the context of the aims for this research and helped to form the hypotheses included at the end of this chapter.

1.6 Thesis outline

Synthetic molecules that target the major groove in a sequence-selective way are a major goal in molecular medicine. Recently a major step has been taken toward achieving this goal: a novel cylinder has been developed that binds strongly into the major groove of DNA. Experimental techniques have provided some information regarding the binding strength and preferred binding sites of various cylinders on DNA as summarized in the earlier section. From the experimental data the parent cylinder is considered to bind in the major groove and is able to induce dramatic conformational changes in the DNA; these are unprecedented effects with synthetic DNA binders. However, gaining molecular level information in such a macromolecular system is challenging. Molecular dynamics (MD) simulations can provide information at the molecular level that is complementary to experiment and therefore is an ideal way to get a better understanding of this system. In this work we present the results of various MD simulations designed to probe the DNA-cylinder system interaction by exploiting the differences in charge and shape between $^{P}C_{Fe}{}^{4+}$, $C_{Fe}{}^{4+}$ and $C_{Cu}{}^{2+}$ to enable us to better understand DNA binding interactions.

In the context of the literature review above and the technical descriptions that follow the following hypothesis have been proposed for consideration during this research.

1. If CHARMM22 has been refined to more accurately model the equilibrium between B-DNA and A-DNA then the reference simulations will be stable with the DNA maintaining its original B-Form during extended uncomplexed simulations.

2. If the size of a supramolecular cylinder is related to its DNA binding mode then increasing the cylinders size will affect its effectiveness at binding in the major groove of B-DNA.
3. If the ability of a supramolecular cylinder to cause conformational changes to DNA is related to the electrostatic charge of the supramolecular cylinder then a greater change in the DNA conformation should be seen for DNA-cylinder systems when C_{Fe}^{4+} is present compared with C_{Cu}^{2+}.
4. If the stability of the DNA is related to the base pair sequence of which it is made up then $d(ATATATATATAT)_2$ will respond more to the presence of supramolecular cylinders than $d(CGCGCGCGCGCG)_2$.

Following on from the review presented in chapter 1, chapter 2 will develop the theoretical background behind computer simulations of DNA and DNA binding systems. Equipped with the understanding of how to perform molecular dynamics simulations the discussion in chapter 3 explains a variety of analytical methods that are available and will be used to quantitatively analyse the output from the simulations. The results will be presented in chapters 4 and 5 initially discussing the variations in the available force-fields for this project. In these chapters the results from the initial high temperature docking of supramolecular systems through to the analysis of multi-nanosecond trajectories, up to 5ns, will be discussed. Chapter six is included as an example of research used to gain an insight into a more simplistic system and the water within it with a view to developing the concepts to analyse the much more complex DNA, DNA-cylinder systems which form the heart of this thesis.

2 Computational Methods

2.1 Statistical mechanics

Statistical mechanics (SM) is the branch of physical sciences that studies macroscopic systems from a molecular point of view. The goal is to understand and to predict macroscopic phenomena from the properties of individual molecules making up the system. The system could range from a collection of solvent molecules to a solvated protein-DNA complex. Within SM the thermodynamic state of a system is defined by the parameters, the temperature, T, the pressure, P, and the number of particles, N. Other thermodynamic properties may derive from the equations of state and other fundamental thermodynamic equations.

Molecular Dynamics (MD) is a method for exploring a set of microstates. The connection between microscopic simulations and macroscopic properties is made via SM which provides the rigorous mathematical expressions that relate macroscopic properties to the distribution and motion of the atoms and molecules of the N-body system. MD simulations provide the means to solve the equation of motion for the particles and evaluate these mathematical formulas. With MD simulations, one can study both thermodynamic properties and time dependent phenomenon.

The microscopic state of a system is defined by the atomic positions, q, and momenta, p, these can also be considered as coordinates in a multidimensional space called phase space. For a system of N particles, this space has 6N dimensions. A single point in phase space, denoted by $\boldsymbol{\Gamma}$, describes the state of the system. An ensemble is a collection of points in phase space satisfying the conditions of a particular thermodynamic state. A molecular dynamics simulation generates a sequence of points in phase space as a function of time; these points define an

ensemble (the microcanonical ensemble for Newtonian MD), and they correspond to the different configurations of the system and their respective momenta. Several different ensembles are described below.

- Microcanonical ensemble (NVE): The thermodynamic state characterised by a fixed number of atoms, N, a fixed volume, V, and a fixed energy, E. This corresponds to an isolated system, and is the natural ensemble for MD, since *E* is conserved by the dynamics of any conservative system.
- Canonical Ensemble (NVT): This is a collection of all systems whose thermodynamic state is characterised by a fixed number of atoms, N, a fixed volume, V, and a fixed temperature, T.
- Isobaric-Isothermal Ensemble (NPT): This ensemble is characterised by a fixed number of atoms, N, a fixed pressure, P, and a fixed temperature, T.
- Grand canonical ensemble (μVT): The thermodynamic state for this ensemble is characterised by a fixed chemical potential, μ, a fixed volume, V, and a fixed temperature, T.

A physical experiment is usually performed using a macroscopic sample that contains an extremely large number of atoms or molecules sampling an enormous number of conformations. In statistical mechanics, averages corresponding to experimental observables are defined in terms of ensemble averages. An ensemble average is an average taken over a large number of replicas of the system being considered simultaneously.

In statistical mechanics, average values are defined as ensemble averages. The ensemble average is given by

$$\bar{A} = \frac{\int A e^{-\beta H(q_1 q_2, \ldots q_M, p_1, p_2, \ldots p_N)}}{\int e^{-\beta H(q_1 q_2, \ldots q_M, p_1, p_2, \ldots p_N)}}$$

Equation 2-1

where,

$\bar{A}$ is the ensemble average of the system property A,

β is $\frac{1}{kT}$, known as thermodynamic beta,

H is the Hamiltonian (or energy function) of the system in terms of the set of coordinates *qi* and their conjugate generalised momenta *pi*, and

$d\tau$ is the volume element of the classical phase space of interest.

The denominator in this expression is know as the partition function, and is denoted by the letter Z.

In SM the partition function Z is an important quantity that encodes the statistical properties of a system in thermodynamic equilibrium. It is a function of temperature and other parameters. Most of the thermodynamic variables of the system, such as the total energy, free energy, entropy, and pressure, can be expressed in terms of the partition function or its derivatives.

There are actually several different types of partition functions, each corresponding to different types of statistical ensemble. The canonical partition function applies to a canonical ensemble, in which the system is allowed to exchange heat with the environment at fixed temperature, volume and number of particles. The grand canonical partition function applies to a grand canonical ensemble, in which the system can exchange both heat and particles with the environment, at fixed temperature, volume, and chemical potential.

For a canonical ensemble of a thermodynamically large system that is in constant thermal contact with the environment, which has temperature *T*, with both the volume of the system and the number of constituent particles fixed. The exact states (microstates) that the system can

occupy can be labelled as j (j = 1, 2, 3, …), and denote the total energy of the system when it is in microstate j as Ej. Generally, these microstates can be regarded as discrete quantum states of the system.

The canonical partition function is

$$Z = \sum_j e^{-\beta Ej}$$

Equation 2-2

Where the "inverse temperature" β is conventionally defined as

$$\beta \equiv \frac{1}{k_B T}$$

Equation 2-3

With k_B denoting Boltzmann's constant.

In classical statistical mechanics, it is not really correct to express the partition function as a sum of discrete terms. In classical mechanics, the position and momentum variables of a particle can vary continuously, so the set of microstates is actually uncountable. In this case, some form of coarse graining procedure must be carried out, which essentially amounts to treating two mechanical states as the same microstate if the differences in their position and momentum variables are not too large. The partition function then takes the form of an integral. For instance, the partition function of a gas of N classical particles is

$$z = \frac{1}{N!h^{3N}} \int \exp[-\beta H(p_1 ... p_N, x_1 ... x_N)] d^3 p_1 ... d^3 p_N d^3 x_1 ... d^3 x_N$$

Equation 2-4

where h is some infinitesimal quantity with units of action (usually taken to be Planck's constant), and H is the classical Hamiltonian.

The partition function is a function of T and microstate energies E_1, E_2, E_3, *etc.* The microstate energies are determined by other thermodynamic variables, such as the number of particles and

the volume, as well as microscopic quantities like the mass of the constituent particles. This dependence on microscopic variables is the central point of statistical mechanics. With a model of the microscopic constituents of a system, one can calculate the microstate energies, and thus the partition function, which will then allows all other thermodynamic properties of the system to be calculated.

The partition function can be related to thermodynamic properties because it has a very important statistical meaning. The probability *Pj* that the system occupies microstate *j* is

$$p_j = \frac{1}{Z} e^{-\beta Ej}$$

Equation 2-5

This is a well known Boltzmann factor. The partition function thus plays the role of a normalizing constant ensuring that the probabilities add up to one.

$$\sum_j P_j = \frac{1}{Z} \sum_j e^{-\beta E_j} = \frac{1}{Z} Z = 1$$

Equation 2-6

This is reason for calling Z the "partition function": it encodes how the probabilities are partitioned among the different microstates, based on their individual energies. The letter Z stands for the German work Zustandssumme, "sum over states".

2.1.1 Calculating the thermodynamic total energy

The thermodynamic value of the total energy is the ensemble average for the energy, which is the sum of the microstate energies weighted by their probabilities:

$$< E >= \sum_j E_j P_j = \frac{1}{Z} \sum_j E_j e^{-\beta E_j} = \frac{1}{Z} \frac{\partial}{\partial \beta} Z(\beta, E_1, E_2, ...) = -\frac{\partial \ln Z}{\partial \beta}$$

Equation 2-7

Or, equivalently,

$$\langle E \rangle = k_B T^2 \frac{\partial \ln Z}{\partial T}$$

Equation 2-8

Incidentally, one should note that if the microstate energies depend on a parameter λ in the manner,

$$E_j = E_j^{(0)} + \lambda A_j \qquad \text{for all } j$$

Equation 2-9

where λ is a coupling parameter. And the Ensemble average, A, is

$$< A >= \sum_j A_j P_j = -\frac{1}{\beta}\frac{\partial}{\partial \lambda} \ln Z(\beta, \lambda)$$

Equation 2-10

This then allows calculation of many microscopic quantities. First the quantity is artificially added to the microstate energies (or Hamiltonian), and then the new partition function is calculated before setting λ to zero in the final expression.

The relationship of the partition function to other thermodynamic variables is as follows.

Thermodynamic energy, see Equation 2-7.

The variance in the energy (or "energy fluctuation") is

$$< \delta E^2 > \equiv < (E - < E >)^2 > = \frac{\partial^2 \ln Z}{\partial \beta^2}$$

Equation 2-11

The heat capacity is

$$C_v = \frac{\partial < E >}{\partial T} = \frac{1}{k_B T^2} < \delta E^2 >$$

Equation 2-12

The entropy is

$$S \equiv -k_B \sum_j P_j \ln P_j = k_B(\ln Z + \beta < E >) = \frac{\partial}{\partial T}(k_B T \ln Z) = -\frac{\partial A}{\partial T}$$

Equation 2-13

Where A is the Helmholtz free energy defined as A = U – TS, where U = <E> is the total energy and S is the entropy, so that

$$A = < E > -TS = -k_B T \ln Z$$

Equation 2-14

2.2 Molecular Dynamics

One of the principle tools in the theoretical study of biological molecules is the method of molecular dynamics simulations (MD). This computational method calculates the time dependent behaviour of a molecular system. MD simulations have provided detailed information on the fluctuations and conformational changes of proteins and nucleic acids. These methods have been used in this work to investigate the structure, dynamics and thermodynamics of nucleic acid systems.

The molecular dynamics simulation method is based on Newton's second law or the equation of motion, F=ma, where F is the force exerted on the particle, m is its mass and a is its acceleration. From a knowledge of the force on each atom, it is possible to determine the acceleration of all atoms in the system. Integration of the equations of motion then yields a trajectory that describes the positions, velocities and accelerations of the particles as they vary with time. From this trajectory, the average values of properties can be determined. The method is deterministic, once the positions and velocities of each atom are known, the state of the system can be predicted at any time in the future or the past. Molecular dynamics simulations can be time consuming and computationally expensive. However, computers are getting faster and cheaper making versions of this method feasible for larger system.

Newton's second equation of motion is given by;

$$F_i = m_i a_i = m\frac{\delta^2 r_i}{\delta t^2}$$

Equation 2-15

Where m_i is the mass of particle i, a_i is its acceleration, $r_i(x_i, y_i, z_i)$ its positions, and F_i is the force acting on particle $_i$; t is the time.

In most cases, the force on each particle depends upon its position relative to other particles. This coupled nature of the particles motions gives rise to a many-body problem which cannot be solved analytically. In these cases the equations of motion must be integrated numerically, usually using a finite difference method.

2.2.1 Integration Algorithms

The potential energy is a function of the atomic positions (3N) of all the atoms in the system. Due to the complicated nature of this function, there is no analytical solution to the equations of motion for more than 2 interacting particles, and so they must be solved numerically in all applications of interest.

Numerous numerical algorithms have been developed for integrating the equations of motion. A selection of the available algorithms are the Verlet algorithm, Leap-frog algorithm, Velocity-Verlet and Beeman's algorithm. In choosing which algorithm to use it is important to consider the following criteria. The algorithm should conserve energy and momentum. It should be computationally efficient. It should permit a long time step for integration and preserve time-reversal symmetry.

2.2.1.1 The Verlet Algorithm

All integration algorithms assume positions, velocities and accelerations that can be approximated by a Taylor series expansion;

$$r(t+\delta t) = r(t) + v(t)\delta t + \frac{1}{2}a(t)\delta t^2 + ...$$
$$v(t+\delta t) = v(t) + a(t)\delta t + \frac{1}{2}b(t)\delta t^2 + ...$$
$$a(t+\delta t) = a(t) + b(t)\delta t + ...$$

Equation 2-16

Where *r* is the position, *v* is the velocity (the first derivative with respect to time), *a* is the acceleration (the second derivative with respect to time), etc.

To derive the Verlet algorithm one can write

$$r(t+dt) = r(t) + v(t)dt + \frac{1}{2}a(t)dt^2 + \frac{1}{6}b(t)dt^3 + O(\mathrm{D}t^4)$$

$$r(t-dt) = r(t) - v(t)dt + \frac{1}{2}a(t)dt^2 - \frac{1}{6}b(t)dt^3 + O(\mathrm{D}t^4)$$

Equation 2-17

Summing these two equations, one obtains

$$\boldsymbol{r(t+dt) = 2r(t) - r(t-dt) + a(t)dt^2 + 0(\mathrm{D}t^4)}$$

Equation 2-18

The Verlet algorithm uses positions and accelerations at time *t* and the positions from time *t*-δt to calculate new positions at time *t*+δt. The Verlet algorithm uses no explicit velocities. The advantages of the Verlet algorithm are, *i*) it is straightforward, and *ii*) the storage requirements are modest[215]. The disadvantage is that the algorithm is of moderate precision.

The problem with the Verlet algorithm is that velocities are not directly generated. While they are not needed for the time evolution, their knowledge is sometimes necessary as they are required to compute the kinetic energy *K*, which is necessary to test the conservation of the total

energy $E=K+V$. This is one of the most important tests to verify that a MD simulation is proceeding correctly. One can compute the velocities from the positions by using;

$$v(t) = \frac{r(t+\Delta t) - r(t-\Delta t)}{2\Delta t}$$

Equation 2-19

the weakness of this method is that the error associated with this expression is significant and of order Δt^2 rather than Δt^4 . However, the velocities are not used in propagating the trajectory, and so this large error in the velocity does not affect the accuracy of the subsequent trajectory. To obtain more accurate velocities, the leapfrog algorithm is used,

$$\boldsymbol{r_i\left(t-\frac{\Delta t}{2}\right) = r_i\left(t-\frac{\Delta t}{2}\right) + r_i(t)\Delta t}$$

Equation 2-20

The velocities at time t can be also computed from

$$r_i(t) = \frac{r_i(t+\frac{\Delta t}{2}) + r_i(t-\frac{\Delta t}{2})}{2}$$

Equation 2-21

This is useful when the kinetic energy is needed at time t, as for example in the case where velocity rescaling must be carried out during equilibration. The atomic positions are then obtained from

$$r_i(t+\Delta t) = r_i(t) + r_i(t+\frac{\Delta t}{2})\Delta t$$

Equation 2-22

The Leap-frog algorithm has advantages in avoiding round-off errors as it involves a difference of two powers of τ rather than adding terms, $O(\tau^{n-1})$ to $O(\tau^n)$ as in other methods which introduce greater error.

To overcome this difficulty, some variants of the Verlet algorithm have been developed. They give rise to exactly the same trajectory, and differ in what variables are stored in memory and at what times. The *leap-frog* algorithm is one of such variants where velocities are handled somewhat better.

2.2.1.2 Finite difference methods

In the context of molecular dynamics simulations finite difference techniques use the known positions, velocities, and other dynamical information at time *t* to obtain the same quantities at a later time, $t + \delta t$. By iterating this procedure the time evolution of a system can be followed generating a molecular dynamics trajectory. During this process the total force on each particle in the system is calculated from the positions using a well defined potential energy function. The forces and masses are used to obtain the acceleration of the particles, which are then taken with the velocities at a time, t to calculate the positions and velocities at a time $t + \delta t$. The new positions are then used to calculate the new forces experienced by the atoms so allowing the velocities and positions at time, $t + 2\delta t$.

The error in the atomic positions is of the order of Δt^4. The velocities are obtained by differentiation with an error in the order of Δt^2

The time step

The time step for a molecular dynamics simulation must be chosen to balance the ability to simulate a real trajectory while sampling sufficient phase space. Choosing too small a time step can mean inefficient sampling of phase space while too large a time step may lead to the system

experiencing high energy overlap of atoms with a resulting instability in the integration algorithm. Unsuitably large time steps mean that the conservation of energy and linear momentum is not maintained and the program will fail due to numerical overflow.

For numerical stability in the integration algorithm the time step is approximately one order of magnitude smaller than the shortest motion. This is a severe restriction, especially as the highest frequency motions in flexible molecules are usually C-H bond vibrations which have very little effect on the overall behaviour of the system. One way to overcome this problem is to 'freeze out' such high-frequency vibrations by constraining bonds using the SHAKE[216] algorithm available in DL_POLY.

2.2.2 Force fields

Theoretical studies of biological molecules permit the study of the relationships between structure, function and dynamics at the atomic level. Considering the number of atoms included in biological systems it is not yet feasible to treat these systems using purely quantum mechanics methods. However, the use of much less computationally demanding empirical potential energy functions makes simulations of these systems viable. The approximations introduced with these methods leads to certain limitations. These are discussed below.

Current potential energy functions provide a reasonably good compromise between accuracy and computational efficiency. They are often calibrated to both experimental results and quantum mechanical calculations of small model compounds. Their ability to reproduce physical properties measurable by experiment is tested; these properties include structural data obtained from x-ray crystallography and NMR, dynamic data obtained from spectroscopy and inelastic neutron scattering and thermodynamic data. The development of parameter sets is a very laborious task, requiring extensive optimisation. This is an area of continuing research and many groups have been working over the past two decades to derive functional forms and

parameters for potential energy functions of general applicability to biological molecules. Among the most commonly used potential energy functions are the AMBER[217], CHARMM[218], GROMACS[219] and OPLS[220]/AMBER[221] force fields.

2.2.2.1 Force field review including CHARMM22, CHARMM27

Since the first force fields were developed there have, in the main, been two strands. The first strand aims to make accurate predictions of molecular structures and properties. This is exemplified by the work of Allinger and co-workers in the development of the Molecular Mechanics force fields. The early members of this family, MM1[222], MM2[223], have been largely superseded by the MM3[224,225,226] and MM4 force fields[227,228]. These have been parameterized against experimental data, specifically heats of formation and vibrational frequencies. To accurately reproduce experimental results a complicated functional form is needed. The MM3 force field has a total of nine different interactions[229], with stretch-bend, bend-torsion-bend, torsion-torsion, torsion-improper torsion, and improper torsion-torsion-improper torsion interactions[230]. In addition the van der Waals is modeled using the Hill potential.

The second strand of force field development is aimed at modeling large molecules such as protein or polymers. Typical force fields in this category are AMBER[231,232], and CHARMM[233]. By necessity these have simpler functional forms, usually only containing the terms described in § 2.2.2.2 with an option to employ the united atom approximation. The interaction potentials for bond stretching and bending tend to be restricted to the harmonic approximation, while van der Waals interactions are modeled using the simpler Lennard-Jones potential. Despite their simplicity force fields such as these can often give good results for condensed phase properties, including thermodynamic properties such as heats of vaporization or densities.

During this research the CHARMM force-field has been employed initially in its earlier form of CHARMM22 and towards the end as CHARMM27. This has allowed comparisons (§ 4.5.2) to

be considered inline with the known differences from the literature which are summarized in the following section.

CHARMM force-field

The first simulation on biological macromolecules in the early 1970's based on empirical force fields were performed in vacuum due to limitations in computer resources. The original nucleic acid force field associated with the program CHARMM,[234,235] called EF2, was designed to be used in vacuo. Since the aqueous environment is an essential part of the structure of biomolecules damping of electrostatics was included implicitly using a distance-dependent dielectric constant. The force field included both extended atom and all-atom models. In the extended atom model the nonpolar hydrogens are represented as being part of the "extended " nonhydrogen atoms to which they are covalently bound while polar hydrogens are represented explicitly. The optimization of EF2 concentrated primarily on the internal terms in the potential energy function, based on geometries, virbrational spectra, and conformational energetics. Nonbonded terms, including an explicit term for hydrogen bonds, were based on analogy with previously available parameters, and were tested via crystal simulations and dimer interaction energies. Application of the force field to small DNA duplexes showed them to be stable in MD simulations, though the stability was strongly dependent on the treatment of the dielectric constant.[236]

The next CHARMM force fields were designed to include an explicit representation of the solvent,[237] based on the TIP3P water model.[238] This CHARMM22 nucleic acid force field was an all-atom representation and put emphasis on optimization of both the internal and nonbonded terms. The previously explicit term for hydrogen bonding, present in the EF2 version, was removed and included within the electrostatic and Lennard-Jones terms. CHARMM22 yielded accurate structural and thermodynamic properties by properly balancing the interaction triad:

the solute-solute, solute-solvent, and solvent-solvent nonbonded interactions. However, it was shown that MD simulations of DNA duplexes in solution would assume the A conformation[239] in high water activity[240,241]conditions where the B form should be stable.

In the next generation of CHARMM, CHARMM27 was a reoptimisation and referred to as the CHARMM27 force field for nucleic acids.[242] In this force field special emphasis was placed on the non-bonded and dihedral parameters. Nonbonded parameters were again optimized to be compatible with the modified TIP3P water potential and were tested with new parameters for sodium. [243] QM minimum interaction energies, geometries and experimental crystal data was reproduced with rigorous optimisation of the nonbonded parameters, particularly charges. For optimisation of the dihedral parameters new target data included extensive results from QM calculations. These calculations included conformational energetics about the rotatable bonds of model compounds designed to mimic the phosphodiester backbone, the sugar, and the glycosyl linkage.[244,245] This information allowed for better optimisation of the dihedral parameters associated with those degrees of freedom. In addition experimental data was used in the dihedral optimisation parameter process. As a result of all of these changes CHARMM27 is a force field that is able to treat DNA in multiple environments including the equilibrium between A and B forms as a function of water activity.[246]

2.2.2.2 The CHARMM potential energy function

The energy, *E*, is a function of the atomic positions, *r*, of all the atoms in the system, these are usually expressed in terms of Cartesian coordinates.

$$V(r) = E_{bonded} + E_{non-bonded}$$

Equation 2-23

E_{bonded}, in turn, is a sum of three terms:

$$E_{bonded} = E_{bond\ -stretc\ h} + E_{angle\ -bend} + E_{rotate\ -around\ -bond}$$

Equation 2-24

which corresponds to three types of atom movement.

The value of the energy is calculated as a sum of internal, or bonded, terms E_{bonded}, which describe the bonds, angles and bond rotations in a molecule, and a sum of external or nonbonded terms, $E_{\text{non-bonded}}$, These terms account for interactions between nonbonded atoms or atoms separated by 3 or more covalent bonds.

The first term in the above equation is a harmonic potential representing the interaction between atomic pairs where atoms are connected by a covalent bond, and is often denoted as a 1–2 interaction. This describes the strain energy induced when the bond length is stretched or contracted away from the ideal bond length, b_0. The force constant, k_b, determines the strength of the bond and is the dominant term in determining the frequency of the bonds stretch. Both ideal bond lengths b_0 and force constants k_b are considered to be specific to the pair of atom-types involved in the bond, *i.e.* depend entirely on chemical "type" of atoms that form the bond.

$$E_{bond\ -stretc\ h} = \sum_{1,2\ pairs} k_b (b - b_0)^2$$

Equation 2-25

Values of force constant are often evaluated from experimental data such as infrared stretching frequencies or from quantum mechanical calculations, though it should be noted that these frequencies are slightly modified by other interactions present within the molecule. Values for bond lengths can be inferred from high resolution crystal structures or microwave spectroscopy data; crystal packing effects on bond lengths are usually negligible.

The second term in the equation is the energy associated with the angle formed between two consecutive bonds, the 1-3 bond angles, and is associated with the alteration of bond angles

theta from ideal values θ_0, which is also represented by a harmonic potential. Values of θ_0 and k_θ depend on chemical type of atoms constituting the angle.

These two terms together describe the deviation from an ideal geometry and are effectively penalty functions that in a perfectly optimised structure would be close to zero. In most structures ideal values are not necessarily consistent within a given molecule, and so there can be considerable strain energy locked into the molecular geometry (e.g. cyclobutane).

$$E_{bond-bend} = \sum_{angles} K_\theta (\theta - \theta_0)^2$$

Equation 2-26

The third term represents the torsion angle potential function which models the steric and electrostatic effects between atoms separated by 3 covalent bonds (1,4 pairs). The motion associated with this term is a rotation, described by a dihedral angle and coefficient of symmetry *n=1,2,3*), around the middle bond. This potential is periodic and is often expressed as a Fourier series, by just a single term within this series. The net torsional potential could also include contributions from vdW and electrostatics as explained later.

In addition to these terms, the CHARMM force field has two additional terms; one is the Urey-Bradley term, which is an interaction based on the distance between atoms separated by two bonds (1,3 interaction), and introduces a degree of anharmonicity into the bond angle potential. The second additional term is the improper dihedral term used to maintain chirality or planarity (depending on the ideal angle φ adopted).

$$E_{rotate\ -around\ -bond} = \sum_{1,4\ pairs} K_\emptyset (1 - \cos(n\emptyset))$$

Equation 2-27

The parameters for the terms, K_b, K_θ, K_ϕ, are obtained from studies of small model compounds and comparisons to the geometry and vibrational spectra in the gas phase (IR and Raman spectroscopy), supplemented with ab initio quantum calculations.

The energy term representing the contribution of non-bonded interactions in the CHARMM potential function has two components, the van der Waals interaction energy and the electrostatic interaction energy. Some other potential functions also include an additional term to account for hydrogen bonds. In the CHARMM potential energy function, these interactions are accounted for by the electrostatic and van der Waals interactions.

$$E_{non-bonded} = E_{van-der-Waals} + E_{electrostatic}$$

Equation 2-28

The Van der Waals interaction between two atoms arises from a balance between repulsive and attractive forces. The repulsive force arises at short distances where the electron-electron interaction is strong. The attractive force, also referred to as the dispersion force, arises from fluctuations in the charge distribution in the electron clouds. The fluctuation in the electron distribution on one atom or molecule gives rise to an instantaneous dipole which, in turn induces a dipole in the second atom or molecule giving rise to an attractive interaction. Each of these two effects is equal to zero at infinite atomic separation r and become significant as the distance decreases. The attractive interaction is longer range than the repulsion but as the distance becomes short, the repulsive interaction becomes dominant. This gives rise to a minimum in the energy. Positioning of the atoms at the optimal distances stabilizes the system. Both values of energy at the minimum E^* and the optimal separation of atoms r^* (which is roughly equal to the sum of van der Waals radii of the atoms) depend on chemical type of these atoms.

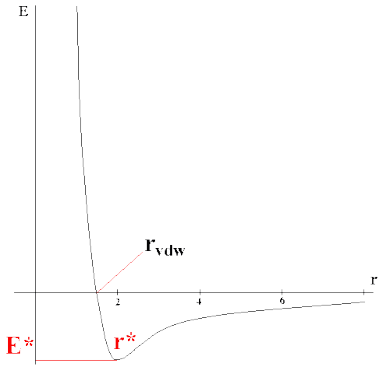


Figure 2-1 The Lennard-Jones 6-12 potential used to model the van der Waals interactions.

CHARMM models these VDW interactions using the Lennard-Jones 6-12 potential which expresses the interaction energy using the atom-type dependent constants *A* and *C*. Values of *A* and *C* may be determined by a variety of methods, like non-bonding distances in crystals and gas-phase scattering measurements

$$E_{van-der-Waals} = \sum_{\substack{nonbonded \\ pairs}} \left(\frac{A_{ik}}{r_{ik}^{12}} - \frac{C_{ik}}{r_{ik}^{6}} \right)$$

Equation 2-29

The van der Waals interactions are one of the most important for the stability of the biological macromolecules.

The fundamental expression in electrostatics, the interactions of stationary electric charges, is the Coulomb potential energy of one charge of magnitude q_i at a distance *r* from another charge q_k :

$$E_{electrostatic} = \sum_{nonbonded \quad pairs} \frac{q_i q_k}{4\pi\varepsilon r}$$

Equation 2-30

q_i and q_k are the charges belonging to atoms a_j and a_k for which the charge is defined by the oxidation state. ε is the effective dielectric constant for the medium between the two charges.

The empirical potential energy function is differentiable with respect to the atomic coordinates; this gives the value and the direction of the force acting on an atom and thus it can be used in a molecular dynamics simulation. The empirical potential function has several limitations, which result in inaccuracies in the calculated potential energy.

One limitation is due to the fixed set of atom types employed when determining the parameters for the force field. Atom types are used to define an atom in a particular bonding situation, for example an aliphatic carbon atom in a sp^3 bonding situation has different properties from a carbon atom found in a Histidine ring. Instead of presenting each atom in the molecule as a unique one described by a unique set of parameters, there is a certain amount of grouping in order to minimise the number of atom types. This can lead to type-specific errors. The properties of certain atoms, like aliphatic carbon or hydrogen atoms, are less sensitive to their surroundings and a single set of parameters may work quite well, while other atoms like oxygen and nitrogen are much more influenced by their neighbouring atoms. These atoms require more types and parameters to account for the different bonding environments.

An approximation introduced to decrease the computational demand is the pair wise additive approximation. This approximation is implicit in the CHARMM force-field with the Lennard-Jones potential that is by definition a pair potential, i.e., interaction energy between one atom and the rest of the system is calculated as a sum of pair-wise (one atom to one atom) interactions, or as if the pair of atoms do not see the other atoms in the system. The CHARMM force-field does not include polarisation so the simultaneous interaction between three or more atoms is not calculated, as a result certain polarisation effects are not explicitly included in the

simulation. This can lead to subtle differences between calculated and experimental results, for example, in the calculation of experimentally observable pK shifts[247] of ionisable amino acid residue side chains induced by the electrostatic field of the whole protein.

Another important point to take into consideration is that this is a potential energy function and not a free energy function. Entropy is included by sampling the energy surface and so is not readily calculated at any point without a full ensemble-sampling. Thus, a minimum value of *E* calculated as a sum of the potential functions at finite temperature does not necessarily correspond to the equilibrium, or the most probable state; this corresponds to the minimum of free energy. Because of the fact that experiments are generally carried out under isothermal-isobaric conditions (constant pressure, constant system size and constant temperature) the equilibrium state corresponds to the minimum of Gibb's Free Energy, *G*. While just an energy calculation ignores entropic effects, these are included implicitly in the sampling provided by a molecular dynamics simulation.

2.2.2.3 Treatment of the nonbonded energy terms

These are numerical issues which relate to the implementation of the force field in molecular simulations, and not to the development or reliability of the force-field itself. The most time consuming part of a molecular dynamics simulation is the calculation of the nonbonded terms in the potential energy function, e.g., the electrostatic and van der Waals forces. Neutral atoms and molecules within the model are subject to these two distinct forces, an attractive van der Waals force, or dispersion force, at long ranges, and a repulsion force, the result of overlapping electron orbitals, called Pauli repulsion. The Lennard-Jones potential, Figure 2-1 **The Lennard-Jones 6-12 potential used to model the van der Waals interactions.**, represents this behaviour[248].

The Lennard-Jones potential is of the form

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

Equation 2-30

Where ε is the well depth and *A* and *B* are the atom hard sphere diameter.

In principle, the non-bonded energy terms between every pair of atoms should be evaluated; in this case, the number increases as the square of the number of atoms for a pair wise model (N^2). To speed up the computation, the interactions between two atoms separated by a distance greater than a pre-defined distance, the cutoff distance, are ignored.

2.2.2.4 Long range electrostatic interactions

A number of experimental studies have demonstrated the importance of long range electrostatic interactions in biological molecules. For example, Fersht and co-workers have studied long range protein – H^+ interactions by detailing the relationship between surface charges of the

serine protease subtilisin and the pKa of the active site residue His-64[249,250]. Other experiments have shown that protein – Ca^{2+} affinity can be modulated by electrostatic interactions occurring over long distances[251]

Inclusion of the longer range electrostatics in a molecular dynamics simulation by simply increasing the cutoff distance can dramatically raise the computational cost. Most often, the long-range electrostatic interactions are ignored, however, in some cases, their neglect introduces a severe approximation; for example in the calculations of dielectric properties or in metal ion – protein interactions[252].

In recent years, a number of models have been introduced which permit the inclusion of long-range electrostatic interactions in molecular dynamics simulation. For simulations of proteins and enzymes in a crystalline state, the Ewald summation is considered to be the correct treatment for long range electrostatic interactions[253]. Variations of the Ewald method for periodic systems include the particle-mesh Ewald method[254]

2.2.3 Treatment of Solvent in a Molecular Dynamics Simulation

Solvent, usually water, has a fundamental influence on the structure, dynamics and thermodynamics of biological molecules, both locally and globally. One of the most important effects of the solvent is the screening of electrostatic interactions. The electrostatic interaction between two charges is given by Coulomb's law,

$$V_{elec} = \frac{q_i q_j}{\varepsilon_{eff} r_{ij}}$$

Equation 2-31

where q_i, q_j are the partial atomic charges, ε_{eff} is the effective dielectric constant and r_{ij} is the relative distance between the two particles. It is important to include solvent effects in an MD simulation. This can be done at several levels. The simplest treatment is to include a dielectric screening constant in the electrostatic term of the potential energy function. In this implicit

treatment of solvent, water molecules are not included in the simulation but an effective dielectric constant is used. Often the effective dielectric constant is taken to be distance dependent, $\varepsilon_{eff} = r_{ij}\varepsilon$, where ε ranges from 4 to 20. Although this is a crude approximation, it is still much better than using unscreened partial charges. Other implicit solvent models have been developed that range from the relatively simple distance-dependent dielectric constants to models that base the screening on the solvent exposed surface area of the biological molecule. The distance-dependent dielectric coefficient attempted to avoid the need to calculate square roots for a cheaper calculation in the early days of computing. This method still introduced inaccuracies into the simulations and resulted in divergent charge-charge interaction energies.

If water molecules are explicitly included in the simulation, then $\varepsilon=1$ in the energy function; the explicit water molecules provide the electrostatic shielding. In this more detailed treatment of the solvent boundary conditions must be imposed, first, to prevent the water molecules from diffusing away from the solute during the simulation, and second to allow simulation and calculation of macroscopic properties using a limited number of solvent molecules. It is also necessary to note that in such a system the electrostatic energy arises as the difference of two infinite parts, and the interactions furthest away are the ones that are most important, hence the need for EWALD or reaction field methods.

2.2.3.1 The TIP3P model

Throughout this study the rigid TIP3P[255] model of water has been used for which CHARMM has been developed to include in calculations. Within this model the bond distances are fixed at 0.957Å for the O-H bond distances and 104.5^{o} for the H-O-H angle[256]. Each hydrogen atom is assigned a charge of +0.417*e* and a charge of -0.834*e* is placed at the site of the electronegative oxygen. A Lennard-Jones site centred on the oxygen atom has a potential of zero when the interatomic distance, σ, has a value of 1.76Å and a well depth, ε, of -0.1521 kcalmol^{-1}. The

water molecule was kept rigid using the SHAKE algorithm, implemented in DL_POLY with a tolerance of 0.0001.

TIP3P water does not include an explicit hydrogen bonding term but enables hydrogen bonding by modelling hydrogen atoms with no defined excluded volume.

2.2.4 Periodic Boundary Conditions

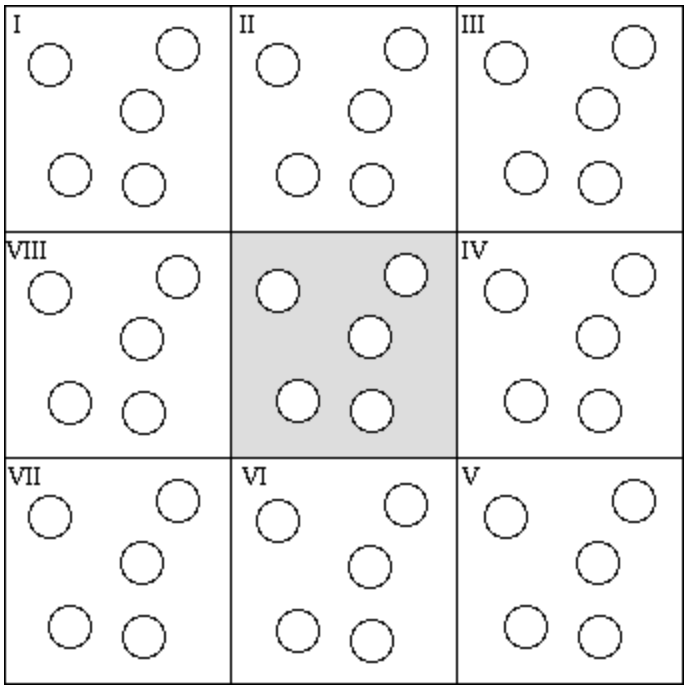


Figure 2-2 Periodic boundary conditions

Periodic boundary conditions enable a simulation to be performed using a relatively small number of particles in such a way that the particles experience forces as though they were in a bulk solution. Figure 2-2, above shows a central box surrounded by eight neighbours. The coordinates of the image particles, those found in the surrounding box are related to those in the primary box by simple translations. The simplest box is the cubic box. Forces on the primary particles are calculated from particles within the same box as well as in the image box. The cut-off is chosen such that a particle in the primary box does not see its image in the surrounding boxes.

2.2.5 The model

The energy function can be written as:

$$U(\vec{R}) = \sum_{Bonds} K(b-b_0)^2 + \sum_{Angles} K_\theta(\theta-\theta_0)^2 + \sum_{UB} K_{UB}(S-S_0)^2 + \sum_{Dihedrals} K_\chi(1+\cos(n\chi-\delta)) +$$

$$\sum_{\text{Impropers}} K_{imp}(\varphi-\varphi_0)^2 + \sum_{Nonbond} \in \left(\left[\left(\frac{R_{\min,ij}}{r_{ij}} \right)^{12} - 2\left(\frac{R_{\min,ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi e r_{ij}} \right)$$

Where K_b, K_θ, K_{UB}, $K\chi$, and K_{imp} are the bond, bond angle, Urey-Bradley, dihedral angle, and improper dihedral angle force constants respectively; *b, θ, S, χ,* and *φ* are the bond length, bond angle, Urey-Bradley 1,3 distance, dihedral torsion angle, and improper dihedral angle respectively, with the subscript zero representing the equilibrium values for the individual terms. Coulomb and Lennard-Jones 6-12 terms contribute to the non-bonded interactions; ϵ is the Lennard-Jones well-depth, $R_{\min}$ is the distance at the Lennard-Jones minimum, q_i is the partial atomic charge, and *e* is the dielectric constant.

2.2.6 Atom types and parameters

The energy function includes bonding terms for bond stretching, angle bending, torsional rotation, and improper (out-of-plane) motion. Non-bonded interactions are represented by a Lennard-Jones 6-12 term for the vdW repulsion and dispersion interaction and a Coulomb term for the charge-charge interactions, which includes a Coulomb term arising from the interaction of partial point charges on each interaction site.

Intramolecular and nonbonded terms for the organic compound were taken from CHARMm. Both CHARMm22 and CHARMm27 were used during this research to get an idea of the force field dependence. The [MeN_i] unit was modelled as a rigid body, with CHARMm force-field terms for the N atoms (typed according to the whole ligand); the Me component of the cylinder was described purely by a charge of +2*e* (or +1*e*, depending on the system being modelled),

with no non-bonded attraction and the use of the rigid [MeN_i] unit removing the need to model the metal ion repulsion. Radial distribution functions were monitored and confirmed that, even with the more open tetrahedral metal centres, the steric bulk of the ligands was sufficient to avoid any close non-bonded contacts with the metal.

Using the ESP method, the charge distributions were initially calculated on isolated ligand molecules. The partial charges were then chosen to reproduce the electric field from the calculated charge distribution, as experienced by a probe ion located on each point in a grid around the molecule. Having calculated the charges and defined the rigid body the van der Waals parameters were taken from CHARMm (22 or 27) for Nucleic acids and proteins for which they had been fitted to electronic structure calculations, enthalpies of sublimation and spectroscopic data[233,234,235].

The bond and angle terms are taken from CHARMm and represented by simple (diagonal) harmonic expression. CHARMm used r_{eq}, θ_{eq}, K_r and K_θ as starting values and adjusted to reproduce experimental normal mode frequencies. Dihedral angles are represented by a simple set of parameters, often specified only by the two central atoms. Van der Waal interactions are only calculated between atoms in different molecules or for those atoms in the same molecule separated by at least three bonds. The electrostatic model employed in this force field relies upon the use of 6-31G* derived electrostatic potentials, multiple molecules, multiple conformations and the ESP fitting approach.

Within this research the ligand cylinders were modelled using standard CHARMm parameters. In order to maintain the structure of each cylinder design, containing two positive metal centres in close proximity, a rigid metal-coordination shell was defined in both the parent cylinders and derivative cylinders. The geometry and charge did depend on the metal centre (Cu^{+} or Fe^{2+}),

but was otherwise the same for all cylinders modelled with each metal ion. Designing the simulations to include models of the various cylinders in this way has maintained a consistency in method with work already carried out in the group[214] and therefore focussed attention on differences between CHARMm 22 and CHARMm 27, and on the effect of cylinder geometry, rather than on the details of the metal-organic force-field development. The use of a rigid metal/coordination shell complex obviated the need to refine repulsive metal/organic interactions, or indeed angle terms including the metal. Further, since the ligands were all chemically rigid in the vicinity of the metal, no torsional terms involving the metal were needed. The metal centre was therefore modelled as a point charge, with magnitude corresponding to the oxidation state; dispersion interactions arising from this cation were ignored. The metal complex force-field was tested in MD simulations of the crystal structure (where these were known) and shown to reproduce the known crystal structure well.

Parameters for all Iron based cylinders were the same as those used in the earlier studies[214]. No such studies existed for the Copper based cylinders, and so these parameters were developed as part of this project. Electrostatic potential charges (ESP) were derived for the Copper cylinder using Density Functional Theory (DFT) calculations, with the B3LYP method and 6-31G* basis set. Again simulation using the refined parameters showed good reproduction of the crystal structure made available by the experimentalists[203].

3 Analysis Techniques

During the production of extended molecular dynamics (MD) simulations, large amounts of data are accumulated as the trajectory evolves. A variety of analytical techniques are available to enable variables to be investigated from the simulated DNA systems. This chapter will describe the methods used in subsequent chapters to extract and display properties of the various DNA reference systems and DNA with bound cylinder.

3.1 Conformational Analysis

3.1.1 MD Toolchest

During the MD simulations undertaken in this study large data files describing the time evolution of the system have been accumulated. A set of analysis programs, collectively called "Molecular Dynamics Analysis Toolchest" has enabled the various simulations to be systematically analysed. This package has been extensively used to undertake time series analysis of the conformational and helicoidal parameters within the DNA systems. The main focus has been on the modules called "Curves, dials and windows".[257 259]

3.1.1.1 Backbone Conformational Parameters

Dials and Windows uses the standard IUPAC defined parameters to describe the backbone conformation of DNA[259]. This phase of the analysis indicates conformational transitions in the backbone and residue torsion angles (Figure 3-1) as well as fluctuations in the variables describing helix morphology during the course of the simulation. The complete analysis allows a full characterisation of the dynamical fine structure of the backbone.

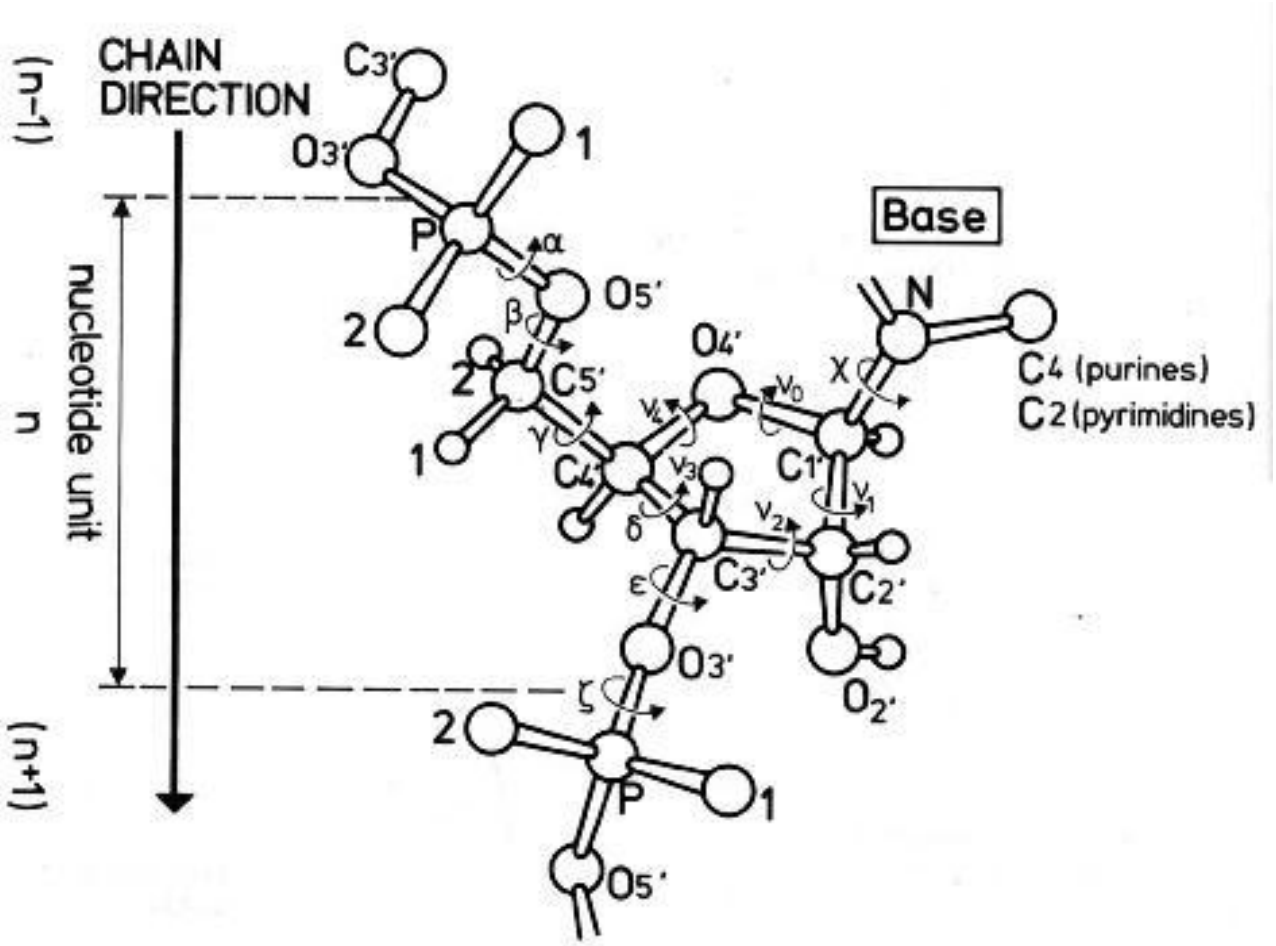


Figure 3-1 Backbone torsion angles in nucleic acid structures. (Diagram copied from Saenger W. 1984. Principles of Nucleic Acid Structure. Springer, New York.)

The backbone conformation is calculated using the Dials and Windows package which takes eight backbone conformational angles and the amplitude of the sugar pucker into account

generating an output in the form of plots called Dials as seen in Figure 3-2.

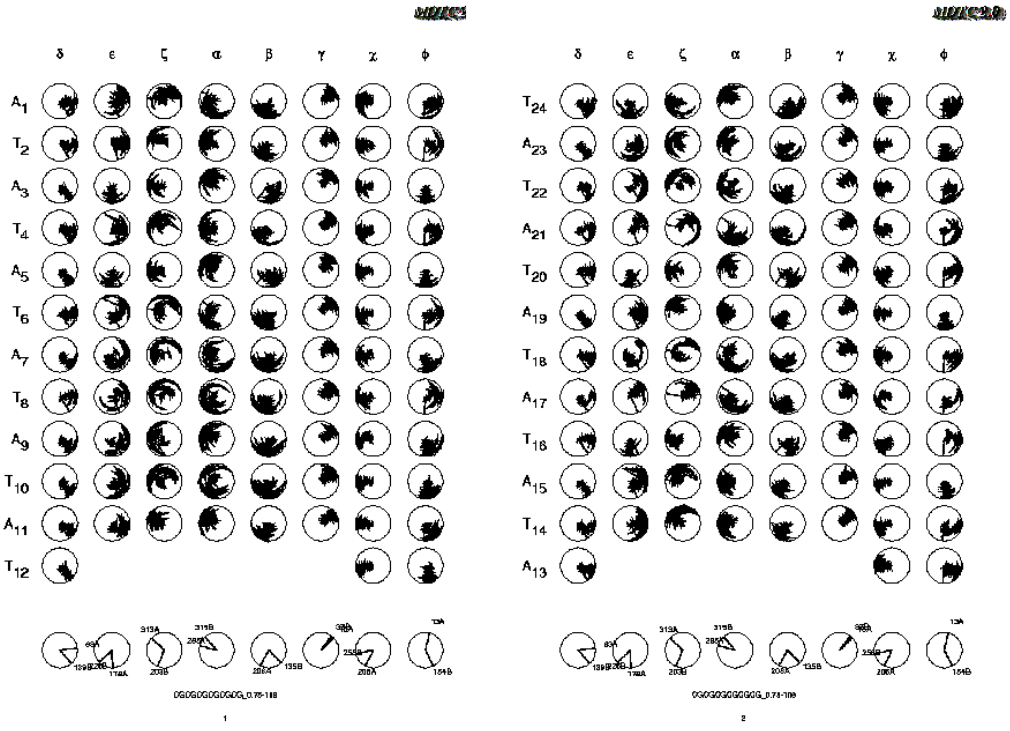


Figure 3-2 Output from Dials and Windows showing the conformational parameters in the form of dials. The expected values for A and B-DNA forms are indicated in the plots at the bottom of the figure.

3.1.1.2 Helicoidal Parameters

There are four categories of helicoidal parameters within nucleic acids. These are inter-base pair parameters, axis-base pair parameters, axis parameters and intra-base pair parameters.

Axis-base pair parameters are X-displacement(XDP), Y-displacement(YDP), inclination(INC) and tip(TIP), Figure 3-3. The first two variables are translational and the last two are rotational.

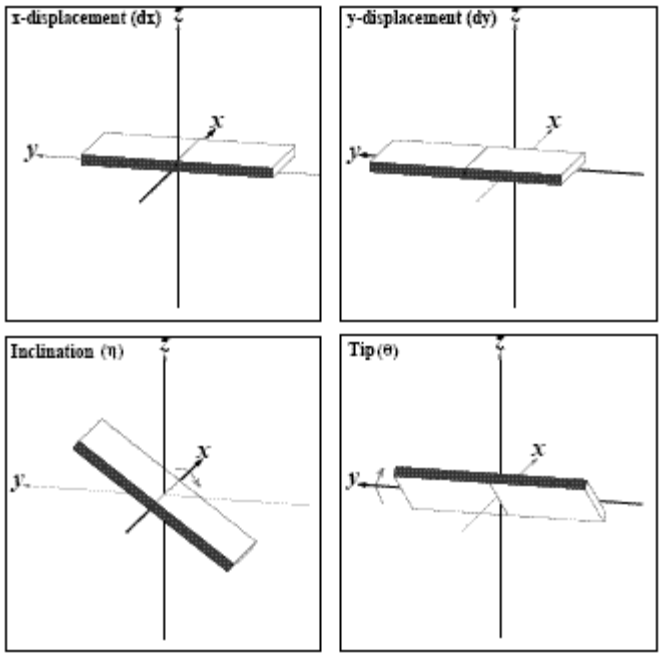


Figure 3-3 Pictorial definitions of parameters that relate base pairs to its helical frame.

Intra-base parameters are, Shear(SHR), Stretch(STR), Stagger(STG), Buckle(BKL), Propeller(PRP) and Open(OPN), see Figure 3-4. The first three variables are translational and the last three are rotational.

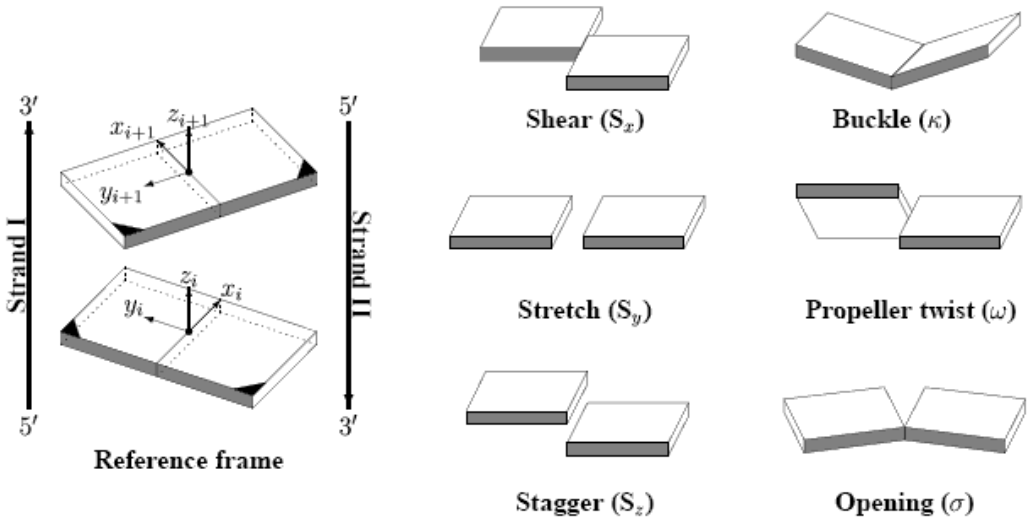


Figure 3-4 Pictorial definitions of parameters that relate complementary base pairs. The base-pair reference frame is constructed such that the x-axis points away from the (shaded) minor groove edge. Images illustrate positive values of the designated parameters.[258]

Inter-base pair parameters are, Shift(SHF), Slide(SLD). Rise(RIS), Tilt(TLT), Roll(ROL) and Twist(TWS), see Figure 3-5. The first three variables are translational and the last three are rotational.

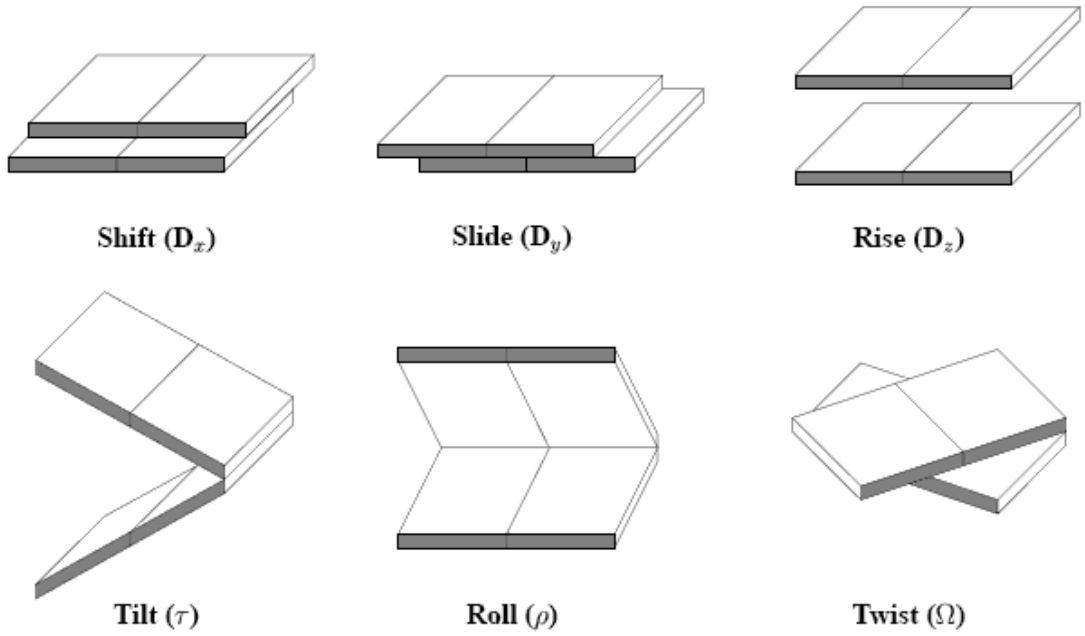


Figure 3-5 Pictorial definitions of parameters that relate sequential base-pair steps. Images represent positive values of the designated parameters. [258]

The parameters displayed pictorially in Figure 3-3, Figure 3-4 and Figure 3-5 are all calculated with the Dials and Windows package. Plots displaying the time dependence of the helicoidal parameters are constructed as explained in Figure 3-7. An example of the data output in the form of windows plots is shown in Figure 3-6.

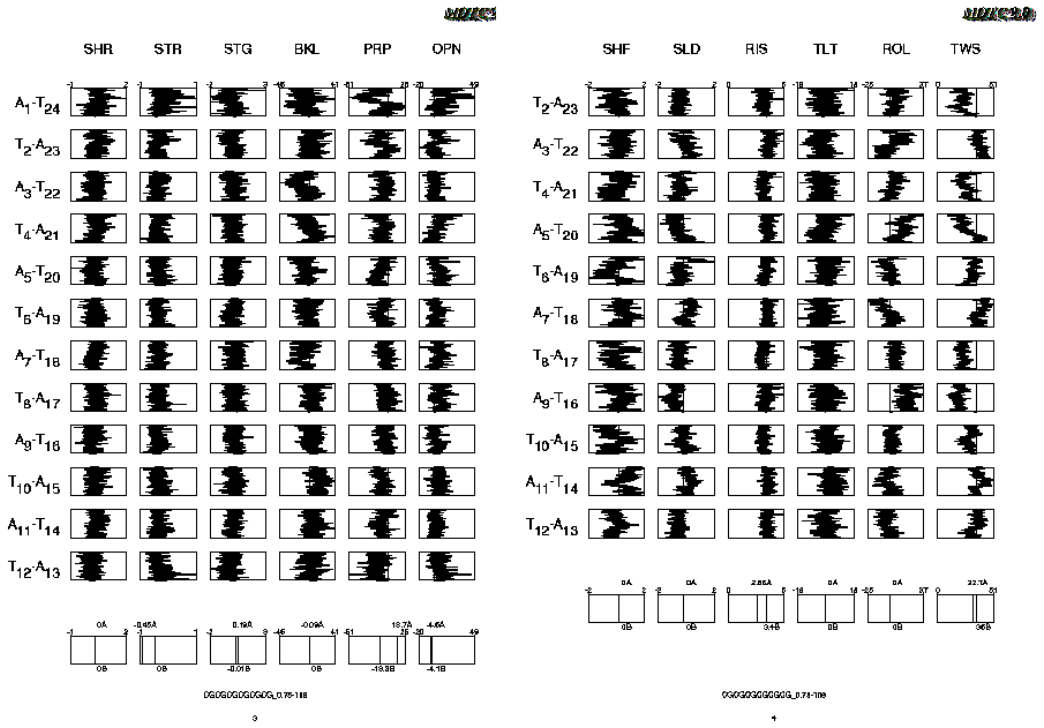


Figure 3-6 Output from Dials and Windows showing the helicoidal parameters in the form of windows. The expected values for A and B-DNA forms are indicated in the plots at the bottom of the figure.

3.1.1.3 Display Format

"Dials and Windows" plots conformational parameters so that each parameter is represented in the form of a dial. An example is shown in Figure 3-2. When displayed in this way the angle of 0^o in the dial is north, 90^o is east, 180^o is south and 270^o is west. The time dependence of the conformational parameters are then presented in circular polar coordinates, (r, ϑ) with time being indicated along the radius, r, and the conformational parameter by the angle, ϑ. A weakness of this method of presenting the output is that points are connected by straight lines rather than by arcs, which disrupts the appearance when large fluctuations occur. The lines generated by this method show the movement of a coordinate as a function of time. One inherent problem with this display format is that the length of lines for earlier time points are smaller than those at the end so artificially causing events at the end of a simulation to dominate the plot.

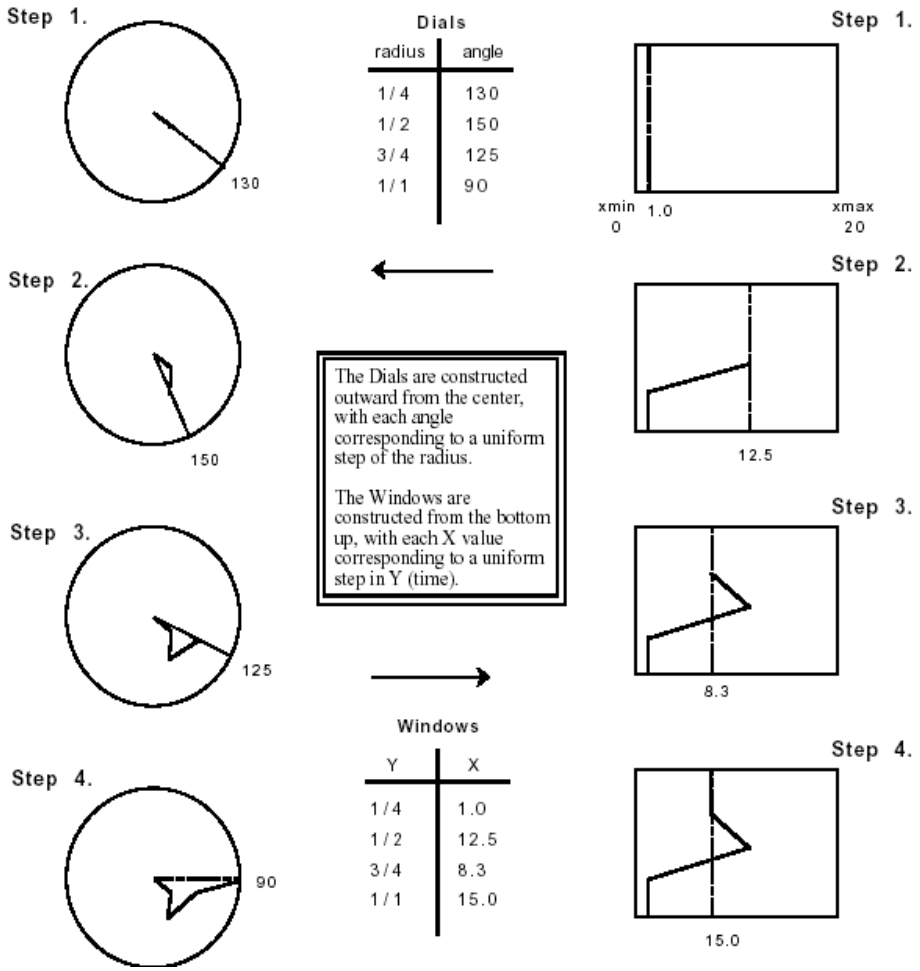


Figure 3-7 An explanation of the dials and windows display format (diagram adapted from the MD toolchest user manual) [259]

Helicoidal parameters can also be displayed in windows where the helicoidal parameter is depicted on the y-axis and time on the x-axis. In the case of a window the method of construction is the same as for the dial but with a linear y-axis removing the issue seen with the dials and making comparison throughout the simulation simpler to visually observe in the output.

3.1.2 3 DNA

As several reference frames are available on which DNA analysis can be based comparison studies,[260] have been undertaken. The conclusion was that the choice of reference frame does have an effect on the calculated parameters. There are two popular reference frames generally

accepted by those analysing DNA. For these methods the reference frame can be defined in two ways. One method is to define the purine C8 atom and the pyrimidine C6 atom and the mean base-pair plane; the other uses the average of two constituent base reference frames in a base-pair.

3DNA utilises the base-centred reference frame, as recommended during an international workshop[261] in 1999, a local helical axis is defined using four variables for each base within the structure, X displacement, Y displacement, Inclination, and Tip. This algorithm allows for exact reconstruction of the structure from a set of parameters which can, in turn, be deduced from the structure. These parameters also define simple geometrical values such as the distance between base-pair centres, $\sqrt{Shift^2 + Slide^2 + Rise^2}$, the bending angle between two base-pairs, $\sqrt{Roll^2 + Tilt^2}$ and the angle between the base-pair normal and local helical axis is $\sqrt{Inclination^2 + Tip^2}$. From the parameters calculated 3DNA can be used to automatically classify a dinucleotide step as A-, B-, or TA-like[262]. This will be employed to observe potential conformational transitions in the ligand-bound DNA during this work.

3.1.2.1 Average Structure

To calculate the average structure of the DNA during a period of the simulation translational and rotational motions are removed from all coordinates of the structure by least squares fitting to a reference structure (the initial structure when t=0)[263]. The average coordinates are then calculated using the following equation for the x coordinates.

$$average(x) = \frac{\sum_{i=1}^{n_{config}} x_i}{n_{config}}$$

Equation 3-1

The average *y* and *z* coordinates are calculated using the same formula and replacing x_i with y_i or z_i respectively.

The flexible nature of DNA with its many vibrations makes it difficult to remove the noise from the signal when considering its motion. For this reason the averaging method is important to enable an average behaviour to be extracted. The advantage of working with an average structure in this way is that it generates a well defined geometry independent of fluctuations. However, the disadvantage is that a geometry calculated to be the average might not ever be realised in practise depending on the sample size or if there were to be a conformational change between two states in the system resulting in the average being neither of them.

To ensure that a true representation of the real system has been made the distribution of values can be observed as calculated within dials and windows, and presented later in chapters 4 and 5.

3.2 Structural Properties

3.2.1 Radial Distribution Functions

The radial distribution function (RDF) shows how the atoms in a system are radially packed around each other and is an example of a pair correlation function. In systems such as liquids where there is continual movement of the atoms, the RDF describes the average structure. The availability of the RDF from X-ray or neutron diffraction studies provides results that enable a direct comparison between experiment and simulation. To construct an RDF, concentric spheres, set at a fixed distance (Δr), are drawn around the central atom. Using the regular intervals of saved data from the simulations trajectory records as snapshots of the system, the number of atoms in each shell is determined. By considering an average, for the whole trajectory, the population of each shell is calculated before dividing by the volume of each shell and the average density of atoms in the system, resulting in a value for the RDF.

$$g(r) = n(\mathrm{r}) / (\rho\, 4\pi r^2 \Delta r)$$

Where $g(r)$ is the RDF, $n(r)$ is the mean number of atoms in a shell of width Δr at distance r, and ρ is the bulk atom number density.

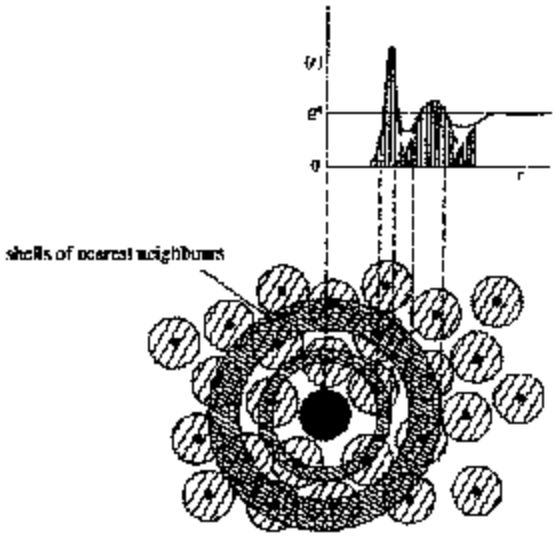


Figure 3-8 Schematic radial distribution around solute molecule.

3.3 Time Correlation Functions

Statistical mechanics makes it possible to calculate the bulk properties of a system, from MD simulations, by using ensemble averaging. Other important properties can be observed by using correlation functions to see how the system is structured, or how it changes with time. For these reasons correlation functions are a powerful means of analysing time dependent data and correlations in space. Both will be used in analysis of data obtained from the study of solvent dynamics. The correlation function reveals the time scale for changes in the atomic motion by correlating the velocity of an atom at a given time, with its velocity at a time later showing the effect of the inter-atomic forces on the atoms motion.

This is possible because a molecular dynamics simulation provides data values at specific times so allowing the time dependent properties within the simulation to be calculated. Time correlation functions have the general form,

$$C_{ab}(t) =< x(t).y(0) >$$

Equation 3-2

And allow for the value of a property at some instant to be correlated with the value of the same, or another property, at a later time *t*. If the properties *x* and *y* are different then the time correlation function is referred to as a cross correlation function, whereas if *x* and *y* are the same property then the time correlation function is referred to as an auto correlation function.

3.3.1 Mean Square Displacement

The fluid property of a liquid is due to the constant motion of molecules which are not constrained in a rigid lattice. This constant motion to diffusion of the constituent molecules within a liquid makes the motion of an individual molecule in a dense fluid complicated. As it travels, the molecule collides with other molecules, and hence changes both its direction and its velocity. This resulting motion describes a random walk and can be analyzed by considering the mean square of the distance traveled by a particle. For diffusion in three dimensions this mean square displacement is proportional to the time elapsed:

$$< r^2 >= 6Dt + C$$

Equation 3-3

Where $<r^2>$ is the mean square displacement and *t* is time. *D* is the diffusion coefficient and *C* is a constant.

When considering the motion of a molecule in one dimension along a straight line it may move either forwards or backwards, both of which are equally likely. When forward motion is defined as positive and backwards as negative, the distance traveled by a molecule over a time

period may be determined by simply adding together the forward and backwards motion. As both forward and backward motion is equally probable the distance traveled sums to zero on average. However, if the squares of each motion both forward and backward are summed all quantities being added will be positive and the square distance will grow linearly with time.

In a real molecular system where a molecule may move in three dimensions the same principle applies. The mean square displacement (MSD) is a result of the average square distance taken over all of the molecules. As a measurable quantity related to diffusion the MSD relates directly to the underlying motion of the molecules.

3.3.2 RMSD

One of the most common mathematical descriptions of a molecules motion within a trajectory is the root mean squared displacement (RMSD) of the atoms in the system with respect to a reference structure. For a stable system, the RMSD would be expected to fluctuate about a constant value. The translational and rotational motions were first removed by least-squares fitting to a reference structure. The RMSD was then calculated from *Equation 3-3*:

$$RMSD_{(i.j)} = \frac{\sum_{k=1}^{N} \sqrt{\left(\left(x_{ik} - x_{jk}\right)^2 + \left(y_{ik} - y_{jk}\right)^2 + \left(z_{ik} - z_{jk}\right)^2\right)}}{N}$$

Equation 3-4

Where N is the total number of atoms selected, x, y, and z are sets of configurations for each molecule in the analysis.

4 Results

4.1 Introduction

In this section the results of molecular dynamics simulations of the interaction between DNA sequences and metallosupramolecular cylinders are presented, see § 1.5.5, for background details about the cylinders being modeled.

4.2 Technical Details

4.2.1 Docking of cylinder derivatives

As the first step towards fully solvated atomistic simulations of DNA and Cylinder complexes, to be discussed later, preliminary investigations into suitable binding sites were undertaken using a high temperature docking protocol. As well as performing docking calculations on the six main systems, that were subsequently developed into extended simulations, a series of high temperature docking calculations for eight other C_{Fe}^{4+} derivatives were performed. For a number of these systems, experimental results from methods such as NMR and X-ray crystallography are as yet unavailable. This study, while not conclusive, helps to provide insight into potential binding modes or starting configurations for further solvated simulation. By carrying out the docking protocol (§ 4.3.1) high temperature MD in a continuum dielectric solvent with explicit mobile counterions was used to identify potential binding sites on a frozen strand of DNA. Docking calculations were performed at 900 K to give effective sampling. The position of the cylinder was then adjusted relative to the frozen DNA to locate conformations of low energy for extraction and further simulation in explicit solvent. Manual docking was also used to check the validity of the binding sites found. This simple method was successful in allowing the dicopper (II) metallo-supramolecular cylinder, $[Cu_2(C_{25}H_{20}N_4)_2]Cl_2$, tetracationic bis iron (II) metallo-supramolecular cylinder, $[Fe_2(C_{25}H_{20}N_4)_3]Cl_4$ and derivative cylinders, Table 4-3, to comprehensively sample many potential binding sites.

It is important to note that the DNA, which would normally be flexible, is immobilised during the docking calculation to prevent degradation at 900 K. As a part of this study three MD Simulations were performed with C_{Fe}^{4+} bound to DNA dodecamer sequences $d(CGCGCATATACG)_2$, $d(CGCGCGCGCGCG)_2$ and $d(ATATATATATAT)_2$. A further set of simulations of C_{Cu}^{2+} bound to the same three DNA sequences were also undertaken. To enable comparisons, from a reference system, simulations of uncomplexed DNA in the presence of water and counterions were also performed for each of the three DNA sequences above.

4.2.2 Comparison of CHARMM22 and CHARMM27 results

Initial reference simulations of $d(CCCCCTTTTTCC)_2$ using the CHARMM22 force-field and $d(CGCGCATATACG)_2$ using the CHARMM27 force-field were performed without the presence of a cylinder. These simulations were run to accumulate a trajectory of 2ns, the period previously seen to include the major dynamical response of the DNA. The variation in sequence also gives insight into the increased stability of a sequence made up of alternating base pairs, $d(CGCGCATATACG)_2$, compared with a DNA strand where consecutive base pairs are of the same kind, $d(CCCCCTTTTTCC)_2$. The initial DNA structure used in the simulation using the CHARMM22 force-field is inherently less stable with the opportunity for inter base pair movements such as rise causing mismatches with equivalent neighboring bases.

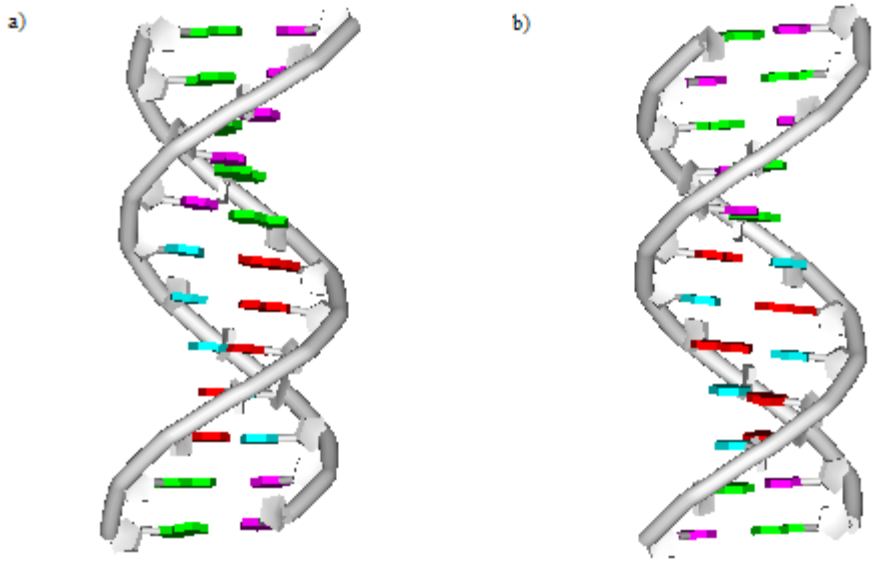


Figure 4-1 Schematic diagrams of , a) $d(CCCCCTTTTTCC)_2$ starting configuration, simulated using the CHARMM22 force-field b) $d(CGCGCATATACG)_2$ starting configuration, simulated using the CHARMM27 force-field.

Following protocol changes during the course of the project the salt concentration and time step were reduced as well as changing the force-field making it difficult to understand the true effect of using the CHARMM27 force-field on the DNA stability. The two parameters that were changed were a reduction in the time step from 2 femtoseconds to 1 femtosecond, to fulfill the greater accuracy demands on the CHARMM27 force-field for the H motions, and a reduction in the no. of background ions present in solution from 82 Na^+ and 58Cl^- to 24Na^+ to better simulate the [Na^+] of experiments. Both simulations were started from the canonical B-conformation and all other treatment was equivalent for both systems.

In the comparison simulations of CHARMM22 and CHARMM27 both uncomplexed DNA strands showed a number of interesting changes during the course of the simulation. After the first nanosecond of simulation the average structure for each 0.25ns portion of simulation was calculated using the first 25 steps.

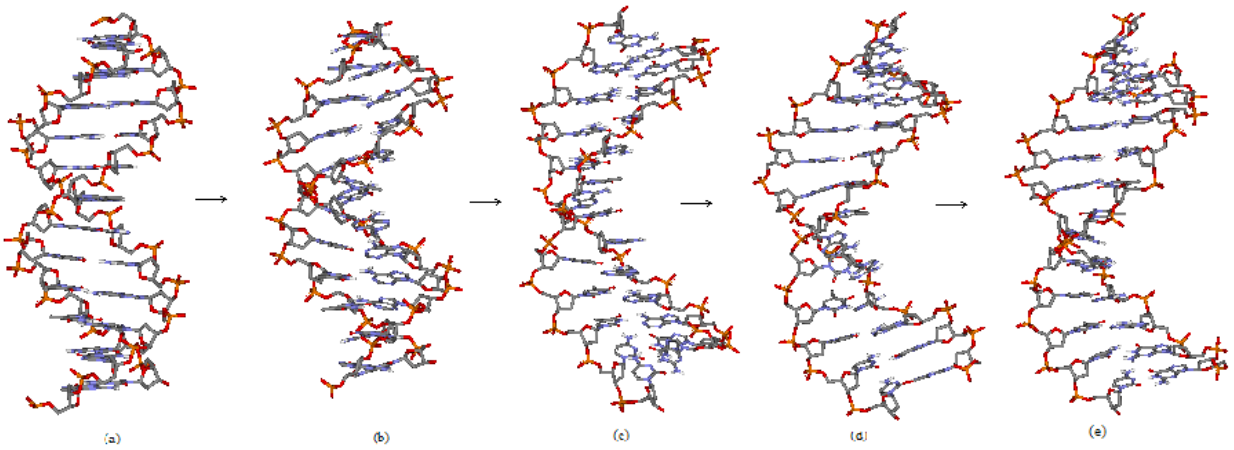


Figure 4-2 Average structures extracted from the CHARMM22 simulation of d(CCCCCTTTTTCC)$_2$ at (a) 0ns (b) 0.25ns (c) 0.50ns (d) 0.75ns (e) 1.00ns .

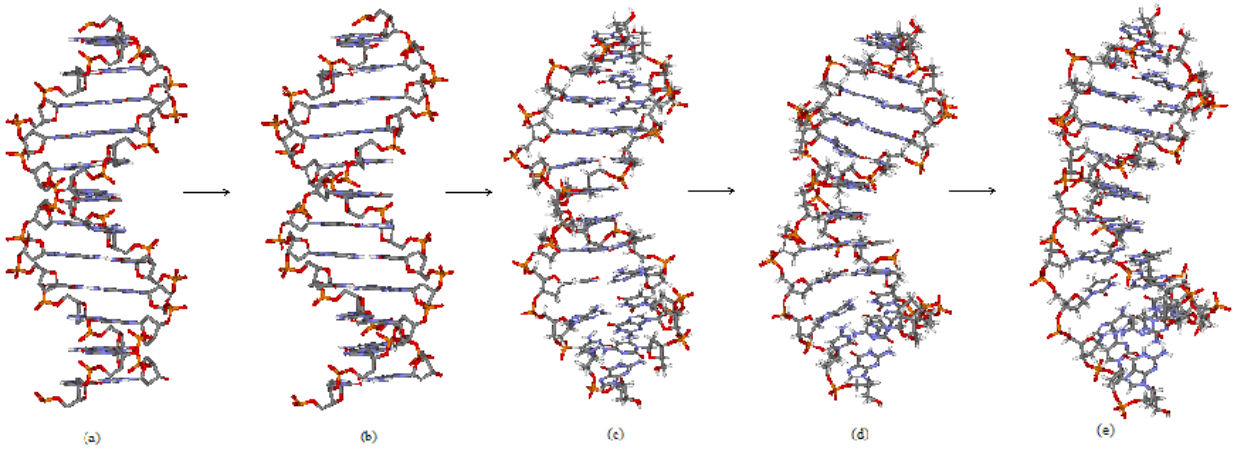


Figure 4-3 Average structures extracted from the CHARMM27 simulation of d(CGCGCATATACG)$_2$ at (a) 0ns (b) 0.25ns (c) 0.50ns (d) 0.75ns (e) 1.00ns, H atoms have been omitted for clarity.

The MD Toolchest suite of analytical packages was used to calculate helicoidal and backbone conformational parameters. These methods probe the structural order of the DNA[264] with an output that displays the conformational parameters as a series of plots known as dials[265], §3.1.1.3. Dynamical behavior of the helicoidal parameters is presented as a series of plots know as windows[20]. Expected values of these parameters for canonical A and B DNA are shown in the following table.

Conformational Parameter	A-DNA	B-DNA
δ / °	83	139
ε / °	178	226
ζ / °	313	203
α / °	319	285
β / °	208	135
γ / °	45	37
χ / °	206	258
φ / °	13	154

Table 4-1 Table to show the expected values of the conformational parameters for canonical A- and B-DNA.[266]

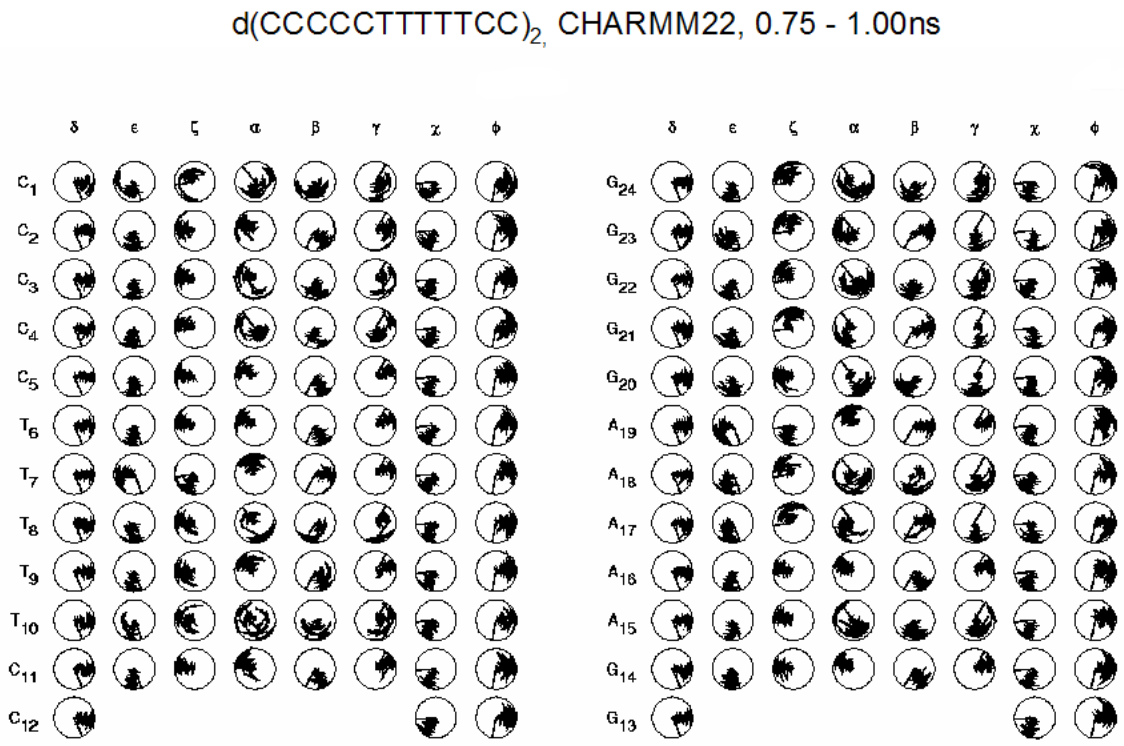


Figure 4-4 Conformational parameters displayed as Dials for the final 0.25ns of the first 1ns of simulation for uncomplexed $d(CCCCCTTTTTCC)_2$. The base to which each plot refers is listed at the side of the figure and the conformational parameter at the top.

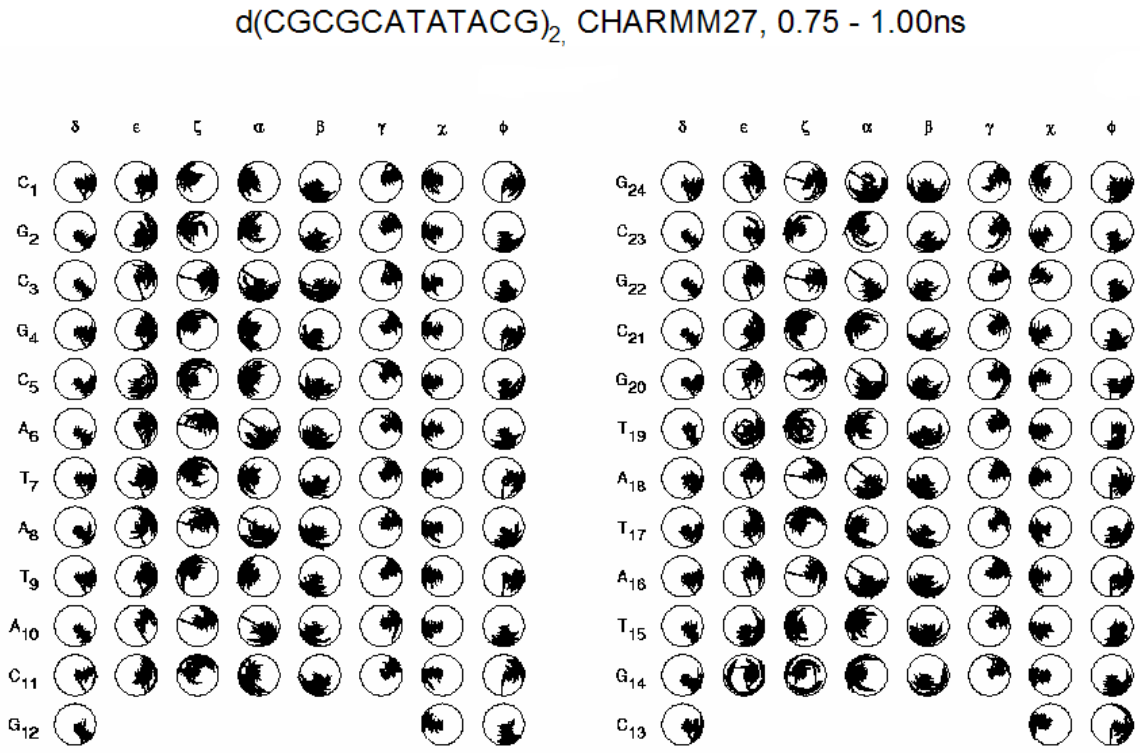


Figure 4-5 Conformational parameters displayed as Dials for the final 0.25ns of the first 1ns of simulation for uncomplexed d(CGCGCATATACG)$_2$.

Analysis of the dials presented in Figure 4-4 shows some interesting transformations during the first nanosecond, evidence that the CHARMM22 force-field is causing the conformation of d(CCCCCTTTTTCC)$_2$ to change from the canonical B-DNA form of the starting configuration. One conformational parameter that demonstrates this particularly clearly is the exocyclic sugar-base (glycosyl), χ, torsion angle, that is seen to be fluctuating around the value of 206^o consistent with that expected for canonical A-DNA. There are other dramatic events within this portion of the simulation specifically relating to the α^o γ^o-flips, an example of which is seen for nucleotide T_8, Figure 4-4, where following this relatively quick event the trajectory is seen to develop with α in the *t* state, rather than the usual *g* $^-$state and γ^o being *t* rather than the usual g^+, Figure 4-6. The interruption of the base stacking in this fashion demonstrates the flexibility allowed by the CHARMM22 force field and the subsequent characteristics seen in further analysis below which ultimately result in the transition from B-DNA to A-DNA.

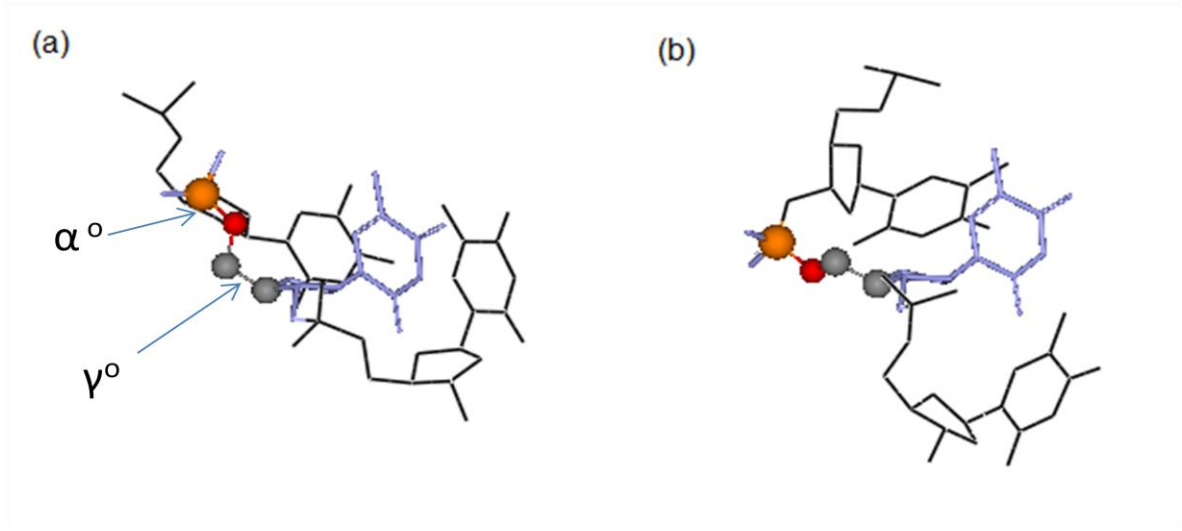


Figure 4-6 T_7, T_8 (highlighted by stick representation) and T_9 at (a) 0.25ns (b) 1ns, showing the α, γ-flip with P (orange), O (red), C5' and C4'(gray, ball and stick representation) .

The system simulated with the CHARMM27 force-field demonstrates the ability for CHARMM27 to maintain the B-DNA form.

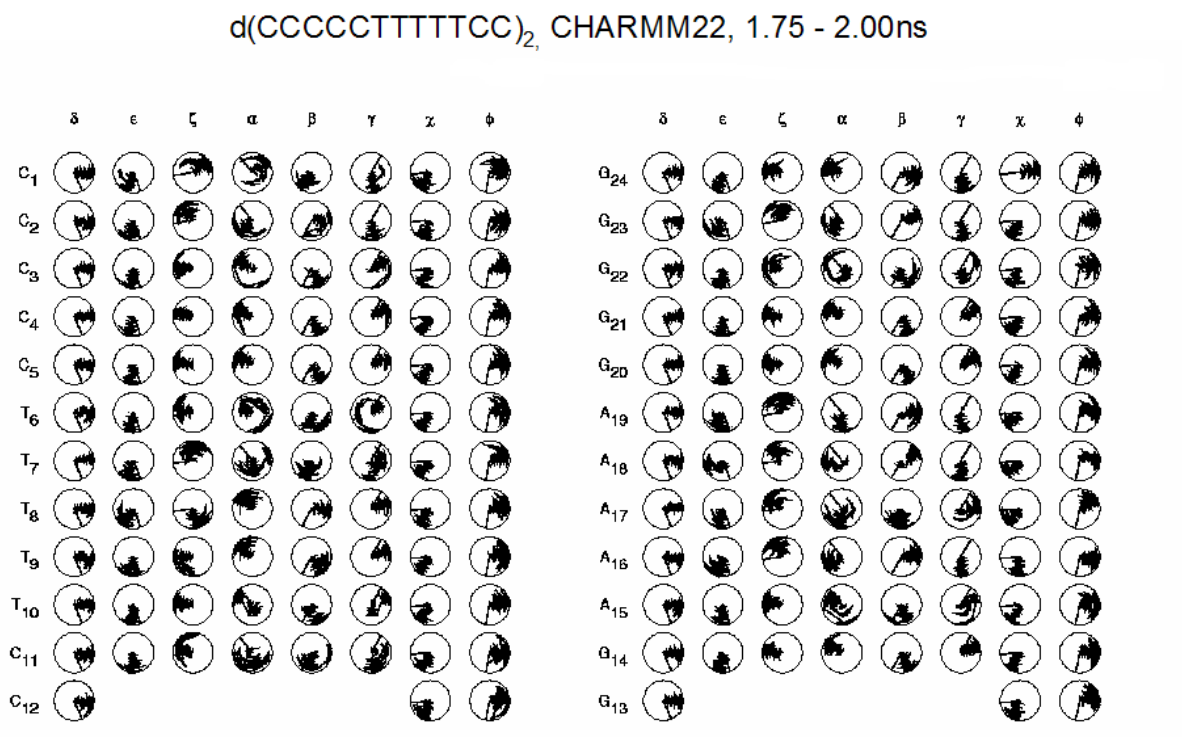


Figure 4-7 Conformational parameters displayed as Dials for the final 0.25ns of the second nanosecond of simulation for uncomplexed d(CCCCCTTTTTCC)$_2$ DNA.

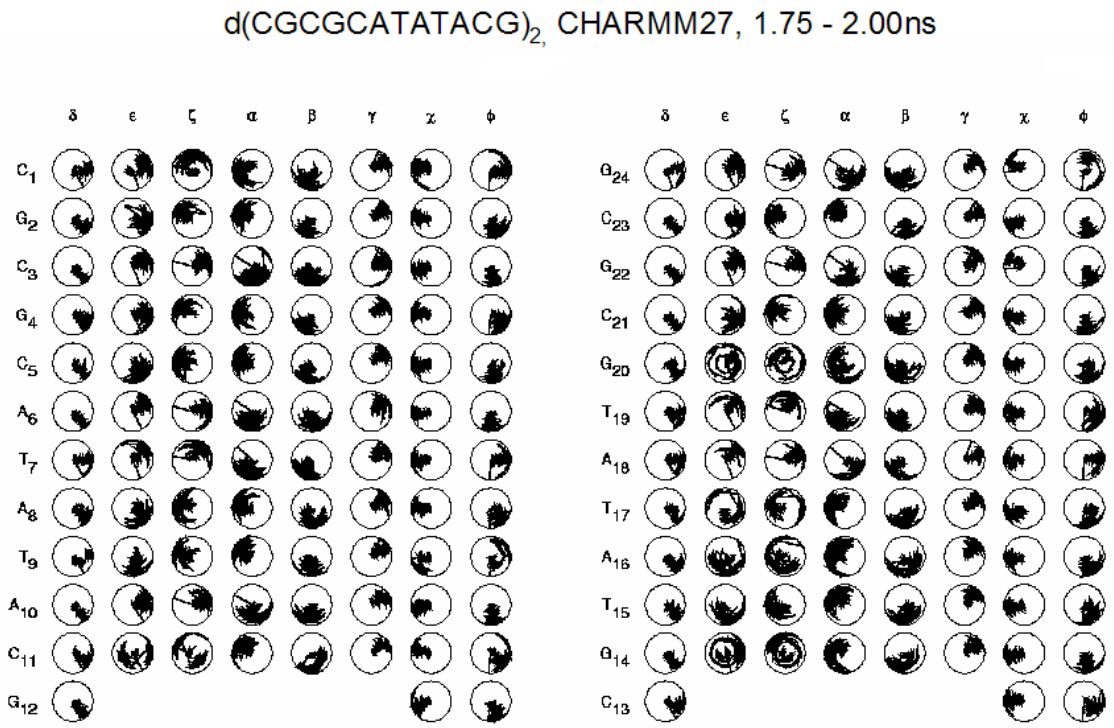


Figure 4-8 Conformational parameters displayed as Dials for the final 0.25ns of the second nanosecond of simulation for uncomplexed d(CGCGCATATACG)$_2$ DNA.

The second nanosecond of these reference simulations extends the trends started during the first nanosecond. The simulation using the CHARMM22 force field continues with a high proportion of the helical parameters displaying A-DNA like properties. For d(CGCGCATATACG)$_2$, being modeled with the CHARMM27 force-field, parameters including γ, χ, φ, and δ show the majority of values calculated to have remained inline with the values for B-DNA. There are a number of anomalies such as the φ value for T_7, T_9 and C_1 that show a shift towards the A-FORM. Changes due to end effects for C_1 G_{24} might be expected during an extended simulation where no constraints are placed on any base-pair configurations.

Further information relating to the dynamics of the system can be obtained from the Complementary Base-pair Parameters and Base-pair Step Parameters which are calculated and displayed in windows as follows.

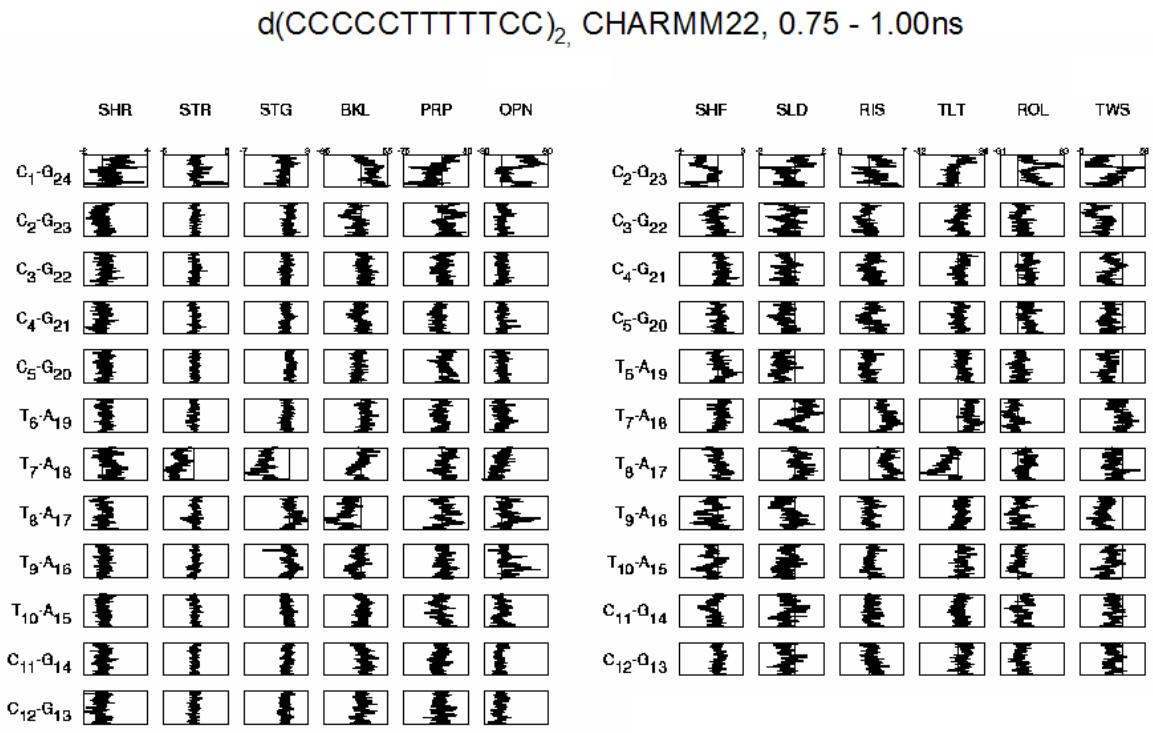


Figure 4-9 Base-pair parameters displayed as Windows for the final 0.25ns of the first nanosecond of simulation for uncomplexed d(CCCCCTTTTTCC)$_2$ DNA.

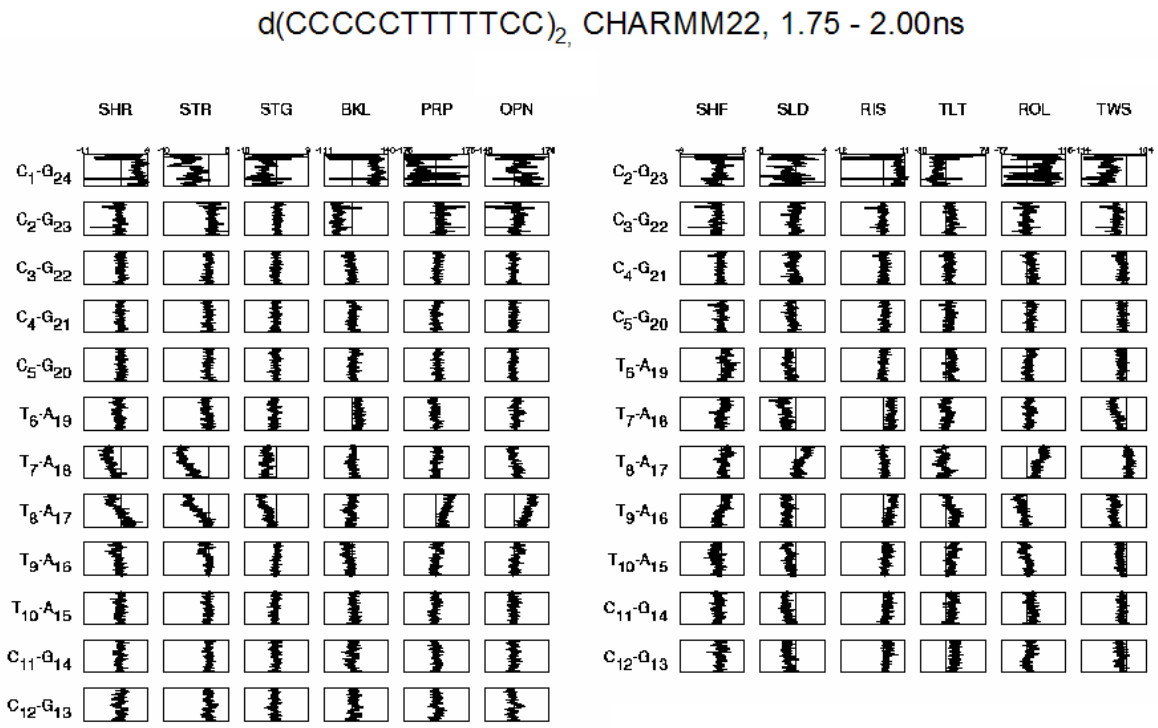


Figure 4-10 Base-pair parameters displayed as Windows for the final 0.25ns of the second nanosecond of simulation for uncomplexed d(CGCGCATATACG)$_2$ DNA.

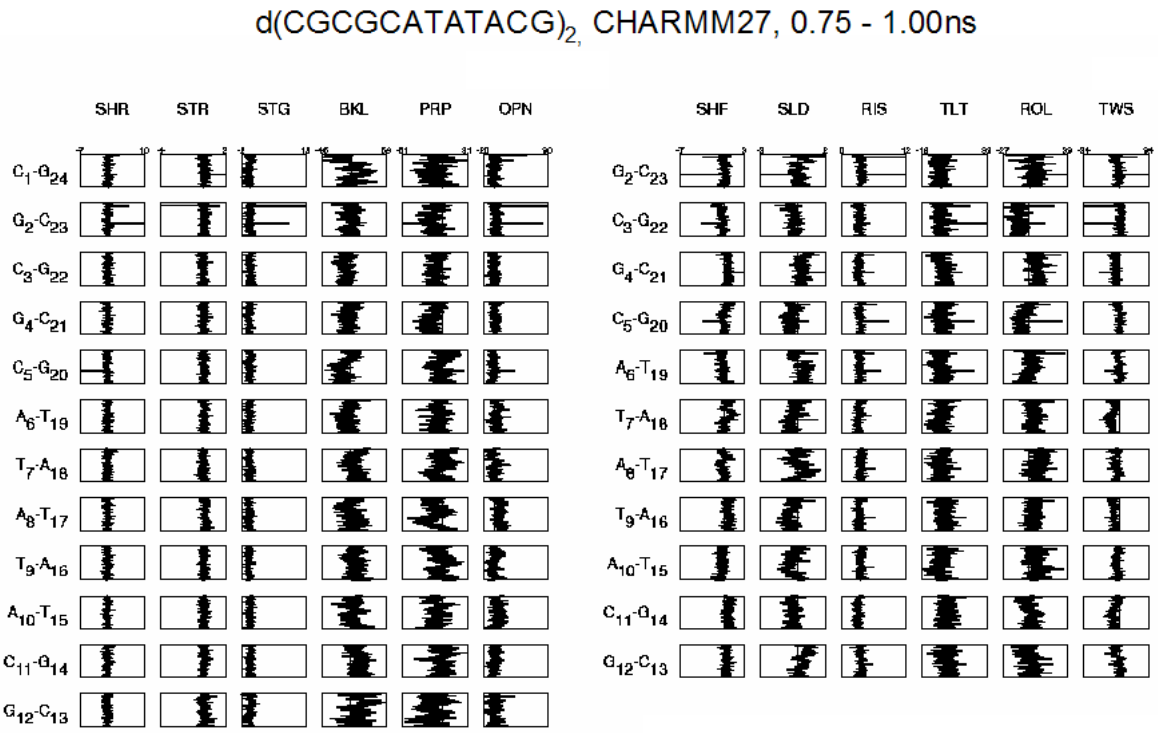


Figure 4-11 Base-pair parameters displayed as Windows for the final 0.25ns of the second nanosecond of simulation for uncomplexed d(CCCCCAAAAACC)$_2$ DNA.

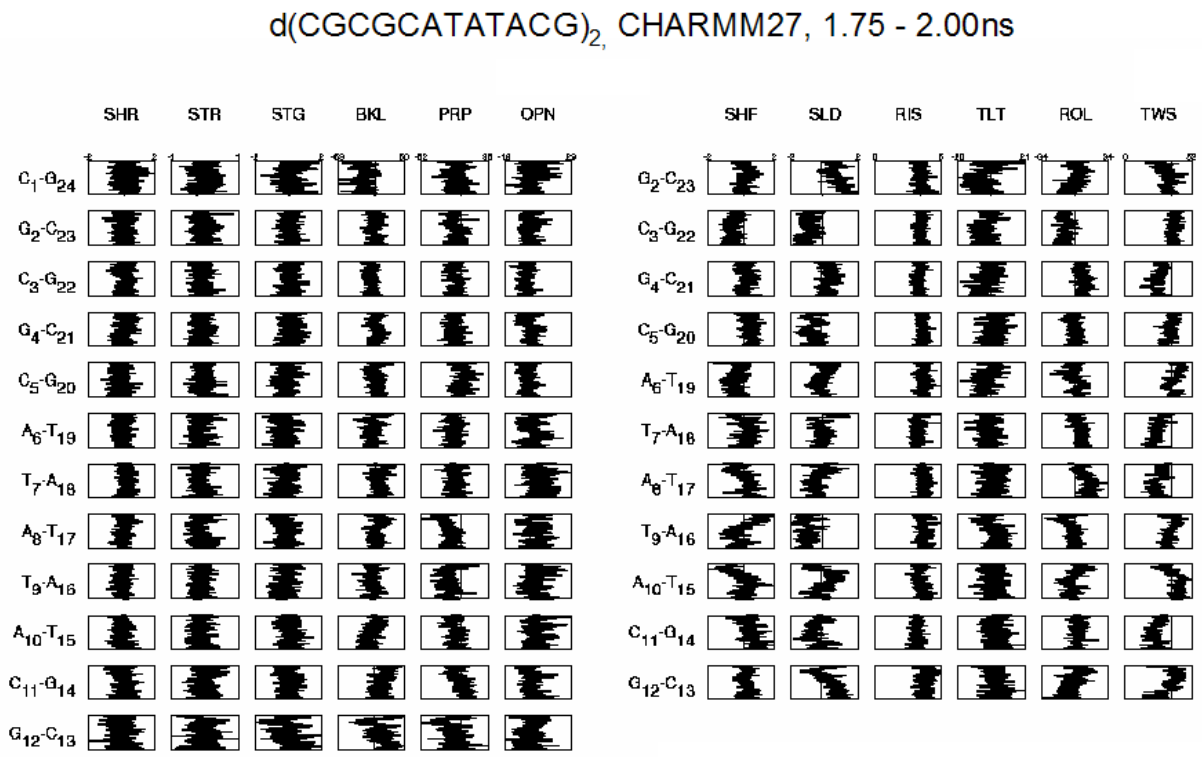


Figure 4-12 Base-pair parameters displayed as Windows for the final 0.25ns of the second nanosecond of simulation for uncomplexed d(CCCCCAAAAACC)$_2$ DNA.

Comparison of the calculated values, with the average values in the following table makes it possible to gain a greater understanding into the finer detail of the DNA dynamics.

Complementary Base-pair Parameters	A-DNA	B-DNA
Shear / Å	0.0	0.0
Stretch / Å	- 0.45	0.0
Stagger / Å	0.19	- 0.01
Buckle / °	-0.09	0.0
Propeller / °	13.7	- 13.3
Opening / °	- 4.6	- 4.1
Base-pair Step Parameters		
Tilt / °	0.0	0.0
Roll / °	0.0	0.0
Twist / °	32.7	36.0
Shift / Å	0.0	0.0
Slide / Å	0.0	0.0
Rise / Å	2.56	3.4

Table 4-2 Average values of the base pair parameters as found in crystal structures.[266,267]

When comparing the plots of base pair parameters note that the scales are not the same. Having noted the scale it is apparent from the plots that d(CCCCCTTTTTCC)$_2$ being simulated with the CHARMM22 force-field experiences a reduction in twist confirmed by the correlated slide values representative of A-DNA and smaller when compared with the equivalent base pairs in d(CGCGCATATACG)$_2$ being simulated with CHARMM27. Other than the extremities of d(CGCGCATATACG)$_2$ the general pattern of the dynamics is for the structure to vary about a mean value consistent with B-DNA, while the distribution plots show a smooth normal distribution about the average value. It is worth noting that the parameter with the greatest variation is that of slide; this is indicative of B-DNA for which greater variability of the value between dimer steps is observed.

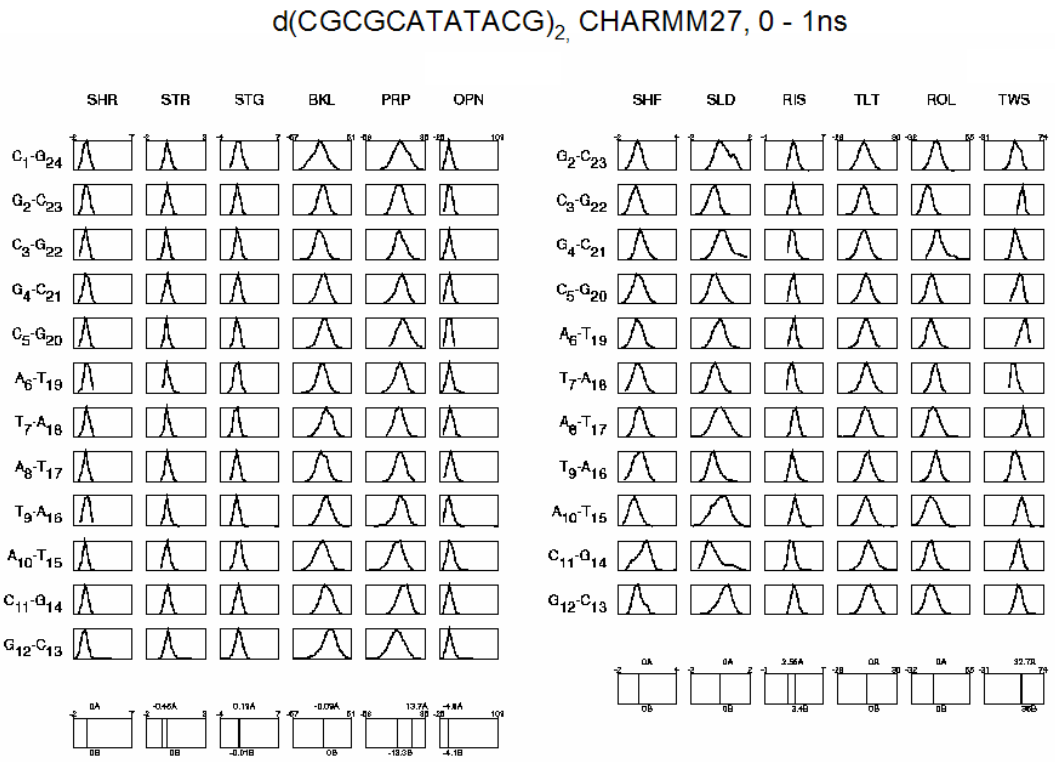


Figure 4-13 Distribution of step parameters calculated over the first nanosecond for d(CGCGCATATACG)$_2$.

Confirmation of the dynamical properties deduced from Figure 4-9 to Figure 4-13 can be obtained from the 3DNA analysis. While 3DNA uses a different reference frame to define the geometry, results seen in the following plots are in agreement with the dials displayed above. In Figure 4-14, the dominant feature of the d(CCCCCTTTTTCC)$_2$ system at two nanoseconds, shown as the light blue line, is indicative of the transformation that has taken place resulting in the uncoiling of the DNA strand and subsequent mismatch. The other area of significant activity is as expected at the end of the DNA strand. Base pairings that were present in the initial conformation are retained in the more stable region between C2-G24 and C5-C20. A mismatch base pair is formed between A18-T8 leaving T7 and A16 unpaired. The rearrangement observed could be characteristic of the DNA strand being modeled which allows base pairs to stagger and reform with equivalent neighboring nucleotides.

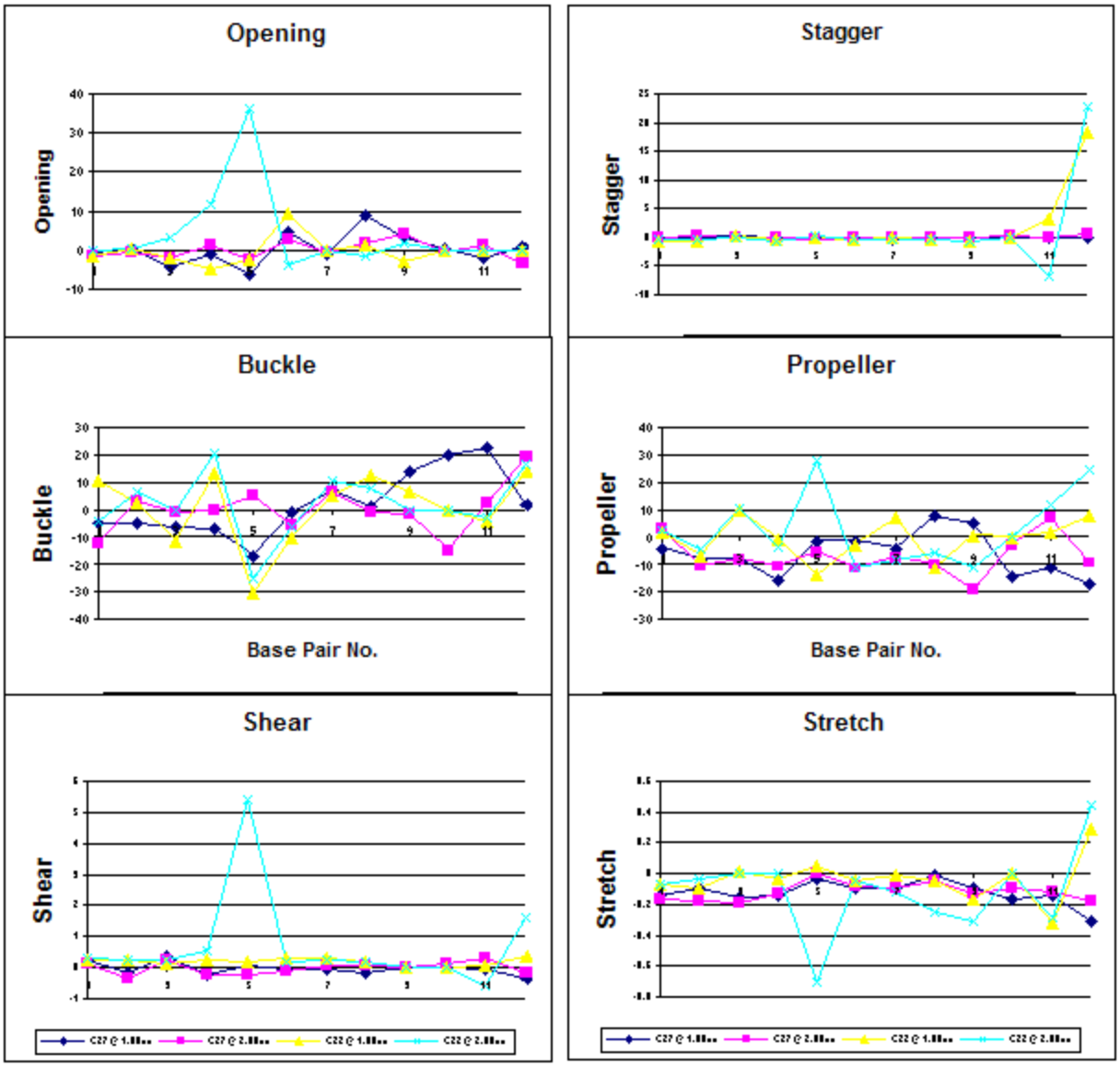


Figure 4-14 Plots from the results of 3DNA calculations for base pair parameters.

A summary of the final nanosecond of this comparison study is shown in Figure 4-15 and Figure 4-16. The simulation of d(CGCGCATATACG)$_2$ with the CHARMM27 force-field has been extended as a reference for the cylinder-DNA complexes.

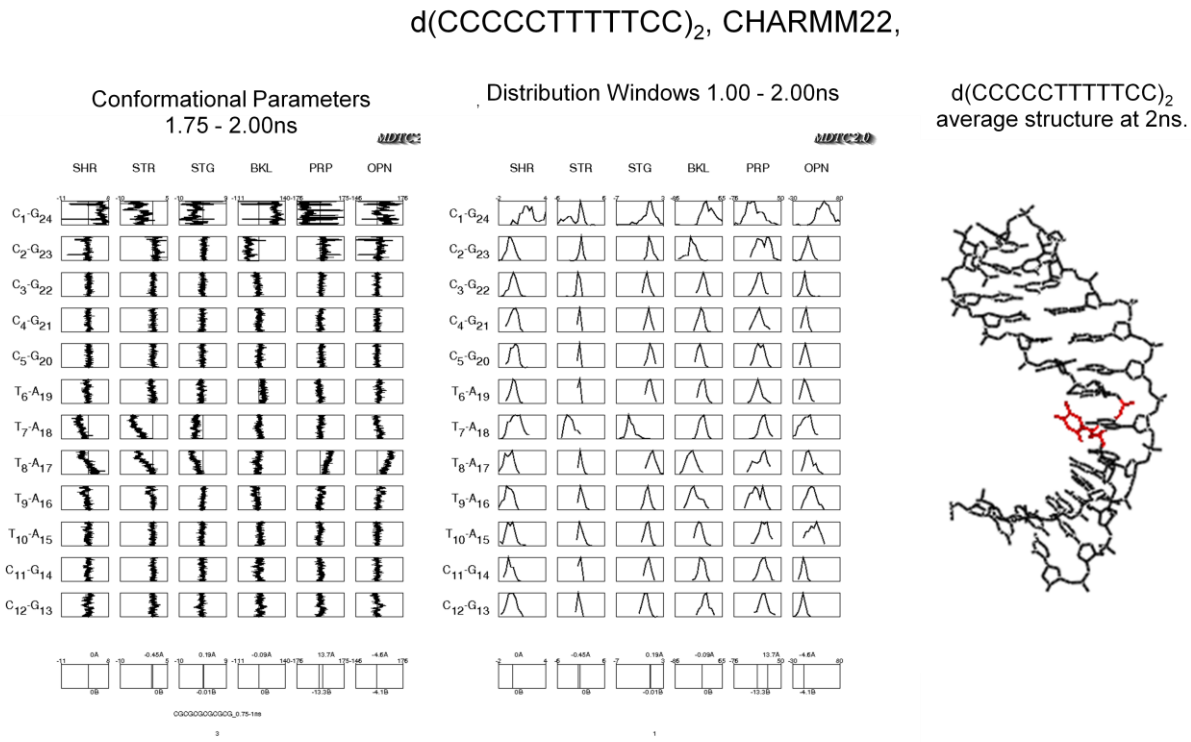


Figure 4-15 Conformational Parameters and Distribution Windows for d(CCCCCTTTTTCC)$_2$ relative to the average structure calculated from the final 25 picoseconds of simulation at 2nanoseconds, T_8 (red) showing the source of the B-Form disruption.

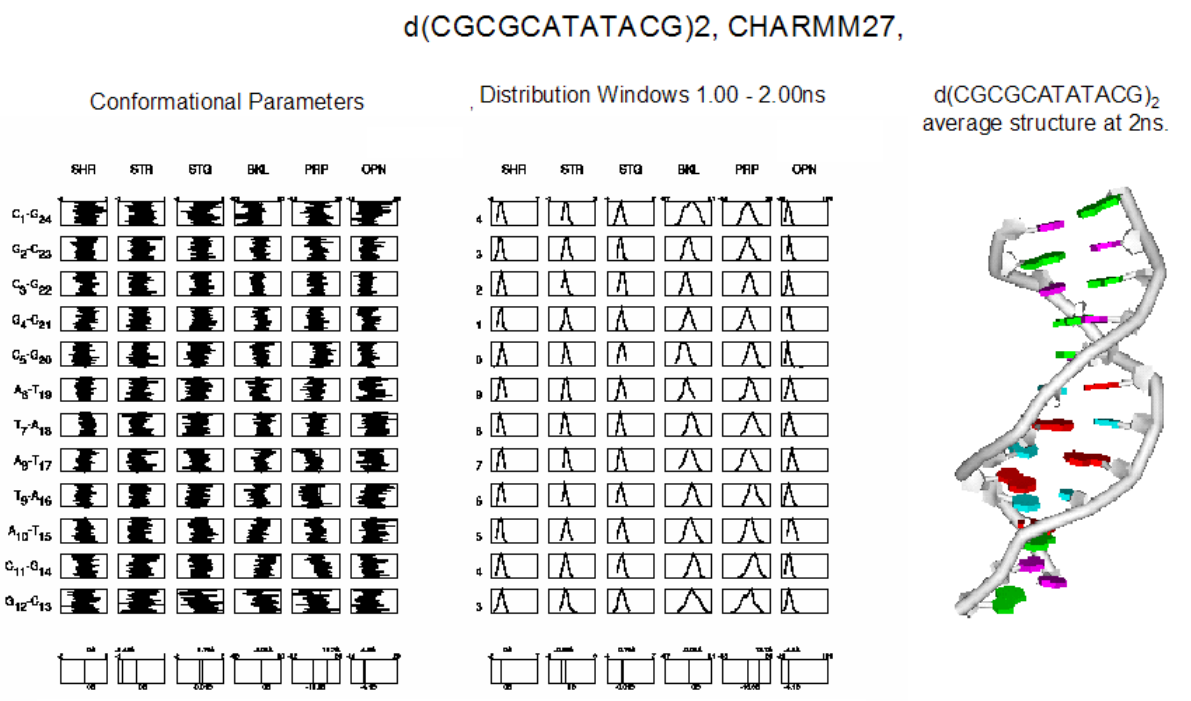


Figure 4-16 Conformational Parameters and Distribution Windows for d(CGCGCATATACG) relative to the average structure calculated from the final 25 picoseconds of simulation at 2nanoseconds.

4.2.3 Simulation overview

Compounds were modeled initially using the CHARMM22 force-field and following the validation experiment with the CHARMM27 all-atom force-field.[268] Significant progress in the development of empirical force fields and molecular dynamics simulation methods has lead to a more reliable description of the structure, energetics, and dynamics of nucleic acids[269]. However, some limitations related to the improper treatment of the equilibrium between the A and B forms of DNA[270] and deviations of helicoidal parameters from canonical B values[271] have been reported. The overstabilization of the A form relative to the B form of DNA with the CHARMM22 force field is well documented in the literature[272] and confirmed by the reference system simulated in this work. However, reoptimization of the CHARMM22 all-atom nucleic acid force field has created the newer CHARMM27 with small but important changes in both the internal and interaction parameters relative to CHARMM22[273]. CHARMM27 appears to treat well the equilibrium between A and B forms[274] of DNA and the influence of the environment, such as the water activity[275].

While the CHARMM27 has been chosen as the main force field for this research, due to practical compatibility and familiarity with current procedures within the group, it is important to have an understanding of improvements with other force fields to understand what is available while at the same time being able to compare and interpret current discussions in the literature correctly. A revised and improved version of AMBER4.1 showed better agreement with experimental data as a result of adjustment of internal force field parameters[276] before being superseded with PARMBSC0[277]. The modified force field has enabled stable trajectories above the 10-ns range with the correct representation of backbone parameters to maintain the α/γ configuration in a *gauche*$^-$, *gauche*$^+$ state rather than causing the transition to a *gauche*$^+$,

trans geometry with the introduction of severe DNA distortions as seen during longer 50-ns trajectories with the previous version.

An alternative nucleic acid force field is also referenced and compared within the literature and is called the Bristol-Myers Squibb (BMS) force field [274,278]. Both the CHARMM27 and AMBER force field parameters are based on the reproduction of experimental results for nucleic acid oligomers (e.g., condensed phase structural properties of DNA and RNA) and consistency with small molecule results obtained from quantum mechanical calculations and experimental data. The BMS force field was developed, in part, by adaptation of the CHARMM22, QUANTA and AMBER force fields. The backbone angle and dihedral parameters were derived from quantum mechanical calculations with refinements based on a series of MD simulations. All the force fields used condensed-phase MD simulations in the final stage of the parameter optimization. The CHARMM27 force field has also been applied to model compounds to evaluate the contributions from the individual moieties to the overall conformational properties of DNA and RNA[273].

Simulations in this research were performed using DL_POLY[279]. The conversion from CHARMM to DL_POLY force field formats was achieved using a purpose-built program that interpreted the CHARMM .prm and .crd files. The cylinders were modeled by keeping both FeN_6 units rigid in C_{Fe}^{4+} and both CuN_4 units rigid in C_{Cu}^{2+}. Fe centers were assigned charges of +2, charges on all other atoms were smoothed to give an overall cylinder charge of +4 with geometry taken from the crystal structure. The Cu centers were assigned charges of +1, charges on all other atoms were smoothed to give an overall cylinder charge of +2 with geometry taken from the crystal structure.

All hydrogen atoms were assigned a mass of 2 u. This gave good energy conservation in trial NVE simulations with a 2 fs time step. The SPME[280] was used to evaluate the electrostatic interactions. All simulations were performed in the NVT ensemble using the Hoover thermostat with a time constant of 0.1 ps and a time step of 2 fs. Orthorhombic periodic boundary conditions were used to mimic an infinite system. All conditions were the same for both the C_{Fe}^{4+} and C_{Cu}^{2+} systems other than the force field parameters that were calculated with Gaussian to give a better description of the copper cylinder.

4.3 Simulation Protocol

The docking protocol was designed to identify a suitable initial configuration, perform an addition of solvent, relax the system and then perform an extended simulation.

4.3.1 Docking Protocol

The position of where to place the cylinder relative to the DNA was found by performing docking calculations in an ionic atmosphere to identify favorable binding sites. A dodecamer of B-DNA and one cylinder molecule were embedded in a neutralising atmosphere of Na^+ ions within a 45x45x60 Å box, with DNA aligned along the *z* axis. Ions were added using the DLPOLY utility SOLVADD with solvent-solvent and solute-solvent sigma tolerances of 0.56 and 0.64 respectively. The DNA was immobilised and an *NVT* MD simulation performed for 1ns at 900 K. No solvent was included at this stage, but the relative dielectric constant was set to 80. The lowest energy conformations were considered with one such conformation being extracted after having been observed to place the cylinder in a central location. Other similar low energy configurations were also observed that would make suitable starting points for duplicate simulations.

4.3.2 Simulation Protocol

Simulations were performed with the DNA, ligand and ions solvated in a water environment. The water molecules for use in the simulation were taken from an equilibrated liquid water simulation and were added to the configuration that was extracted from the docking stage. Again molecules experiencing unfavorable overlap were removed from the simulation box. Overlap was defined as occurring when the distance between an atom of a water molecule and any atom in the DNA, Cylinder or the background ions (Na^+ or Cl^-) was less than the sum of the respective Van der Waals radii multiplied by a tolerance factor of 0.64, this value of less than one was due to the high electrostatics leading to close contact between ions and water. This left a total of 3720 water molecules in the final system. For the reference systems with no cylinder present 3758 waters were included to achieve the same water density within the simulation cell.

4.3.2.1 Tethering Protocol

When considering delicate super molecular structures such as DNA it is necessary to have a mechanism for allowing relaxation of the solvent and DNA from the initial starting configuration. By utilising the tethering potential in DLPOLY gentle relaxation can be undertaken prior to a completely mobile extended simulation. To do this DL_POLY allows atomic sites to be tethered to a fixed point in space, r^0, their initial position. A harmonic potential is used for this purpose with the energy being defined as,

$$U(r_{io}) = \frac{1}{2}k(r_{io})^2$$

Equation 4-1

Where *k* is the tethering force constant, r_{io} is the distance of the atom from its position at t=0.

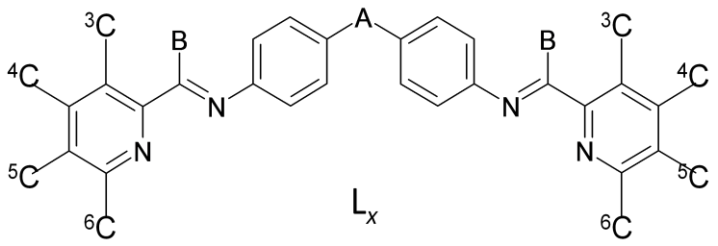
The DNA atoms were tethered to their original positions with a force constant of 100 kcal / mol / Å2, the force constant was gradually reduced over five 5ps tethering runs before the tethering

potentials were then removed completely and the final configuration from the last tethered simulation used to start a multi-nanosecond simulation.

4.4 Docking results of derivative cylinders

The high temperature docking was carried out with each cylinder derivative shown in Figure 1-24 being docked to an immobilised DNA dodecamer, $(CGCGCATATACG)_2$.

Docking Cylinder Derivatives, $[M_2(L_x)_3]^{4+}$



Ligand , L_x	A	B	C
L_1	CH_2	H	^{5}Me
L_2	CH_2	Me	H
L_3	CH_2	Me	^{4}Me
L_4	O	H	H
L_5	O	H	^{3}Me
L_6	O	H	^{4}Me
L_7	O	H	^{5}Me
L_8	O	Me	^{4}Me
L_9	CH_2	H	^{4}Me

Table 4-3 Series of derivatives $[M_2(L_x)_3]^{4+}$, where M = Fe^{2+} and L_x = the ligand derivative with functional groups A, B and C.

Following the docking calculation favorable binding sites of low potential energy were extracted from the trajectory and allowed to relax at 210 K for 1ns. Analysis of the

accumulated trajectory from the docking calculation shows the cylinder to sample many potential binding sites as seen for C_{Fe}^{4+}-L_1 below.

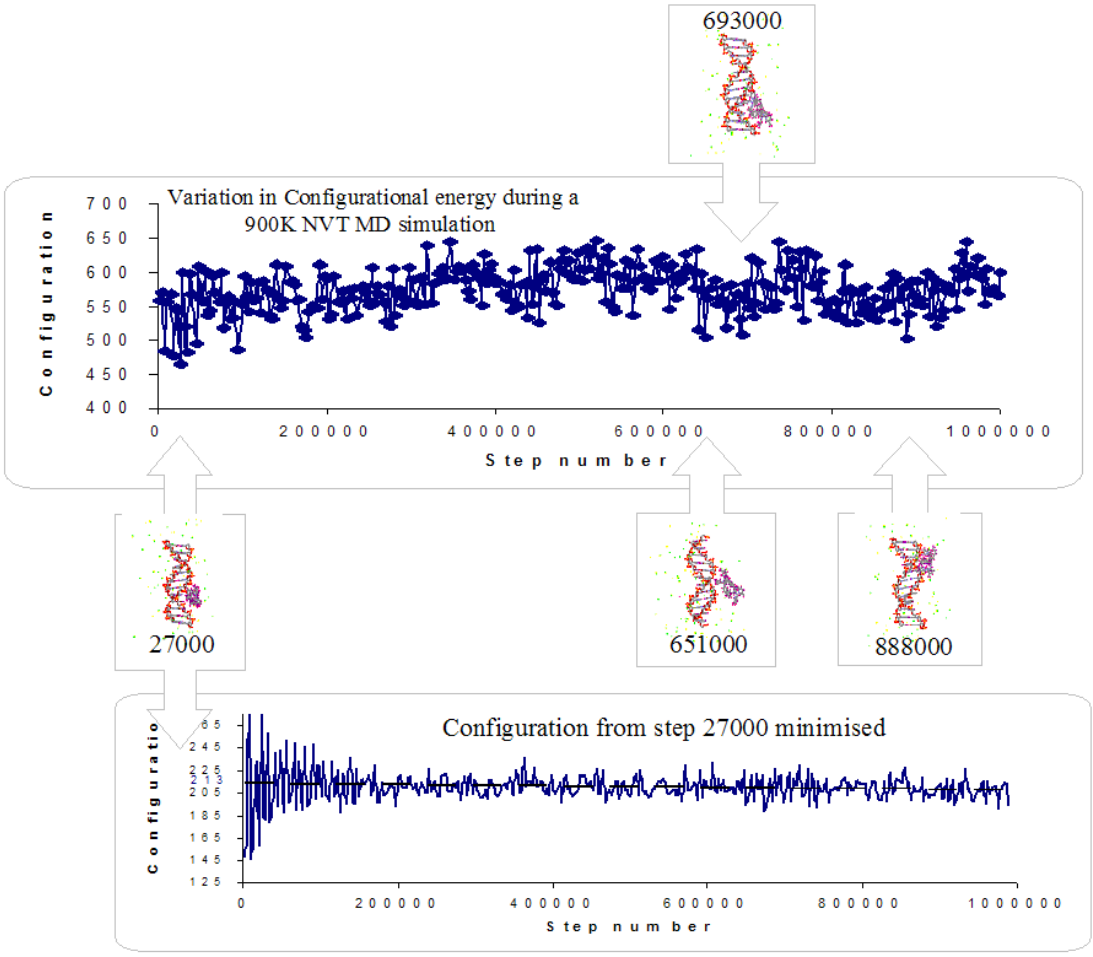


Figure 4-17 Configurational energy plot showing change in configurational energy during the high temperature docking calculation at 900K. Also, the configurational energy plot of time step 27000 extracted from the 1ns docking trajectory and minimised to give an indicator of the final configurational energy of the system.

The binding site chosen for the initial starting configuration was one of relatively low energy and within the central region of the DNA strand to minimise end effects. The presence of more than one potential binding site will enable duplicate simulations to be run from comparable starting configurations to investigate the reproducibility of the final results.

When considering the significance of the values obtained from these calculations it is important to remember that they were not carried out in the presence of explicit solvent. For this reason

none of the values extrapolated from the data are real configurational energies and cannot therefore be used to accurately calculate the binding energy of the DNA and cylinder derivative complex.

For this reason the Binding Energies could be calculated as an extension to this work by performing solvated minimisation calculations from which the binding energies will be accurately calculated and placed in order of strength to further understand the binding properties and mechanisms. On completion of the further study it would be interesting to see if the crude trend, seen in the following table, could be improved to confirm the apparent correlation between the bulk of the cylinder and binding energy.

Ligand	A	B	^{4}C	Configurational Energy DNA + Cylinder (Kcal/mol)	Configurational Energy Cylinder (kcal/mol)	Binding Energy (kcal/mol)
L_1	CH_2	H	^{5}Me	281	200	81
L_2	CH_2	Me	H	490	767	-277
L_3	CH_2	Me	^{4}Me	647	806	-159
L_4	O	H	H	490	573	-83
L_5	O	H	^{3}Me	560	623	-63
L_6	O	H	^{4}Me	639	756	-116
L_7	O	H	^{5}Me	684	605	79
L_8	O	Me	^{4}Me	761	646	115
L_9	CH_2	H	^{4}Me	685	762	-77

Table 4-4 Indicator of configurational energies for the derivatives of CFe^{4+}. Binding energies are to be recalculated as further work with explicit solvent present.

While the simple high temperature docking protocol employed in this work appears to provide extensive sampling it is also worth being aware of the potential limitations. By its nature DNA is an inherently flexible molecule allowing it to accommodate a wider range of potential binding molecules than may be apparent when it is constrained to a ridged structure. To investigate the extent of the limitation to possible binding configurations the high temperature

docking calculation was repeated allowing the DNA some mobility during two different simulations, one where the DNA was loosely tethered and another with more tightly tethered DNA strands. To practically implement the FIELD file was edited to define the 383 atom positions being restrained for each DNA strand along with the tethering force constant of either 17 or 43 kcal/mol/Å^2.

The configurational energy for these two tethered systems enabled us to see that although the configurational energies cannot be directly compared, due to the variable in the DNA's tethered configurational energy; as expected there is a greater range of configurational energies available in the observed oscillatory behavior when considering the more loosely tethered DNA system.

The ability to achieve more effective binding when the DNA is able to flex to accommodate the cylinder is a point worth noting when considering sequence selectivity for which exposure to the base pairs deep in the groove are important. While the high temperature docking protocol with rigid DNA is sufficient for sampling a large range of potential binding sites, and at least as effective as manual docking, it would be beneficial in future work when considering the variety of ligands, L_1 – L_9 to modify the method to ensure that the true properties associated with larger ligands are observed and not just a result of a limited method.

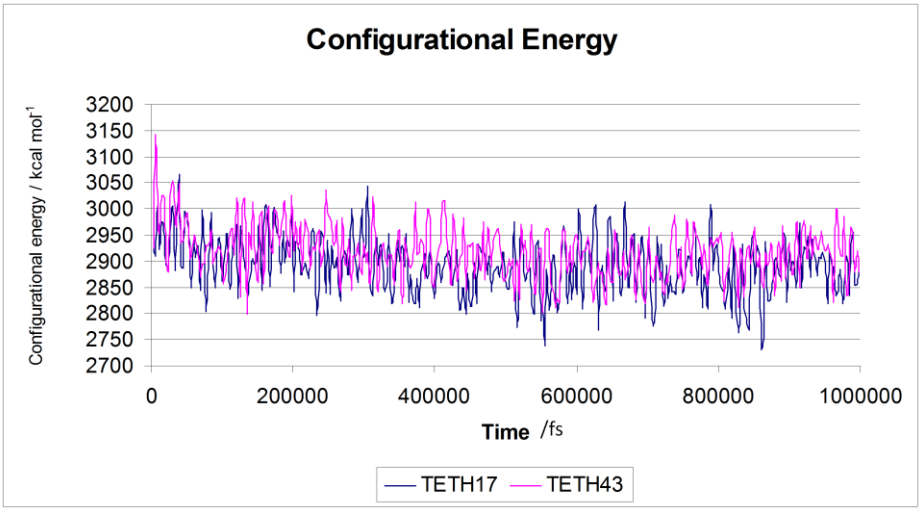


Figure 4-18 Configurational energy plot showing the difference between using a loosely tethered, TETH17 (17kcal/mol/Å^2), and a more tightly tethered, TETH43 (43 kcal/mol/Å^2), DNA strand for the high temperature docking calculations.

Introducing flexibility into the DNA structure during the docking calculation may also enable investigation into any competition that might be present between sites such as minor groove and major groove. While a cylinder might have a preference for one area of the DNA structure utilizing a flexible DNA strand may highlight occasions when an expanded minor groove offers a favorable binding site.

4.5 Reference Systems

4.5.1 Background

Simulations of the uncomplexed DNA solvated in an aqueous solution with sodium counterions present were performed as a reference against which to measure the effect of the cylinder on the DNA in the main simulations. All simulations from this point onwards were performed with the CHARMM27 force-field. Each DNA sequence used in the cylinder DNA systems was simulated as an uncomplexed strand to validate the method used. Reference simulations were initially performed for 1ns and then run in parallel with the production stage of the DNA-cylinder simulations to enable continued comparison for the whole trajectory.

4.5.2 Reference systems results

For all three DNA systems, d(CGCGCATATACG)$_2$, d(ATATATATATAT)$_2$ and d(CGCGCGCGCGCG)$_2$ the uncomplexed DNA remained stable throughout the 1ns simulation and retained an approximately linear B-DNA form. Figure 4-19 shows representative snapshots of the DNA structure taken after 1ns of simulation.

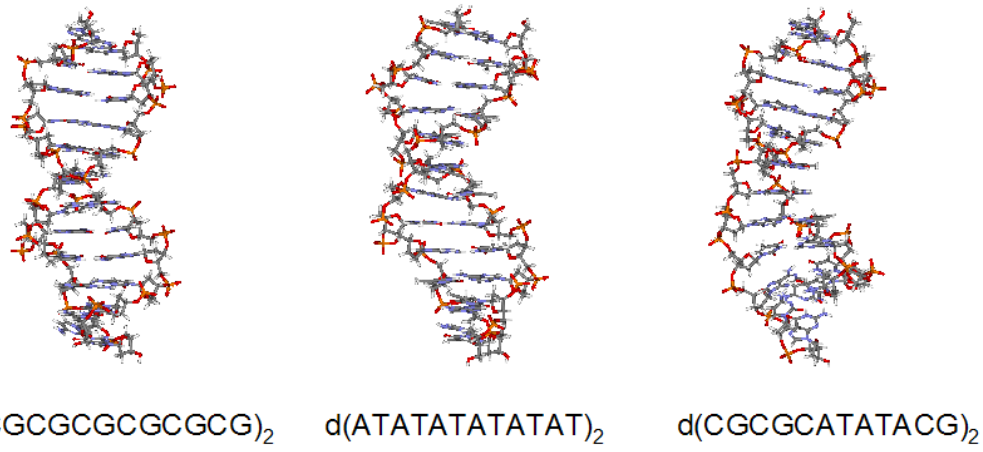


Figure 4-19 Conformation of uncomplexed DNA after 1ns

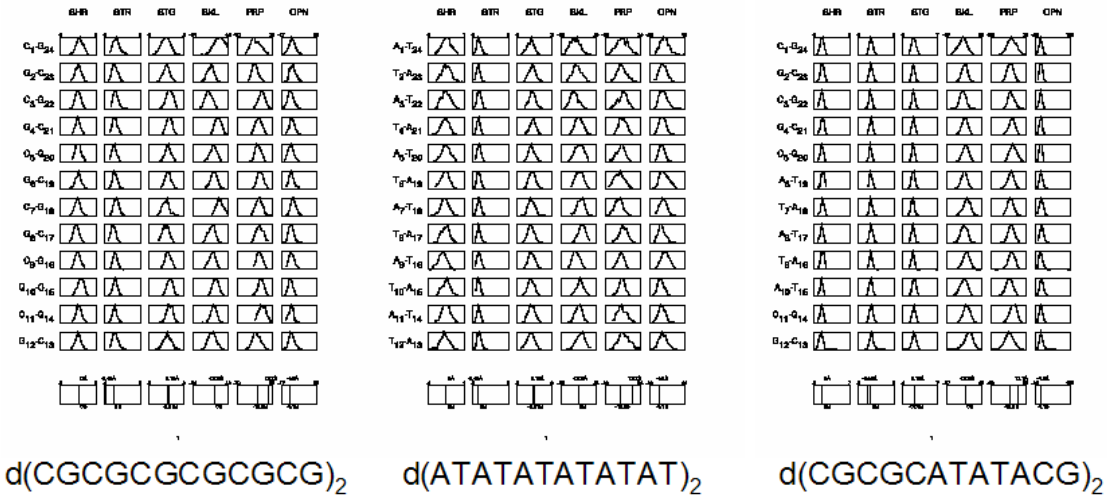


Figure 4-20 Distribution of conformational parameters between the start of the simulation and 1ns.

The distribution of helical conformational parameters were calculated and confirmed the expectation that activity was in the majority of cases distributed evenly about the mean value expected for canonical B-DNA, Table 4-1. The stability observed during this period of simulation is consistent with the findings of the CHARMM22 and CHARMM27 force field comparison and demonstrates that the canonical B-DNA form is stable within solvated systems. A visual inspection of the systems evolution was undertaken by superimposing the starting configuration and average structure calculated from the final 25ps of the 1ns simulation.

$d(CGCGCGCGCGCG)_2$ $d(ATATATATATAT)_2$ $d(CGCGCATATACG)_2$

Table 4-5 Schematic representation of the DNA starting and final average structure superimposed. The red backbone is the starting structure and the blue backbone is the final average structure at 1ns.

In each of the DNA structures the variations seen are small and within the range of the structural fluctuations present in these large flexible complexes.

Following the initial 1ns reference simulations each system has been developed to enable direct comparison with the bound system and ultimately the effect of the cylinders presence.

With the sequence $d(CGCGCATATACG)_2$ accommodating a variety of base pairs and steps docking investigations were primarily focused on this sequence with $d(CGCGCGCGCGCG)_2$ and $d(ATATATATATAT)_2$ being run to further clarify findings in various regions of this strand. Retrospective visualisation of the conformations extracted from the extended simulations once again confirms the stability toward the DNA B-form of the chosen strands when simulated with the CHARMM27 force-field.

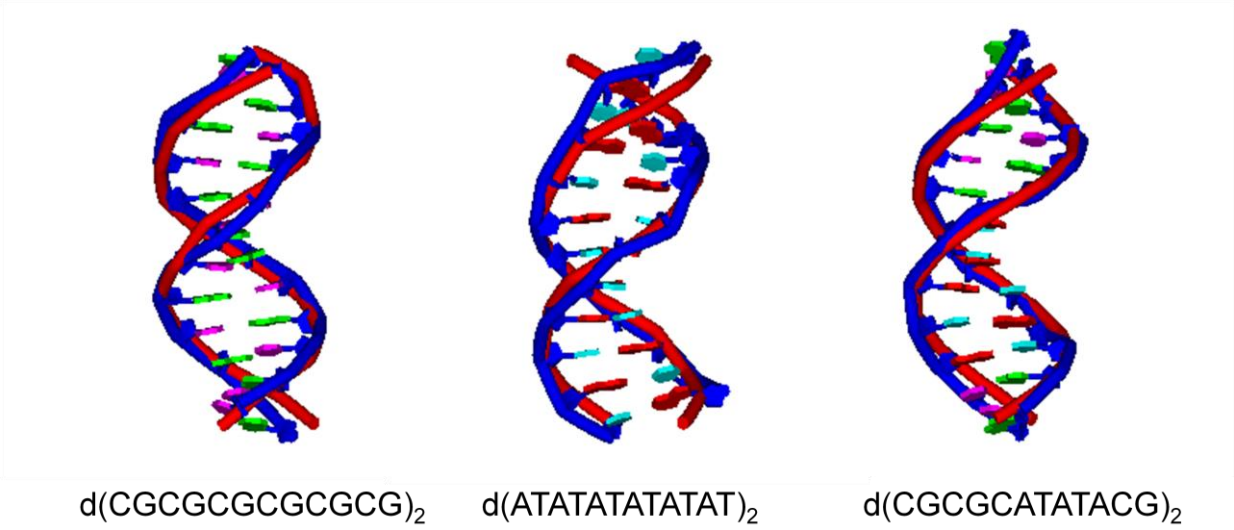


Figure 4-21 Schematic representation of the DNA starting and 4ns average structure superimposed. The red backbone is the starting structure and the blue backbone is the final average structure at 4ns. The bases displayed belong to the blue backbone and are included to indicate the stability of all parts of the DNA structure.

With the increased stability over longer timescales observed during these simulations compared to those conducted with CHARMM22 reduced effects of the cylinder on the DNA conformation might be expected.

5 Results for DNA with groove binding ligands

5.1 C_{Fe}^{4+} and C_{Cu}^{2+} with d(CGCGCATATACG)$_2$

5.1.1 Docking results

Docking calculations were performed with both C_{Fe}^{4+} and C_{Cu}^{2+} using the protocol described above. Energetically favourable configurations were inspected for suitability, with a binding site centrally located along the DNA being chosen to minimise end effects during the extended simulation.

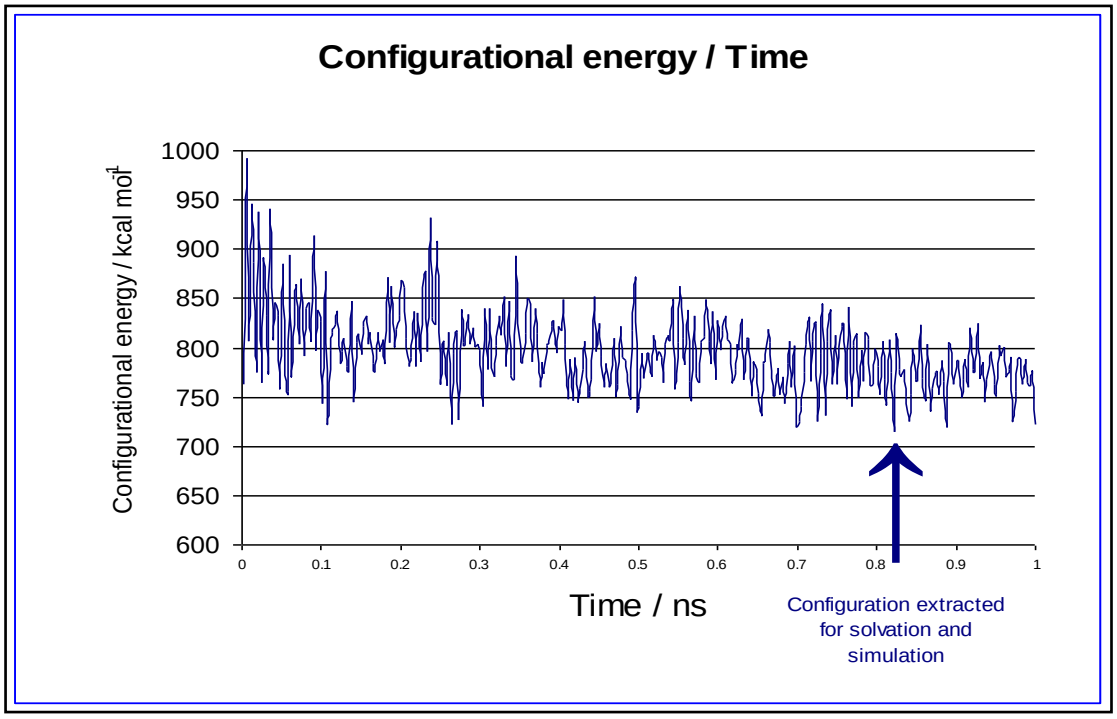


Figure 5-1 A plot showing configurational energy during the 1ns docking calculation of C_{Fe}^{4+} and d(CGCGGCATATACG)$_2$.

The configurational energies for the two cylinders docked with d(CGCGCATATACG)$_2$ as extracted from the docking calculation are in the following table.

Conformation	Configurational Energy (kcal/mol)
d(CGCGCATATACG)$_2$ – C_{Fe}^{4+}	714.9
d(CGCGCATATACG)$_2$ – C_{Cu}^{2+}	700.8

Figure 5-2 Configurational energies of the starting points chosen for the two systems.

Other favourable binding sites were identified and will be used in future work to investigate the reproducibility of findings from these studies.

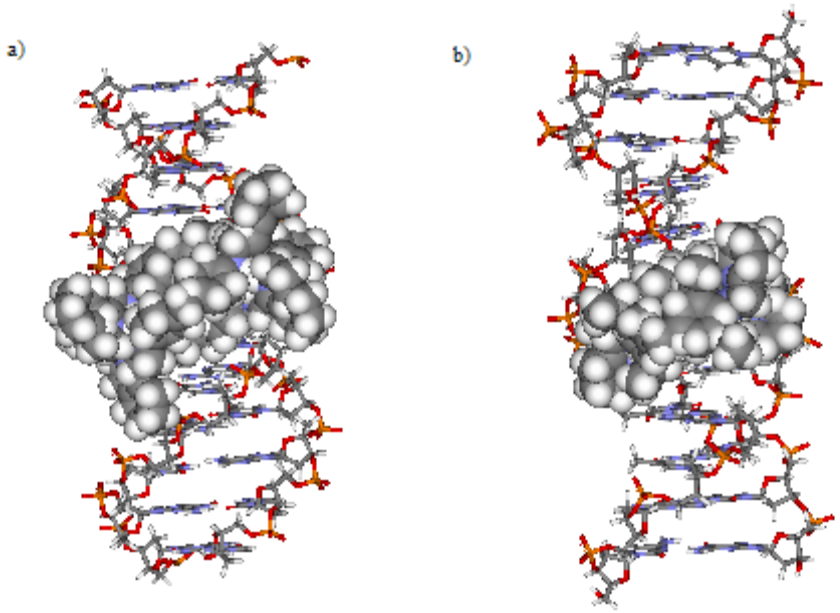


Figure 5-3 Two starting conformations extracted from the docking calculations. (a) C_{Fe}^{4+} (b) C_{Cu}^{2+}.

5.2 C_{Fe}^{4+} system

5.2.1 Overview

The starting conformation shows the cylinder lying symmetrically along the major groove spanning base pairs C3-G22 to T9-A16, Figure 5-3. The DNA remains close to the original canonical B-DNA starting configuration for the duration of the 3ns simulation. While there were no base pair mismatches observed the end base pair did separate demonstrating the ability for fraying at the end of the DNA to which the cylinder was bound. This behaviour was not present with the reference uncomplexed DNA that was seen to progress to 4ns with the end base pairs still intact.

During the course of this simulation the cylinder was seen to migrate from being positioned symmetrically along the major grove to a more exposed position on the spine of the major

groove. A cartoon of this movement by the cylinder can be seen in Figure 5-4. The colour coding system used in Figure 5-4 will be used to indicate the stage of the simulation being shown. In diagrams showing superimposed structures, such as Figure 5-5, visual comparisons can be made of the time-dependence of the system.

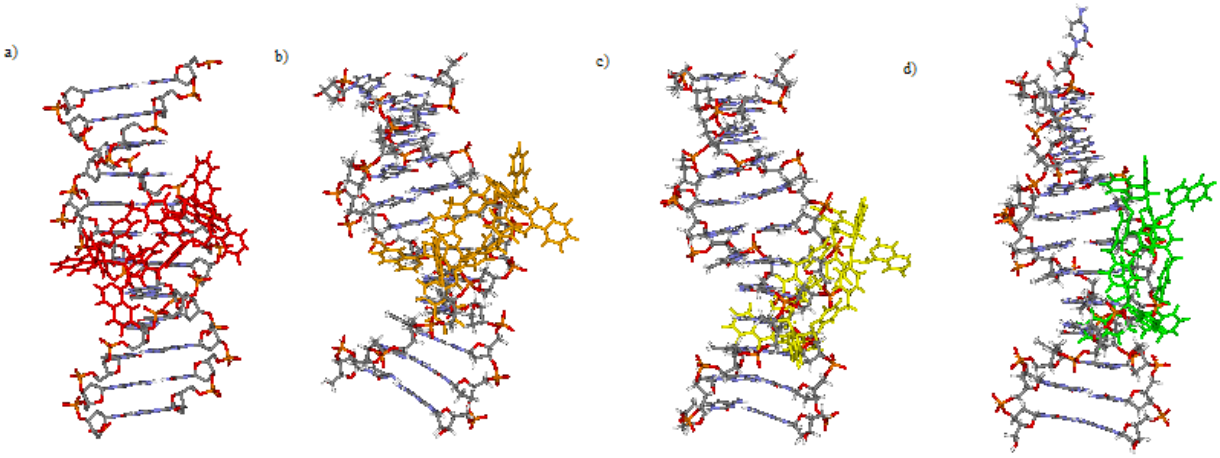


Figure 5-4 The migration of CFe^{4+} from an initial binding site in the major groove to a position on the DNA spine. (a) starting configuration, red (b) 1ns , orange (c) 2ns, yellow (d) 3ns, green.

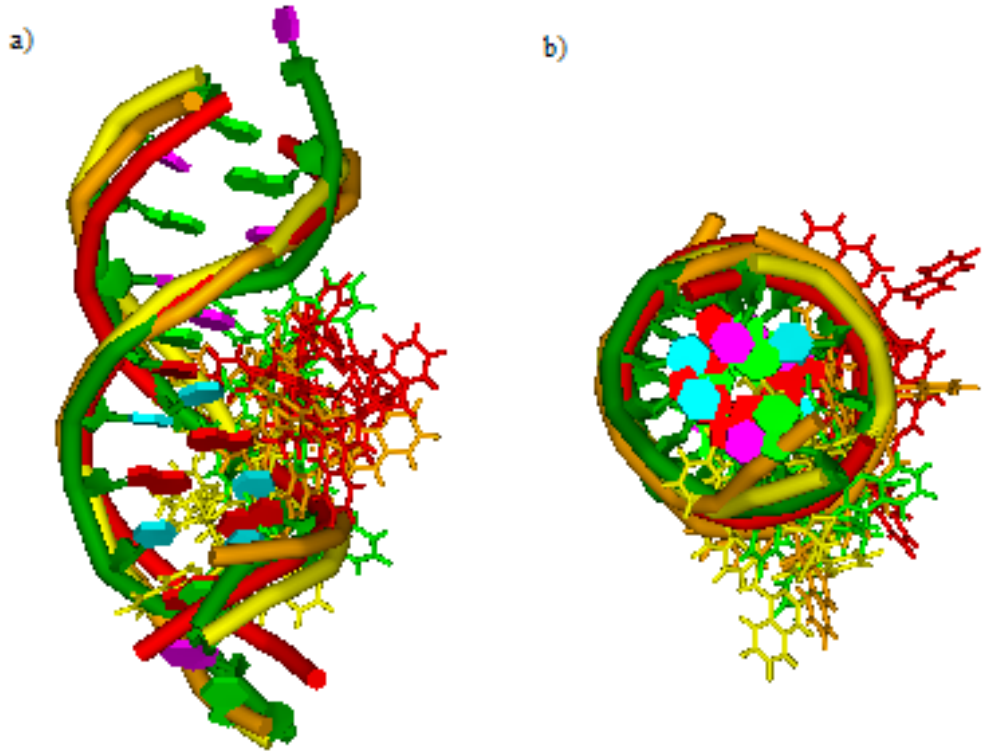


Figure 5-5 A schematic diagram of the four extracts superimposed to show the observed similarities of the DNA over 3ns of simulation. a) side view b) end view.

With the four extracted configurations superimposed the DNA in Figure 5-5 can be seen to maintain the canonical B-form of the starting structure with deviations within the fluctuations due to flexibility.

To clarify the time-dependence of the data, and to facilitate comparison with the uncomplexed DNA simulation, the simulation was dissected into three consecutive 1ns portions, and sometimes into 0.25ns portions, and each of these were analysed separately.

5.2.2 First nanosecond

Conformational and helicoidal parameters were calculated in 0.25ns portions for the trajectory. The conformational parameters for the 0.75 – 1.00ns and 2.75-3.00ns are shown in Figure 5-6 and Figure 5-7.

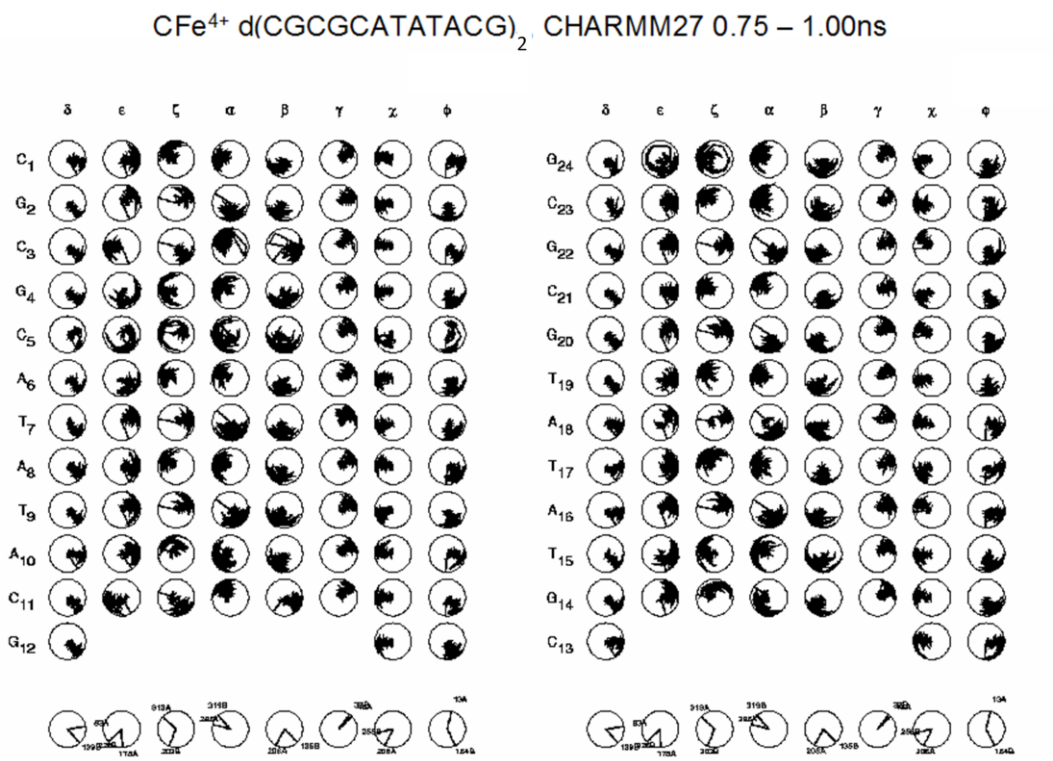


Figure 5-6 Conformational parameters for 0.75 – 1.00ns of CFe^{4+} d(CGCGCATATACG)$_2$.

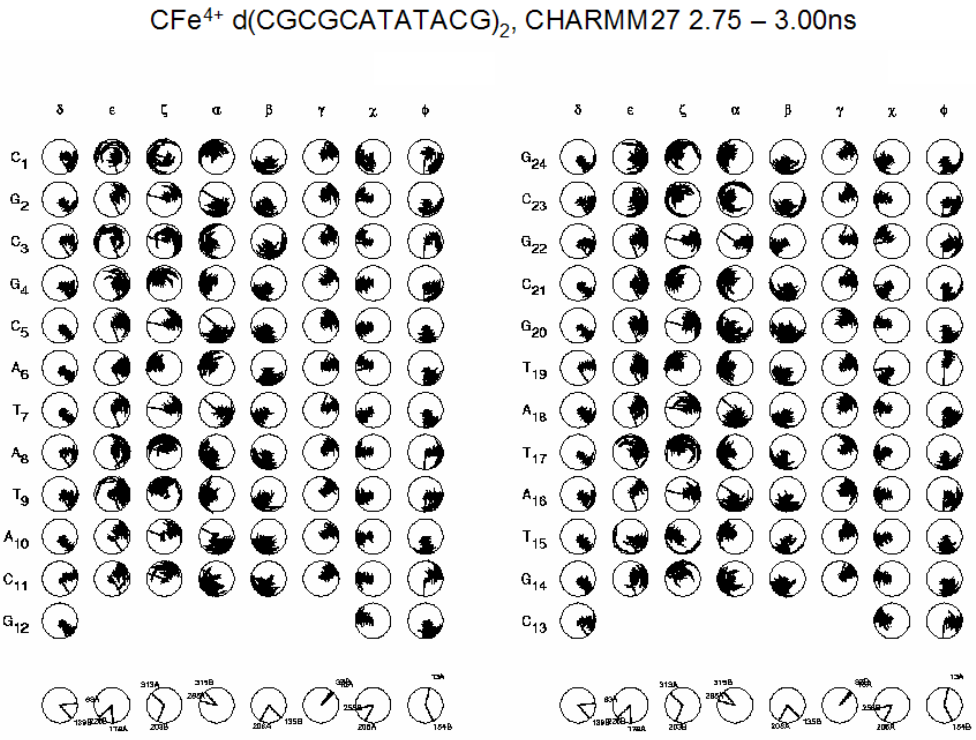


Figure 5-7 Conformational parameters for 2.75 – 3.00ns of CFe^{4+} d(CGCGCATATACG)$_2$

The general trend in the conformational parameters for the 0.75ns-1.00ns, Figure 5-6 and 2.75ns-3.00ns Figure 5-7 were the same. The internal sugar angle, δ was stable in both cases, and consistent with the B-form (Table 4-1). The exocyclic sugar-base (glycosyl) torsion angle χ also remained stable. The phosphodiester torsion angles α and γ showed significant dynamical activity throughout both portions of the trajectory. Rotation of these torsion angles led to an oscillation of the position of the phosphorus atom. The five-membered sugar ring is commonly puckered in an envelope form in which one atom lies out of the plane to one side or other, Figure 5-8. The values displayed in the plots for the sugar puckering parameter show that fluctuations within the range suitable for the C2'-*endo* conformation are maintained throughout the simulation with the exception of T_{19} that can be seen to display C3'endo puckering characteristic of A-DNA. In general the conformational parameters are well defined and consistent with canonical B-DNA and comparable to the reference simulation.

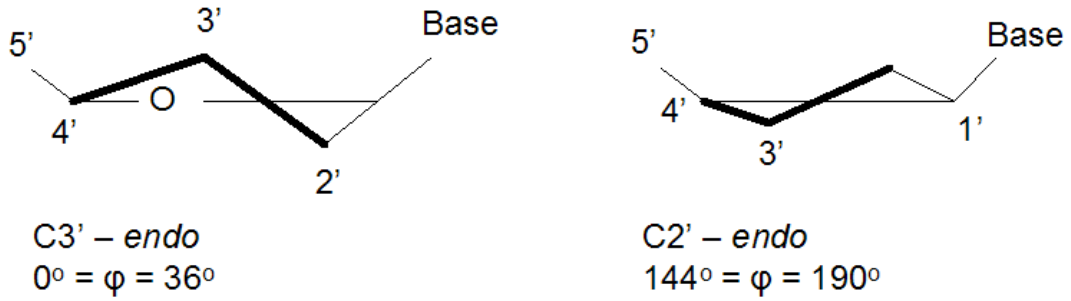


Figure 5-8 Common sugar puckering modes and their associated values of the φ parameter for canonical B-DNA.

The intra-base and inter-base helicoidal parameters for the last 0.25ns of the first nanosecond were stable. However, a number of values for the twist inter-base parameter deviated towards values associated with canonical A-DNA, Figure 5-9 . These coincide with the motion of the cylinder towards the DNA spine from having been cradled in the bottom of the major groove, Figure 5-4 (b).

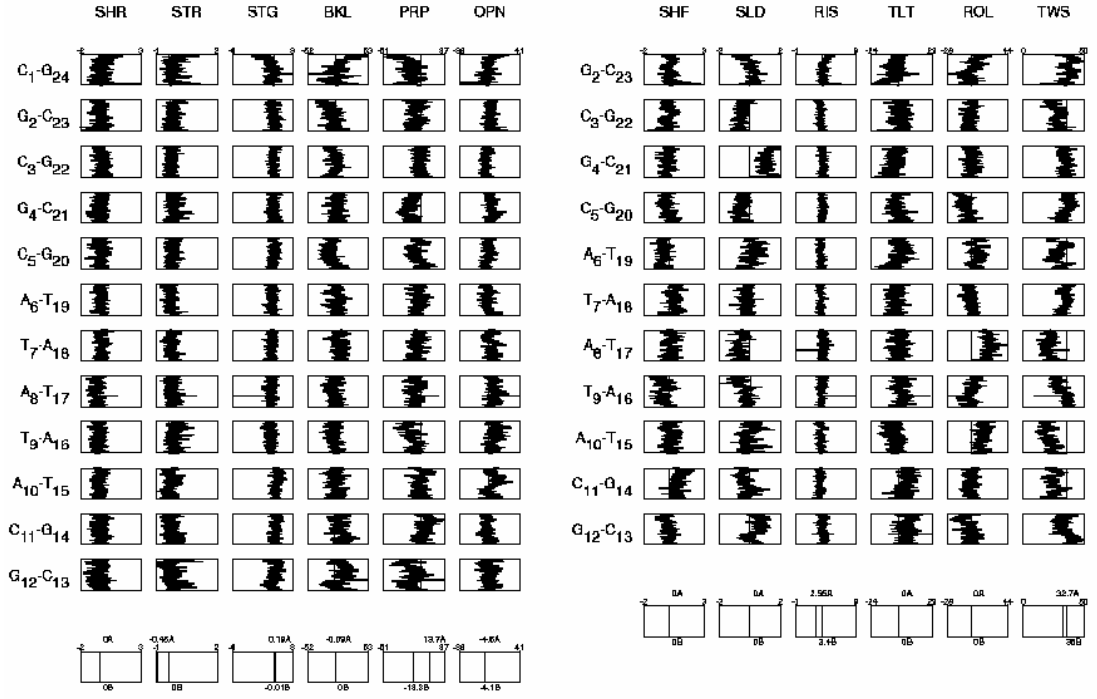


Figure 5-9 Helicoidal parameters for 0.75 – 1.00ns of CFe^{4+} $d(CGCGCATATACG)_2$

The variation seen for, A_8-T_{17} and A_{10}-T_{15} with T_9-A_{16} remaining less perturbed could be related to the points of close contact either side of the void in the cylinder between the propeller ends and body. Similarly this could account for the variation for the twist value of A6-T19 which while less significant than the other two variations is accessible to the end of the cylinder "propeller" at the same time as having bridged T7-A18 at the narrow point of the cylinder. This same regular feature of disruption can be observed during the initial stages of the simulation when effects of the cylinder spanning alternating base pairs can also be seen in the dials plots of conformational parameters, Figure 5-10.

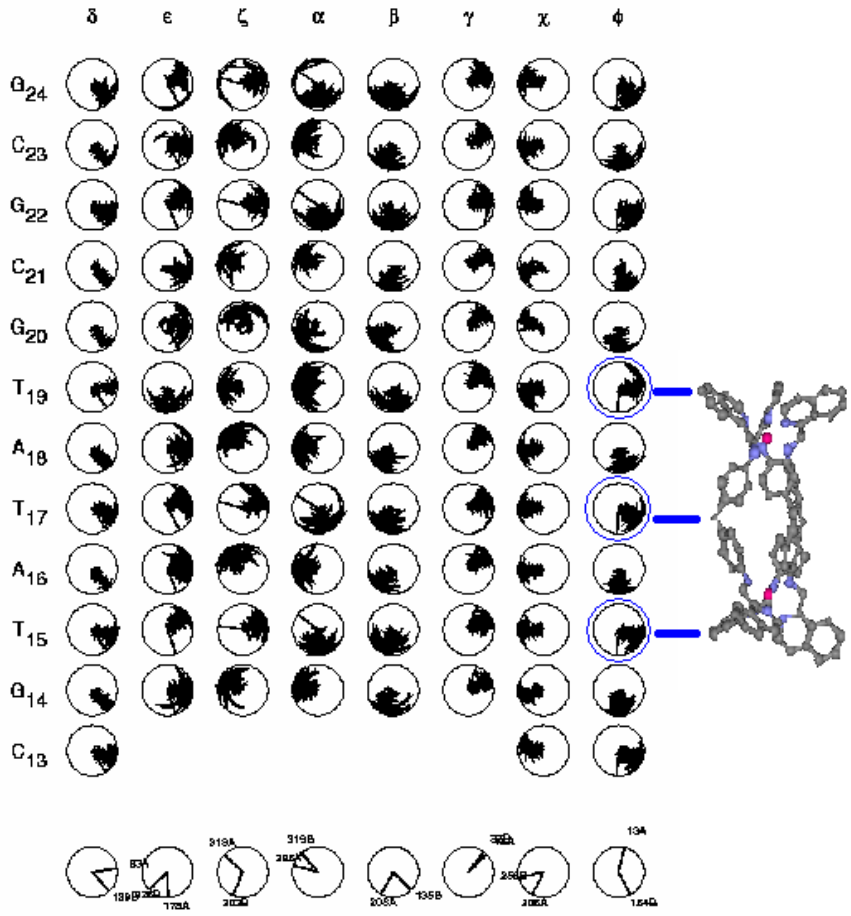


Figure 5-10 Dials for 0 – 0.25ns displaying conformational changes relating to the position of the cylinder relative to the DNA.

5.2.3 Second nanosecond

The helicoidal parameters for the second nanosecond are shown in Figure 5-11. These parameters showed similar trends to those calculated over the first nanosecond. However the ordered disruption observed in the bottom portion of the DNA local to the cylinder is less obvious as the cylinder has tumbled out of the groove by this stage and is bound along the backbone of the DNA spanning the four base pairs T_9-A_{16} to A_6-T_{19}, Figure 5-4 (c). The twist in this region is reduced as the template of the cylinder constrains the DNA to fluctuate about a mean value of 30^o which is more indicative of canonical A-DNA than B-DNA. The variance in slide values is characteristic of B-DNA with the majority being greater than -0.8Å.

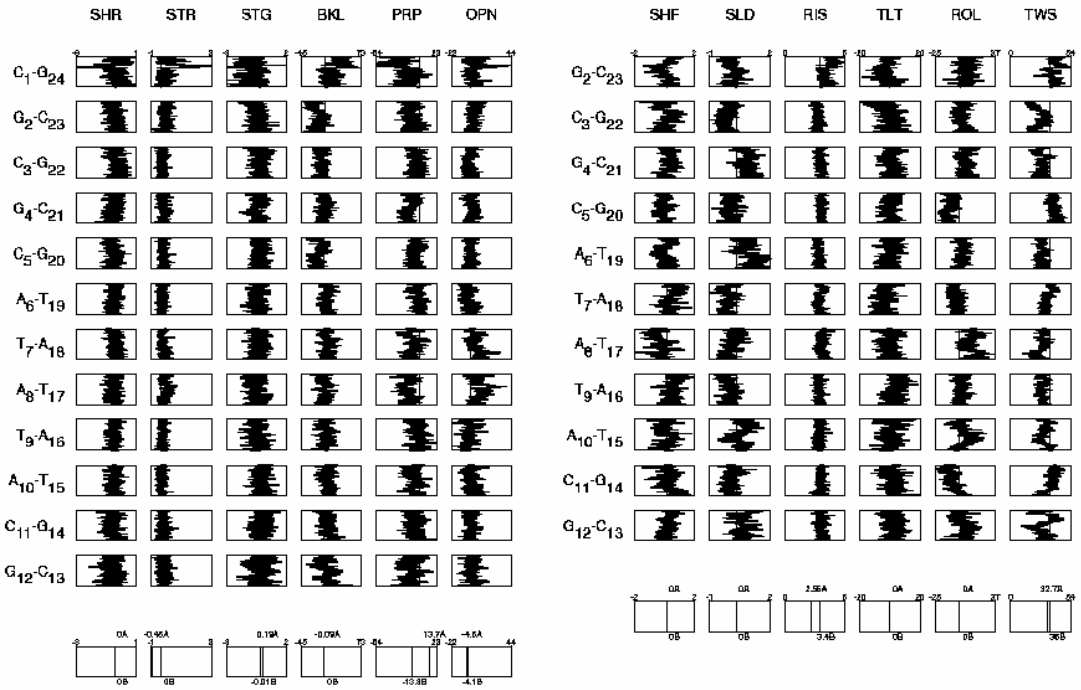


Figure 5-11 Helicoidal parameters for 1.75 – 2.00ns of C_{Fe}^{4+} d(CGCGCATATACG)$_2$

5.2.4 Third nanosecond

The helicoidal parameters calculated over the third nanosecond as shown in Figure 5-12. Due to the separation of the terminal C_1-G_{24} base pair the other features become less clear as the scale is adjusted during plotting to accommodate the large fluctuations. The fraying behaviour at the end is not in common with the reference system and is not mirrored between G_{12}-C_{13} at the bottom end of the DNA strand, although these bases do show less ordered behaviour than those more central to the stand.

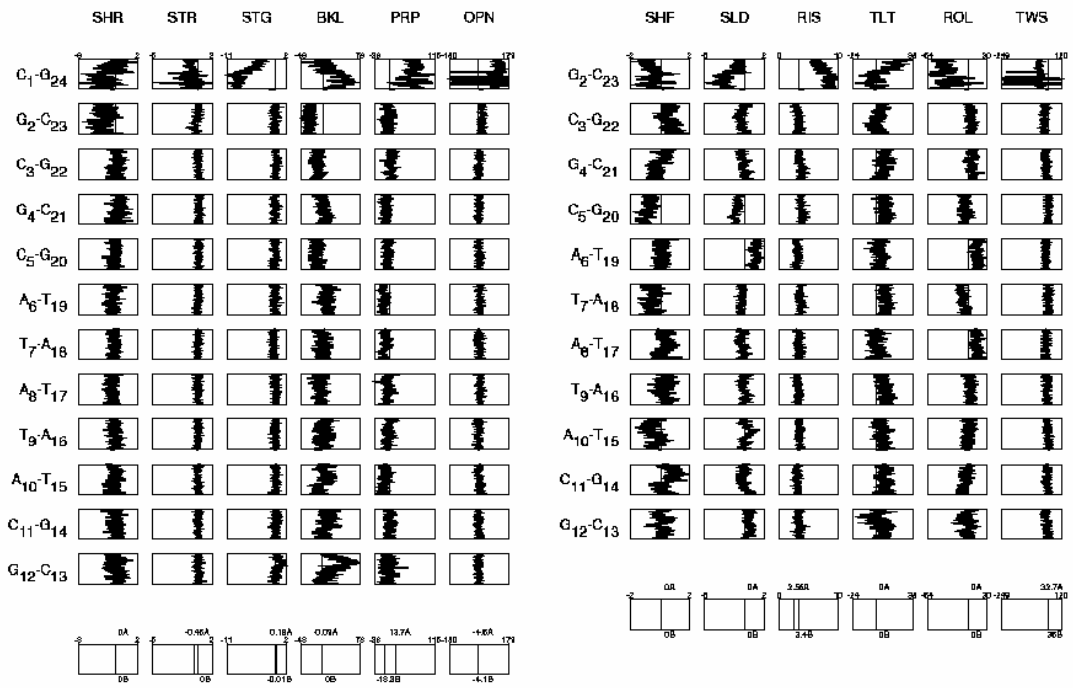


Figure 5-12 Helicoidal parameters for 2.75 – 3.00ns of C_{Fe}^{4+} d(CGCGCATATACG)$_2$

At this stage in the simulation the cylinder is located along the spine utilising the propeller shaped ends of the helix to bind with the DNA, see Figure 5-13. Binding in this fashion leaves one end of the cylinder to be proud from the DNA with contact being made to the DNA with one of the helix propeller blades at G_4-C_{21}.

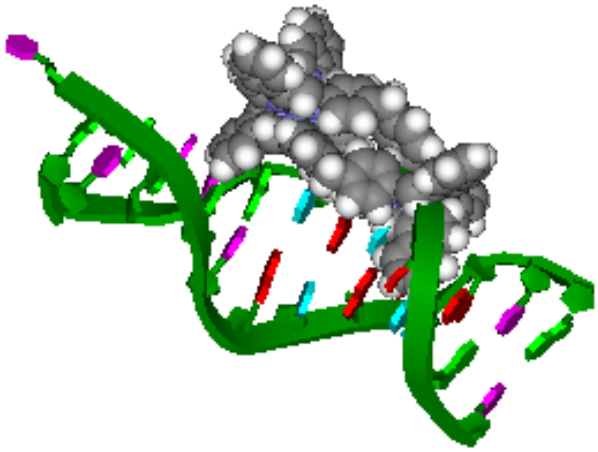


Figure 5-13 C_{Fe}^{4+} binding to d(CGCGCATATACG)$_2$ after 3ns of simulation. See Figure 5-4 (d) for a different view.

Similar to the observation made earlier the cylinder binding along the spine appears to introduce structural changes between the base pairs with which closest contact is made. This causes the parameters of alternating base pairs to show variation from the reference system as the cylinder spans five bases in total.

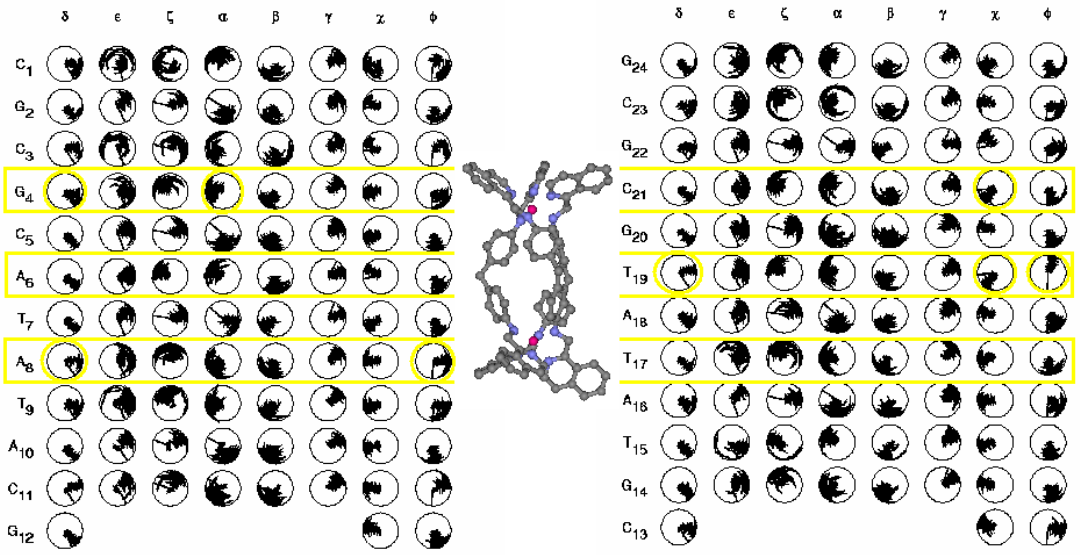


Figure 5-14 Conformational parameters for 2.75 – 3.00ns of C_{Fe}^{4+} d(CGCGCATATACG)$_2$. The three yellow bands show the parameters potentially affected by the points of close contact, parameters showing a value different to the reference system and their closest neighbours are highlighted with a yellow circle.

5.3 C_{Cu}^{2+} system

5.3.1 Overview

A detailed study of the interaction of the C_{Fe}^{4+} cylinder with d(CGCGCATATACG)$_2$ was described in the previous section. In this section the discussion is extended to consider the effect of the charge and shape of the cylinder on its DNA-binding properties and on the response of the DNA. The method of performing docking calculations to identify favourable binding sites followed by fully solvated simulations was followed in the same way as for the simulations discussed earlier in this chapter. This simulation is more extensive than that of C_{Fe}^{4+} with 5ns of simulation having been calculated.

To investigate the significance of charge and shape of the C_{Cu}^{2+} cylinder relative to C_{Fe}^{4+} a low energy major groove configuration was extracted from the docking calculation, see Figure 5-3. While minor groove sites were sampled during the docking calculation the majority of time was spent sampling potential major groove sites. The starting configuration was comparable to that of C_{Fe}^{4+} with the cylinder lying symmetrically along the major groove, Figure 5-15 (a). The low energy starting configuration placed C_{Cu}^{2+} so that it spans base pairs $T_8 - A_{17}$ to C_5-G_{20}. By the end of the simulation the cylinder was bound to the DNA backbone of bases A_{16} to T_{19}, Figure 5-15 (f).

The average DNA structures calculated for the last 25 picoseconds of each nanosecond were extracted with the cylinder present and are presented in Figure 5-15. While the figure for each portion of the simulation in Figure 5-15 has been orientated to enable the reader to see the cylinder it does make a visual comparison of changes in the DNA structure difficult. For this reason a composite schematic diagram of the six snap shots was put together to aid qualitative analysis while showing the migration of C_{Cu}^{2+} from deep in the major groove into the minor groove and then to the backbone binding mode.

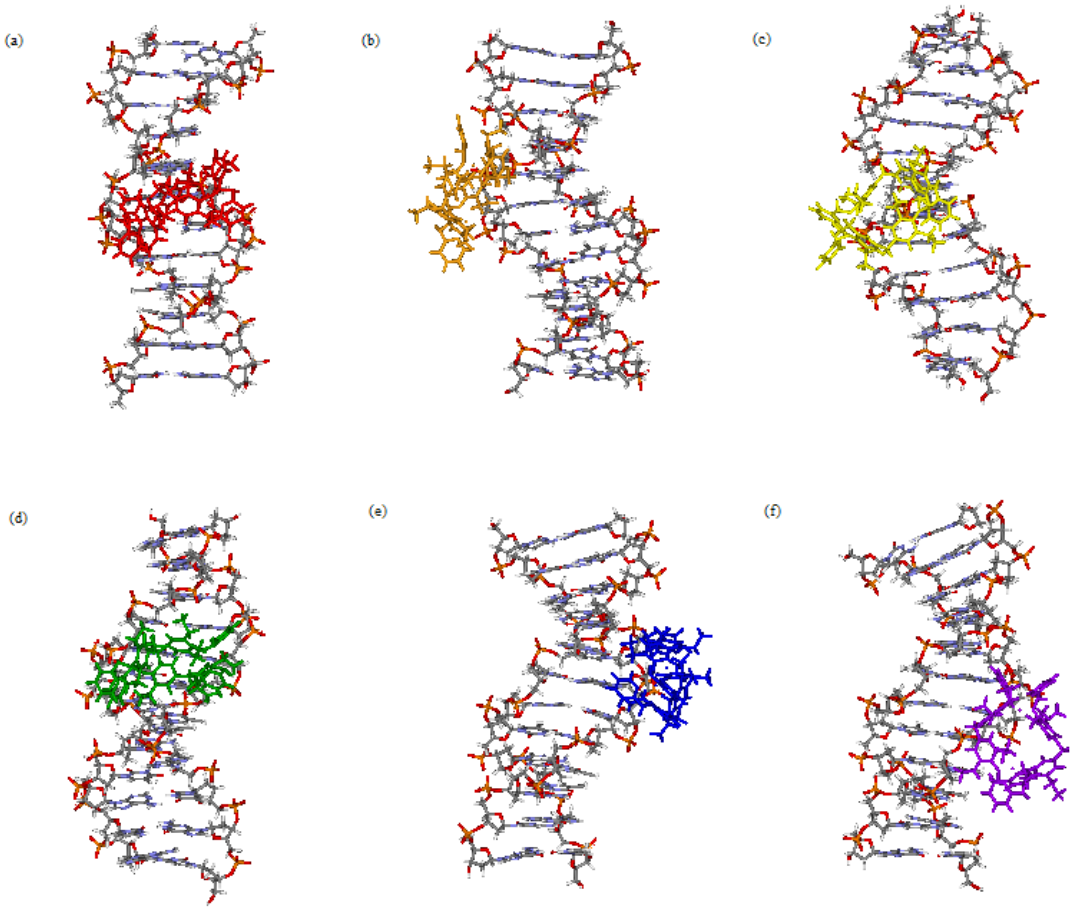


Figure 5-15 The migration of CCu2+ from an initial binding site in the major groove to a position on the DNA spine. (a) starting configuration, red (b) 1ns , orange (c) 2ns, yellow (d) 3ns, green (e) 4ns, blue (f) 5ns, indigo.

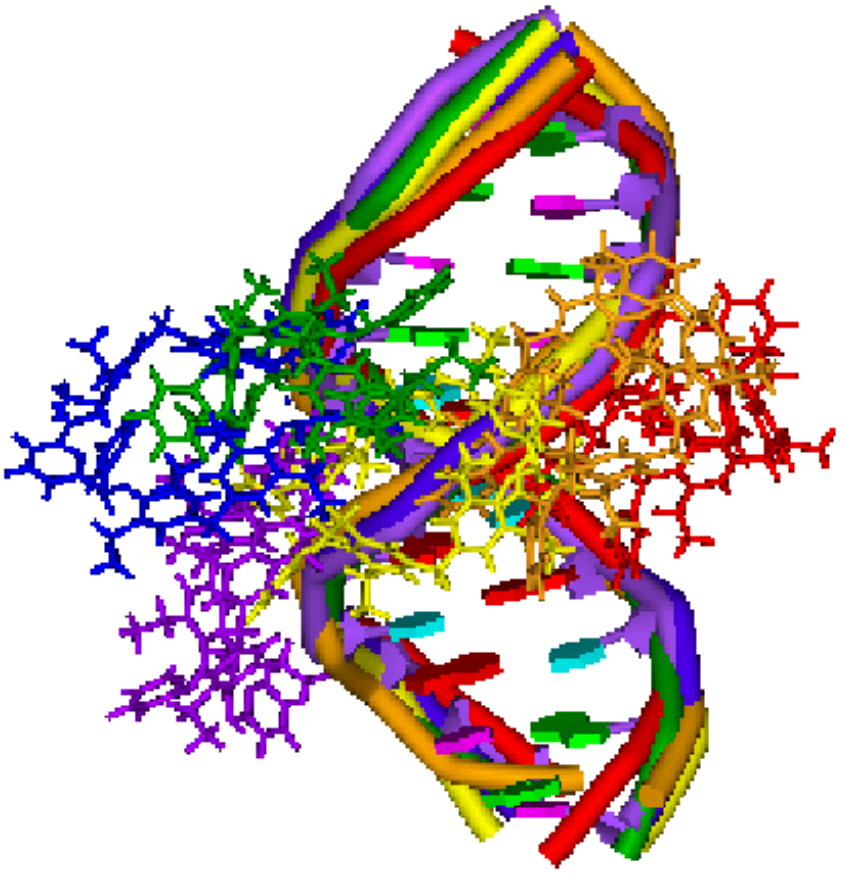


Figure 5-16 A schematic diagram of the extracts superimposed to show the observed similarities of the DNA over 5ns of simulation and the relative motion of the cylinder.

5.3.2 Analysis of results

Helicoidal parameters were calculated for each 0.25ns portion of simulation. The values for the final 0.25ns of each 1ns stage of simulation are presented in Figure 5-17, Figure 5-18 and Figure 5-19.

The helicoidal parameters calculated for the duration of the simulation indicate a stable DNA structure that is generally displaying the characteristics of canonical B-DNA up to 5ns. While there are occasions during the simulation when the terminal base pairs demonstrate greater mobility there are no values to indicate that the base pairs separate. The variations take place

within the tolerances of the DNA flexibility and can be seen to be comparable to central base pairs even at 5ns.

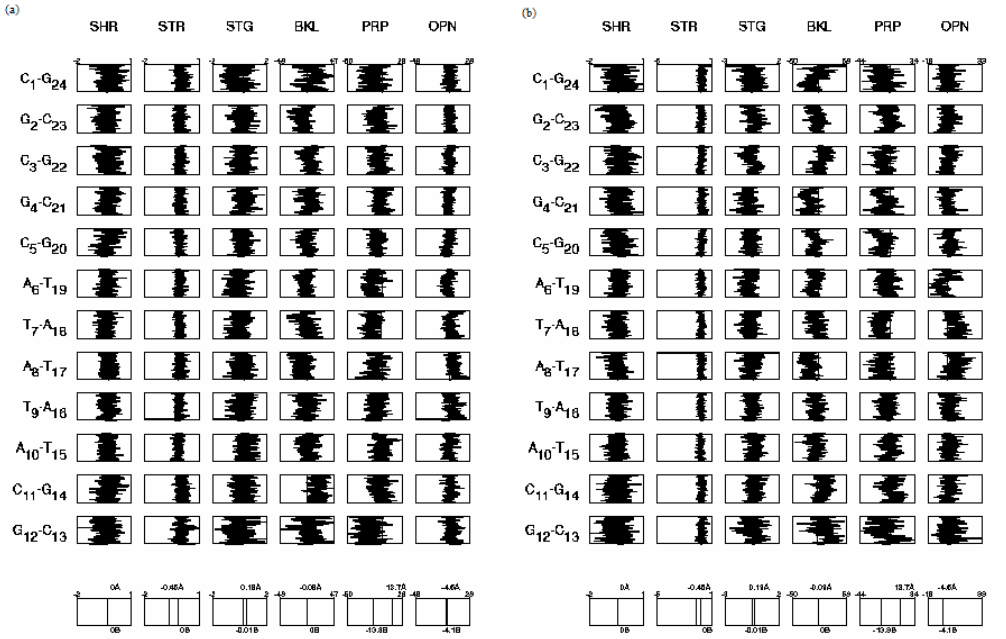


Figure 5-17 Helicoidal parameters calculated for a) 0.75 – 1.00ns and b) 1.75 – 2.00ns.

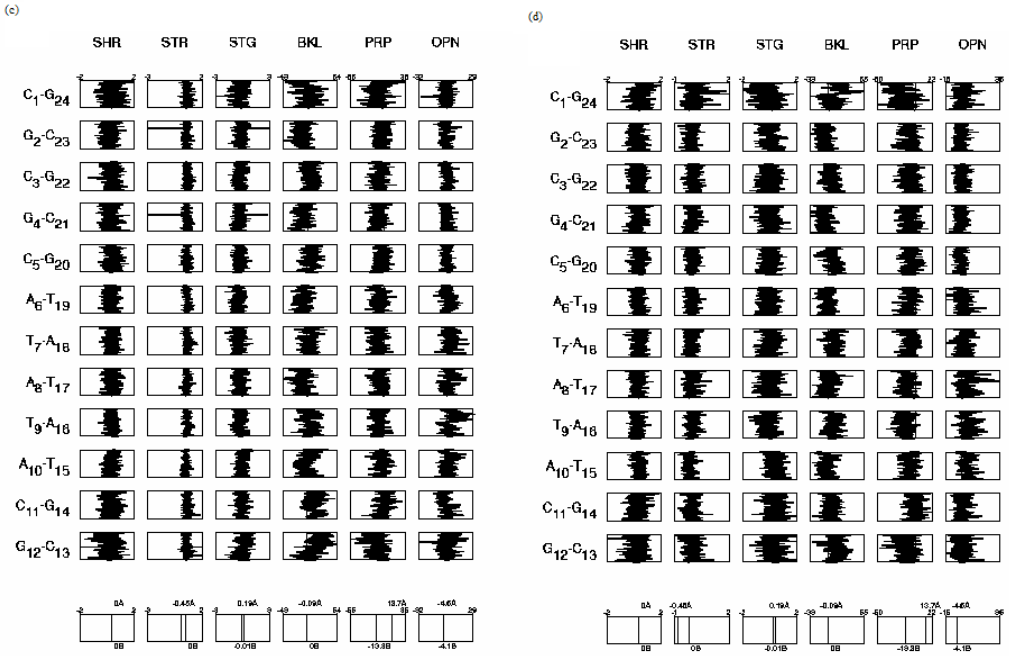


Figure 5-18 Helicoidal parameters calculated for c) 2.75 – 3.00ns and d) 3.75 – 4.00ns.

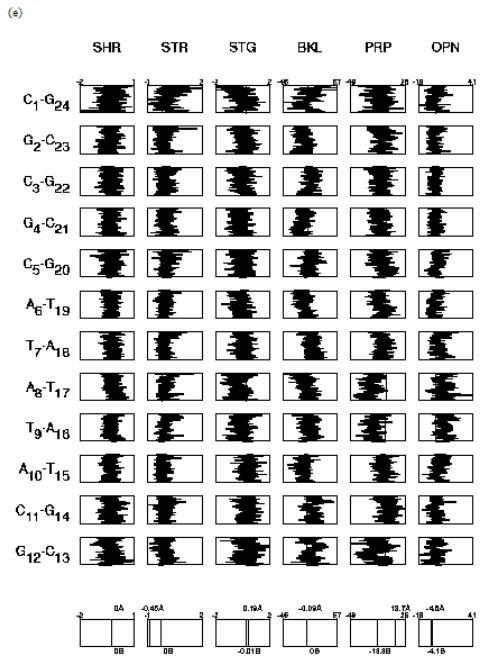


Figure 5-19 Helicoidal parameters calculated for e) 4.75 – 5.00ns

Calculating and plotting the distribution of inter-base pair parameters shows the DNA sequence to develop an interesting feature during the course of the simulation. Although it is difficult to correlate the reduced twist values for alternating base pairs to the presence of C_{Cu}^{2+} the pattern can be seen consistently throughout the strand by 4ns, see Figure 5-20 and Figure 5-21, and is not present to the same extent in either the Reference or C_{Fe}^{4+} simulations.

Similar behaviour is present to a much lesser extent in the CFe^{4+} system but with the variation being local to the cylinder this has been attributed to an effect induced by the cylinder.

The values of roll, twist and slide confirm conformational distinction and have been used to define structure type since early DNA studies[281]. The transformation from B- to A-DNA tends to decrease twist, increase roll and reduce slide. Values of slide below -0.8 Å are typical of A-DNA dimmer steps and those greater than 0.8 Å are found in the majority of B-forms. The values for these parameters during the simulation, as highlighted in Figure 5-21, are inline with

the values for A-DNA. However, the alternating values for the helical parameter δ, lie within the angle range for the ribose ring conformation of O4'-endo. The most prominent conformation is C2'-endo in B-DNA and C3'-endo in A-DNA with angle values of 144^{o} - 180^{o} and 0^{o} – 36^{o} respectively. Both C2'-endo and C3'-endo exist in equilibrium at room temperature with transitions occurring via the 04'-endo state. If the sugar ring is present in an intermediate state it would be interesting to develop this simulation to observe its stability or possibly the ability of C_{Cu}^{2+} to stabilise it or cause the DNA to change conformation.

While steps C_3-C_{22}, C_5-C_{20}, T_7-A_{18}, T_9-A_{16} and C_{11}-G_{14} have undergone conformational change within the sugar rings the neighbouring base-pairs have maintained their structure and continue to be defined as representing B-DNA. The twist values for a number of the B like steps are greater than the mean value for B-DNA. The twist and related parameters of the bases that remain in the C2'-endo form, 144^{o} – 180^{o}, compensate for the reduced twist and so "kinks" are introduced into the backbone of the DNA retaining an overall B-Form for the DNA structure as a whole.

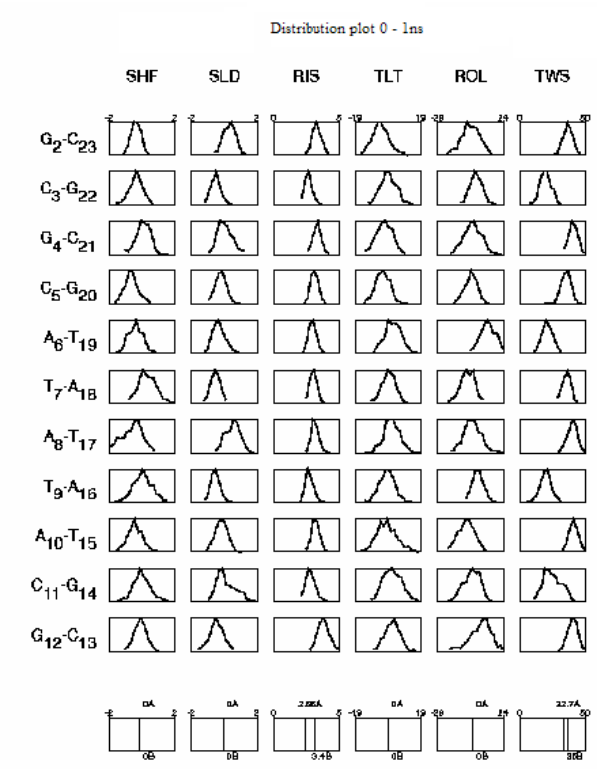


Figure 5-20 Distribution plots of inter base pair parameters for the first nanosecond of simulation.

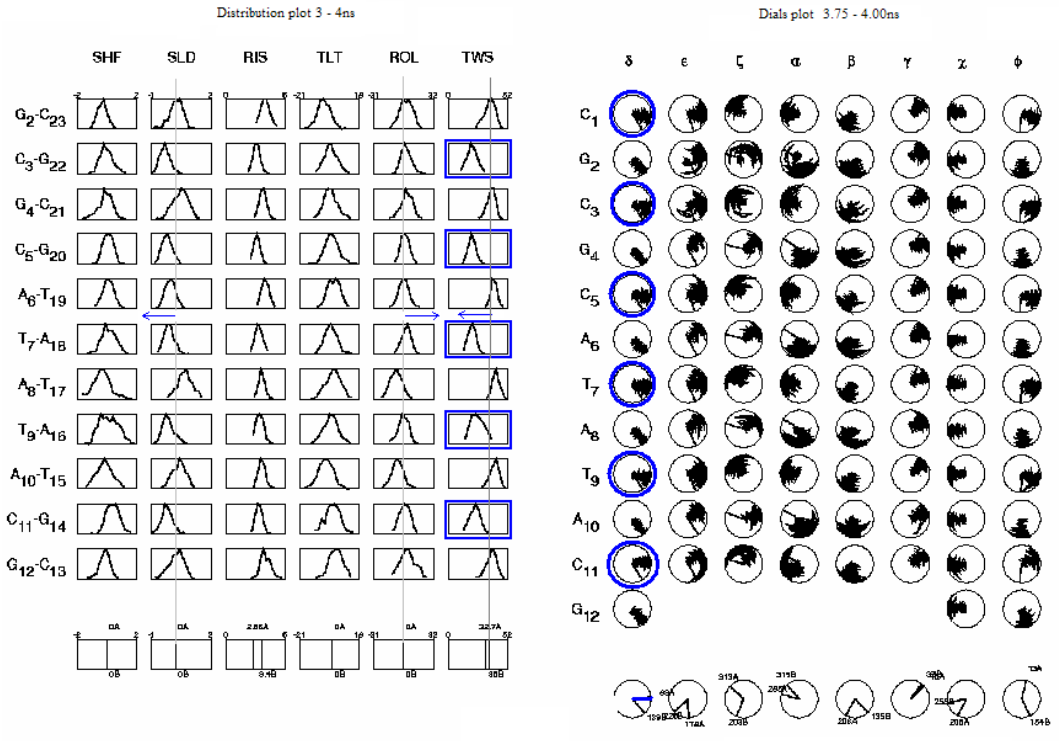


Figure 5-21 Left, distribution of inter base pair parameters calculated for 3 – 4ns. Arrows indicate the relationship between decreased Twist, increased Roll and reduced Slide. Right, helical parameters for the 3.75 – 4.00ns portion of simulation. The highlighted dials show alternating base pairs showing δ values indicative of canonical A-DNA.

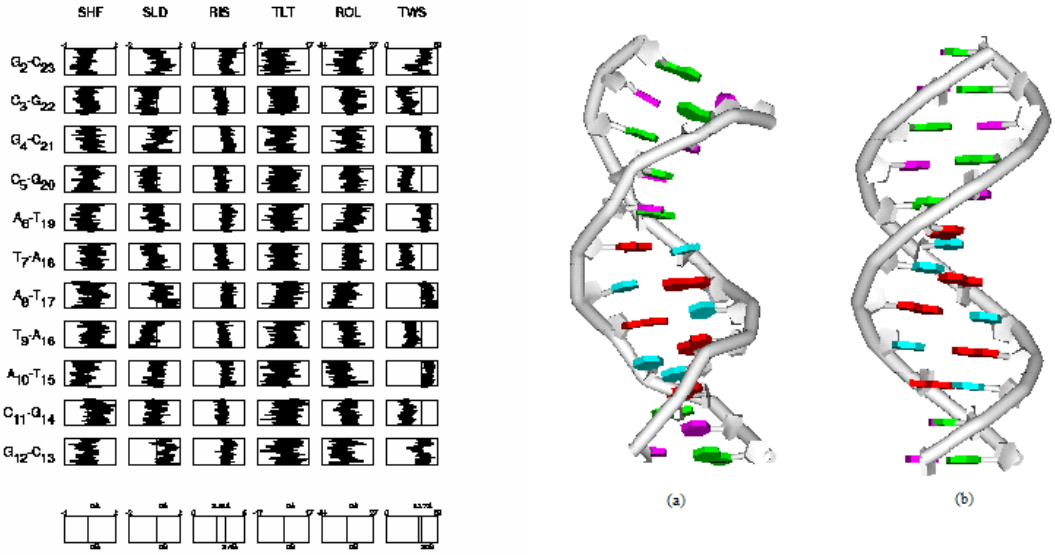


Figure 5-22 Window plots of helicoidal parameters calculated for 4.75 – 5.00ns, the associated schematic diagram of the molecular structure with "kinks" in the backbone (a) and compared with the smooth backbone of the reference system (b).

5.4 C_{Fe}^{4+} d(ATATATATATAT)$_2$ system

As the simulations of the bound DNA progressed the reference systems were extended for the uncomplexed DNA systems. The distribution plots in Figure 5-23 show the DNA complex to be stable and of the B-form throughout 5ns of simulation.

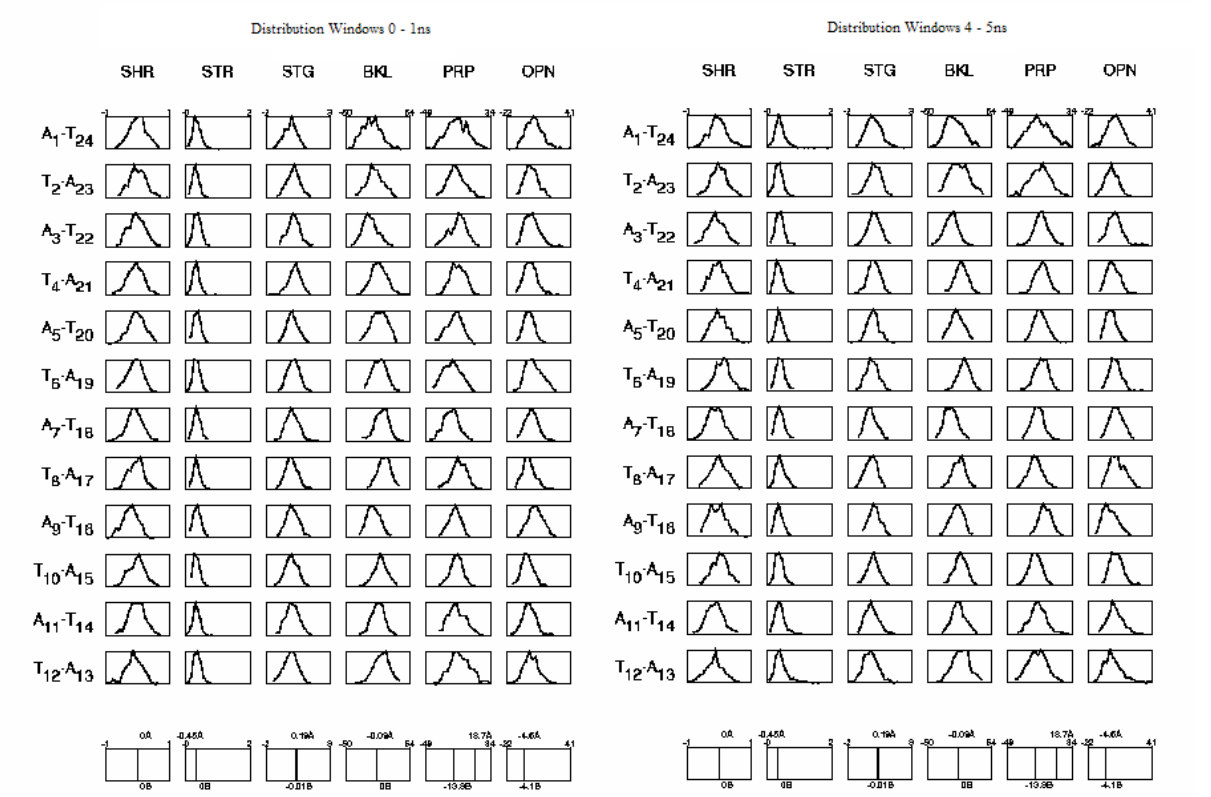


Figure 5-23 Distribution of conformational parameters between the start of the uncomplexed $d(ATATATATATAT)_2$ simulation at 0.75 – 1.00ns (left) and 4.75 – 5.00ns (right).

A detailed study of $d(ATATATATATAT)_2$ was performed to extend the discussion on charge and shape to include variation in shape of the DNA as well as just the cylinder. While $d(CGCGCATATACG)_2$ has two distinctive regions simulating sequences of one base pair only will increase our understanding about the role of the DNA sequence in the context of binding supramolecular cylinders of different charges and shapes.

Docking calculations predicted major groove binding for C_{Fe}^{4+} bound to $d(ATATATATATAT)_2$. Figure 5-24 (a) shows the conformation that was extracted and solvated with the lowest configurational energy. The cylinder spans A_9-T_{16} to A_5-T_{20} while

being located in the major groove before rotating and migrating so that the end of the helicate is bound in the minor grove and the cylinder spans T_4-A_{21} to A_7-T_{18}, Figure 5-24 (e).

During the simulation and the process of the cylinder migrating out of the major groove the DNA undergoes a number of conformational changes resulting in it adopting a structure more like the A-Form.

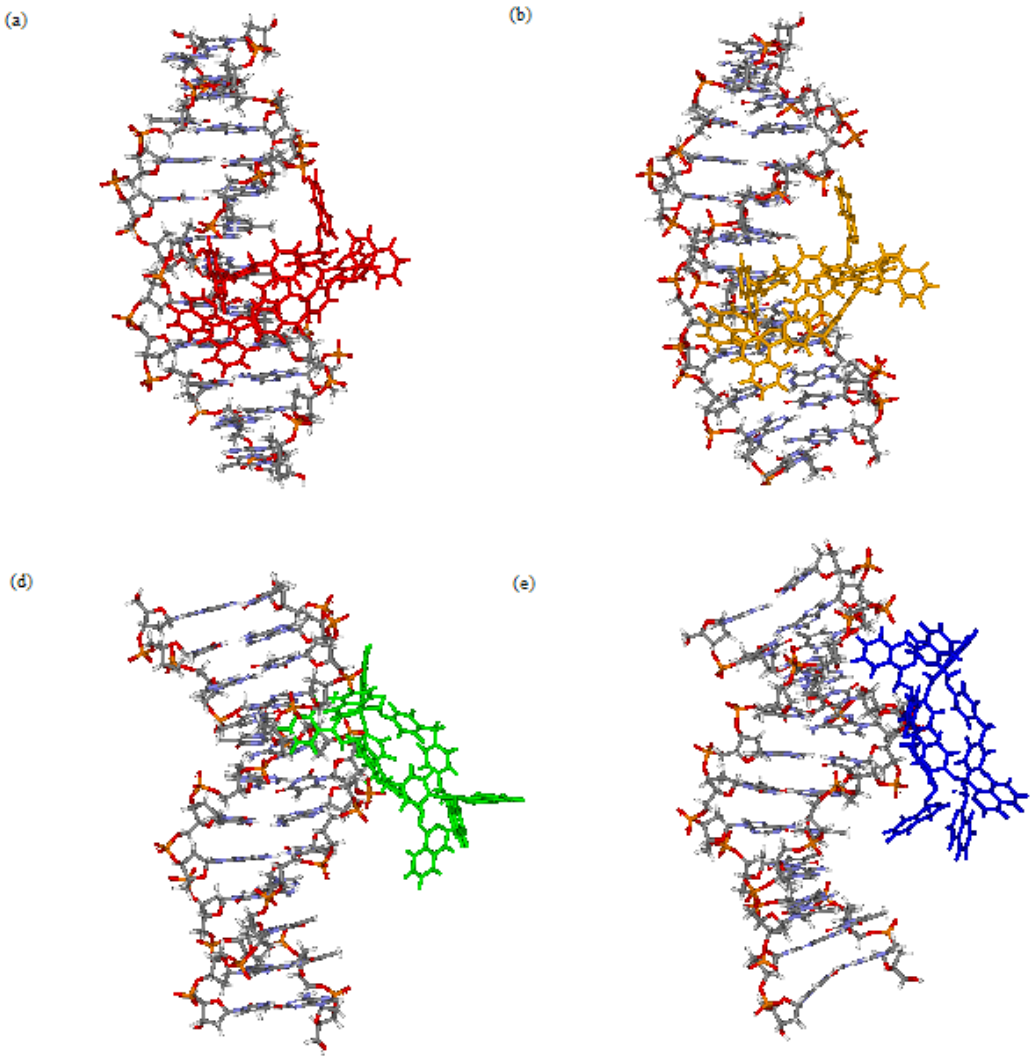


Figure 5-24 The migration of C_{Fe}^{4+} from an initial binding site in the major groove to a binding position with the propeller of the helicate in the minor groove. (a) starting configuration, red (b) 1ns , orange (c) not available due to with data transfer error (d) 3ns, green (e) 4ns, blue.

Comparison of the superimposed structures in Figure 5-25 shows the DNA coil to be slackening as the simulation progresses with no change to base pairings or fraying of the ends.

The helicoidal parameters calculated show variations in the δ and φ parameters within the first 0.5ns which are indicative of sugar pucker and a transition to A-DNA. Initially some of these changes are short lived, less than 0.25ns, but become more abundant as the simulation progresses spending more time in the changed state.

The changes seen in the inter-base pair parameters during this simulation may help to explain the observations made in the C_{Cu}^{2+} - $d(CGCGCATATACG)_2$ simulation particularly as the transitions of twist, roll and slide parameters were initially seen more in the AT region of the DNA, see Figure 5-20.

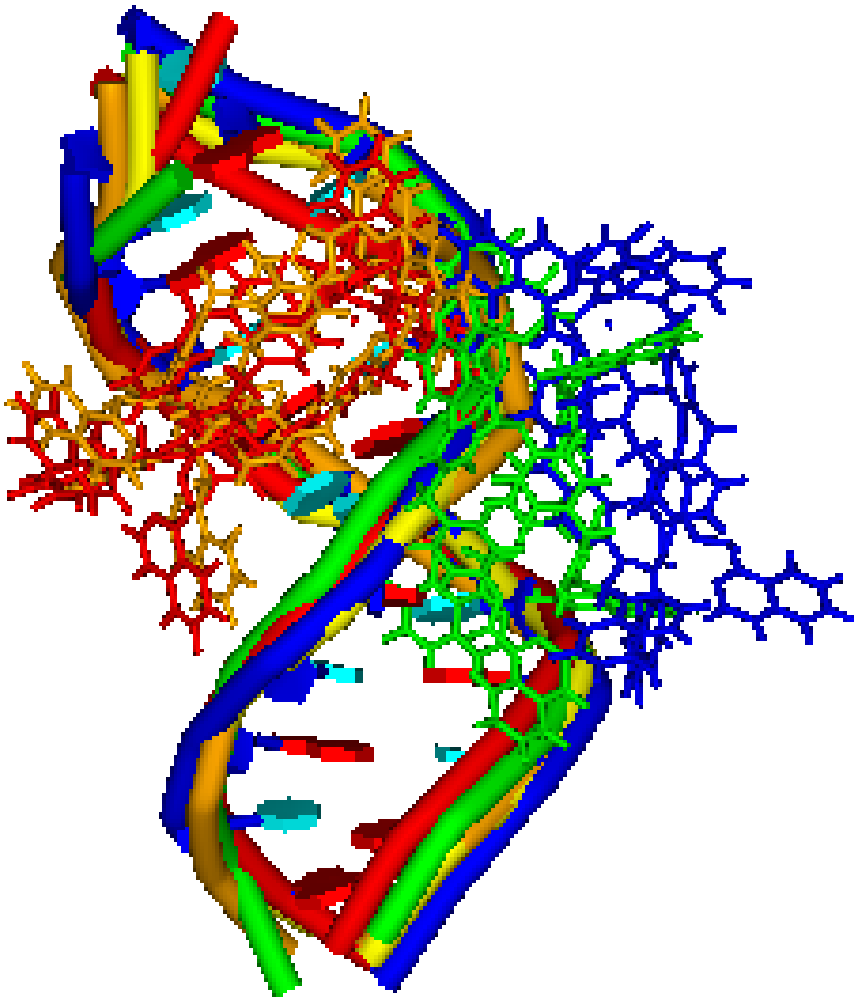


Figure 5-25 A schematic diagram of the five extracts superimposed to show the observed similarities of the DNA over 4ns of simulation and the relative motion of the cylinder.

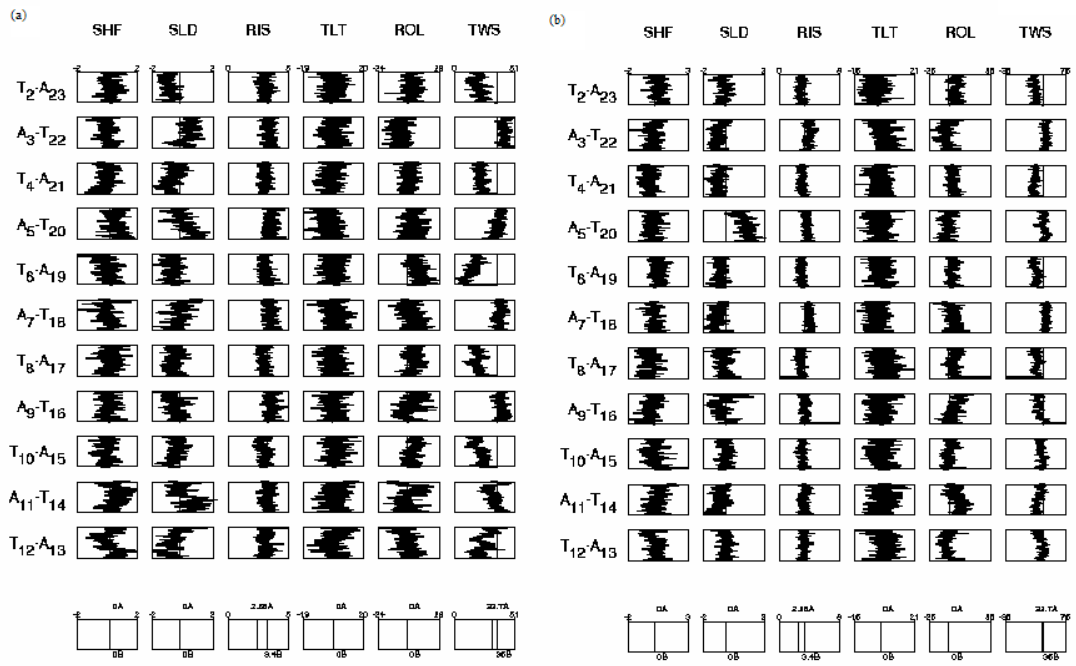


Figure 5-26 Inter-base pair parameters calculated for (a) 1.50 – 1.75ns and (b) 3.75 – 4.00ns to show the transition from B-DNA to A-DNA.

By 1.5ns the twist values in Figure 5-26 can be seen to alternate in a way that is comparable to Figure 5-21 for the C_{Fe}^{4+} - $d(CGCGCATATACG)_2$ system.

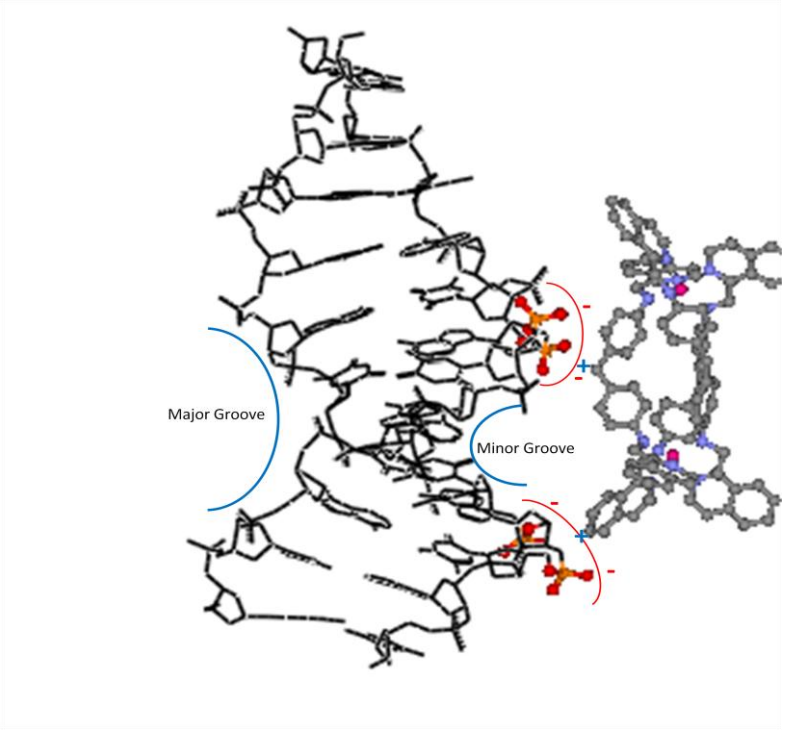


Figure 5-27 C_{Fe}^{4+}-$d(ATATATATATAT)_2$ after 4ns and migration of C_{Fe}^{4+} from major to minor groove.

However, the difference for this system is that by 4ns, Figure 5-26, values for eight of the eleven base pair steps have changed sufficiently to no longer be classified as B-DNA with the related roll parameters having increased and those for slide having reduced. At this stage in the simulation the Minor groove is also calculated as having widened while the Major groove has narrowed.

During this analysis no evidence was found that either cylinder disrupted the base pairings at any stage during the simulations. Evidence of the preserved pairings is confirmed with at least one of the hydrogen bond lengths being less than 3.0 Å in every base pair, while only the end pair had a bond length greater than 3.15Å. This is a major difference from the CHARMM22 reference system and previous simulations[37] carried out with the same C_{Fe}^{4+} where the +4 *e* cylinder charge induced base-pair mismatches within just 0.25ns. Given the strongly ionic character of DNA and its flexibility this apparent invariance to the ligand charge is unexpected. One property that does show differences between the two force fields is the groove widths. Relative to CHARMM22 the major groove of CHARMM27-DNA contracts in the presence of C_{Fe}^{4+}. For this reason we note that CHARMM27 uses a more strongly charged DNA backbone than does CHARMM22, with the PO_4 net charge being -1.2 *e*, compared with -1.0 *e* for CHARMM22. This increased negative charge may lead the DNA backbone to contract more strongly onto the tetracationic cylinder which may be able to act as a template for the minor groove, resulting in stiffening and straightening of the DNA backbone and thereby inhibiting the subsequent bending of the DNA. A similar response to ligand charge has been seen with netropsin bound in the minor groove,[282] but greater scope for DNA relaxation might have been anticipated when the ligand is initially docked and bound in the major groove. Instead the tetracationic ligand unable to induce base pair mismatches favours sampling of the more negative DNA backbone and any increased solvent structuring that this may be responsible for.

Comparison with the other systems simulated enables us to understand the force field dependence.

If this is the characteristics of a mechanism by which the DNA changes form it could be possible that the earlier C_{Cu}^{2+} - d(CGCGCATATA)$_2$ system is in some form of intermediate which, with the right conditions may be seen to change form if the simulation was extended. The C_{Cu}^{2+} could be contributing to this transition in an otherwise stable system by acting as a spacer in the minor groove making it to become more like that of the A-Form.

5.5 C_{Cu}^{2+} d(ATATATATATAT)$_2$ system

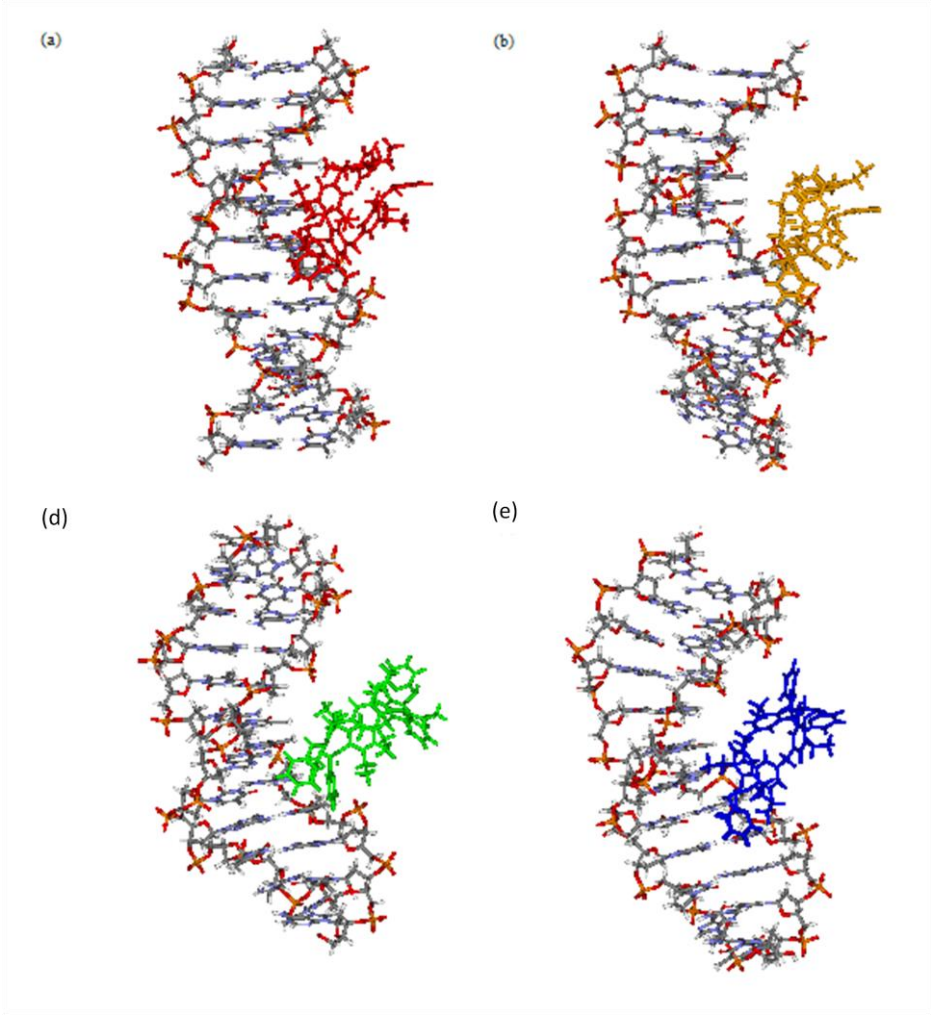


Figure 5-28 The configurations of C_{Cu}^{2+} from an initial binding site in the major goove of d(ATATATATATAT)$_2$ (a) starting configuration, red (b) 1ns , orange (c) not available due to with data transfer error (d) 3ns, green (e) 4ns.

Docking calculations predicted major groove binding for C_{Cu}^{2+} bound to $d(ATATATATATAT)_2$. Figure 5-28(a) shows the conformation that was extracted and solvated. As can be seen from the time evolution of the system Figure 5-28(a)-(d) the cylinder resides in the major groove during the 4ns of simulation. When considering the cylinder behavior this is rather different from C_{Fe}^{4+} which appeared to be quite mobile over the same period and in the presence of the same DNA sequence. While it would be interesting to extend the simulation of C_{Fe}^{4+} to see if it will return to the major groove or find a favorable binding site in the minor groove it is apparent that C_{Cu}^{2+} is less able to leave the major groove and sample the DNA.

The helicoidal parameters calculated initially appear to show random fluctuations with a major event between T_{12}-A_{13} at about 1.5ns before stabilizing with an element of structure added to the upper half of the strand. The terminal base pair T_{12}-A_{13} appears to fluctuate about the average values for twist, roll and slide before undergoing a step change from a value near to 30^o to one nearer -25^o as associated with C2'-exo sugar conformation. This was a rapid change at a time when the cylinder was not local to this end of the strand. This behavior was not seen in the C_{Fe}^{4+} system or the references system. As can be seen in Figure 5-29 the bend at the bottom of the orange strand returns to a natural position for the remainder of the simulation, having only stayed in this conformation for about 0.25ns. During the remainder of the simulation, 1.5 – 4.0ns evidence of structure forming is seen throughout the top seven base pairs. Base pairs T_6-A_{19} up to A_1-T_{24} once again demonstrate an alternating trend of twist values being defined as A-DNA and B-DNA and related changes to the correlated roll and slide values. In these final stages of the simulation the DNA appears to be accommodating the cylinder with a feature in the conformation of a suitable size and orientation to complement the cylinder, see Figure 5-3

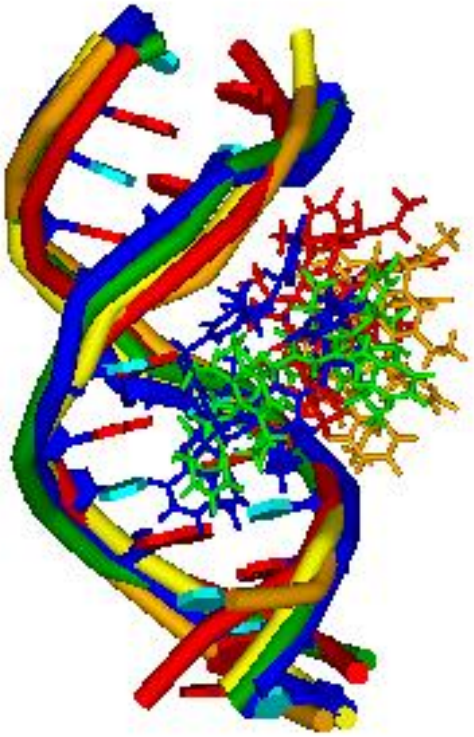


Figure 5-29 5-30 A schematic diagram of the five extracts superimposed to show the observed similarities of the DNA over 4ns of simulation and the relative motion of the cylinder

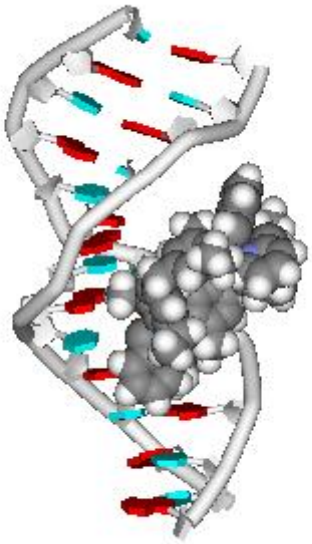


Figure 5-31 d(ATATATATATAT)$_2$ kinking to accommodate C_{Cu}^{2+}

5.6 C_{Cu}^{2+} d(CGCGCGCGCGCG)$_2$ system

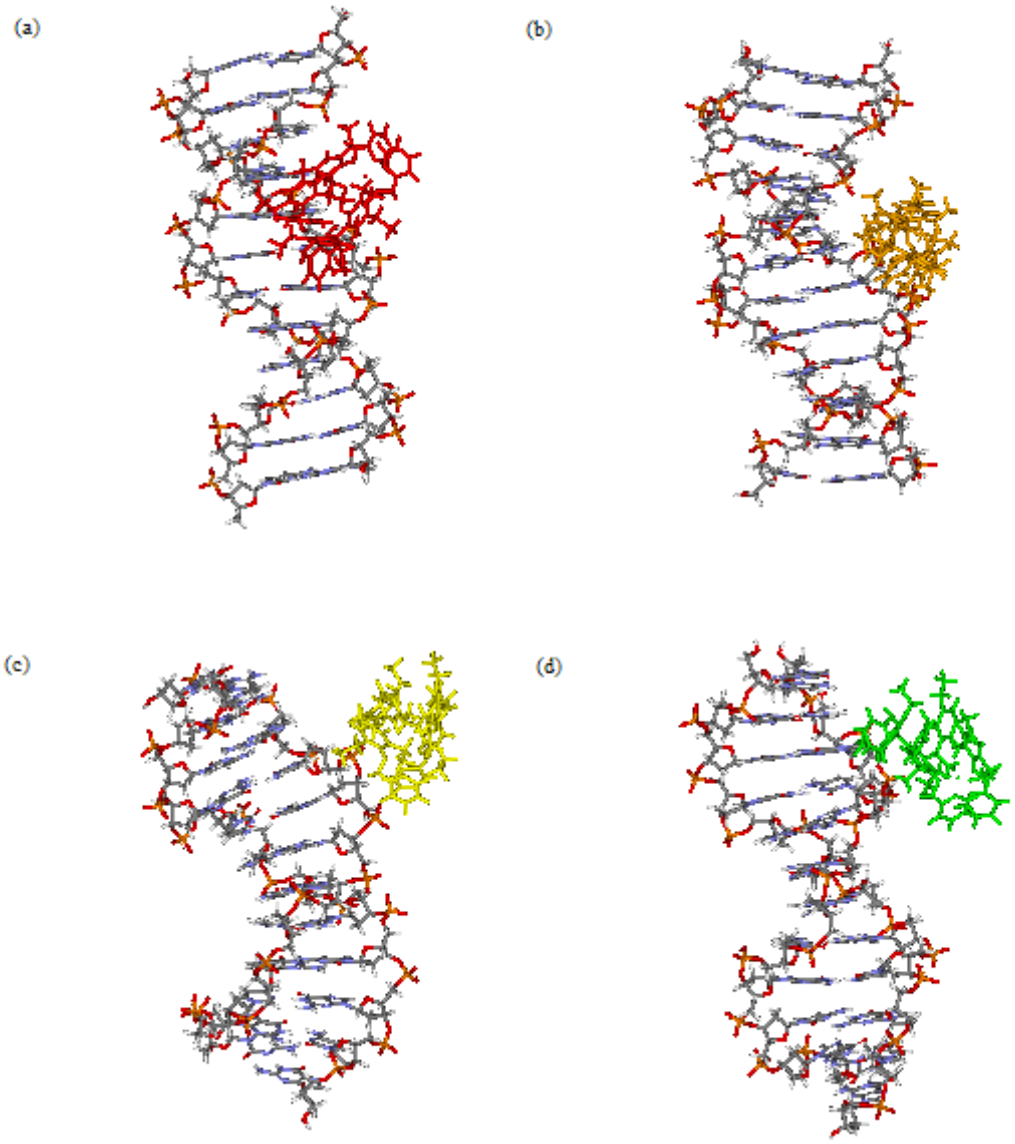


Figure 5-32 The configurations of C_{Cu}^{2+} from an initial binding site in the major groove of d(CGCGCGCGCGCG)$_2$ (a) starting configuration, red (b) 1ns , orange (c) 2ns, yellow(d) 3ns, green.

By taking the lowest energy configuration from the docking calculation Figure 5-32(a) was extracted and solvated as a starting configuration with the cylinder docked in the major groove. During this three nanosecond simulation of C_{Cu}^{2+} with d(CGCGCGCGCGCG)$_2$ the cylinder

was seen to be mobile resulting in a configuration where the cylinder is spanning the backbone of the DNA. Throughout the simulation all base pairs in the entire strand, including the terminal pairs, maintain their partnerships. The helical parameters show apparently random fluctuations in values as a number of changes that do take place are short lived. Figure 5-33 is an extract taken from the dials plots for this simulation. As a series of plots it helps to show the timescales over which the random changes occurred. For the base C21 the values for the helical parameter φ is seen to be defined as A-DNA at 3.00 – 3.25ns before flipping to the B-DNA value towards the end of the 3.25 – 3.50ns segment of the simulation. In the portion 3.50 – 3.75ns there is no indication of its history as the configuration is stable in the B-DNA form.

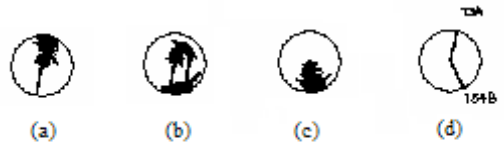


Figure 5-33 Extracts taken from dials at (a) 3.00 – 3.25ns (b) 3.25 – 3.50ns and (c) 3.50 - 3.75ns. (d) is the reference values for canonical A-DNA (13°) and canonical B-DNA (154°).

A number of the occasions where changes such as that described have taken place could be explained as being a result of the localized effects induced by the cylinder as it has migrated during the simulation. From the current final stage of this simulation it is unclear if the cylinder is sampling the DNA for a more appropriate docking site where it will settle or if it will return to the major groove. It is clear from the helicoidal parameters that to this point no large features as a result of the cylinder have developed. This could be expected as the C·G base pairs are stronger than those of the A·T base pairs so reducing the ability of the less charged C_{Cu}^{2+} to effect the DNA conformation with the CHARMM27 force-field. These simulations will be compared to the C_{Fe}^{4+} with $d(CGCGCGCGCGCG)_2$ simulations to give insight into the effect of charge on the observable characteristics, Section 5.7.

5.7 C_{Fe}^{4+} d(CGCGCGCGCGCG)$_2$ system

The starting conformation shows the cylinder lying symmetrically along the major groove spanning base pairs G10-C15 to G6-C19, Figure 5-34 (a). During this 4ns simulation there were no prolonged periods of base pair mismatching as expected due to the relatively high level of hydrogen bonding that contributes to the stable nature of the d(CGCGCGCGCGCG)$_2$ DNA sequence.

During the course of this simulation the cylinder was seen to reside for the most time in the major groove before indicating that localised binding between the propeller ends of the cylinder and DNA spine could be the beginning of a slow migration from the major groove. A cartoon of this behaviour can be seen in Figure 5-34.

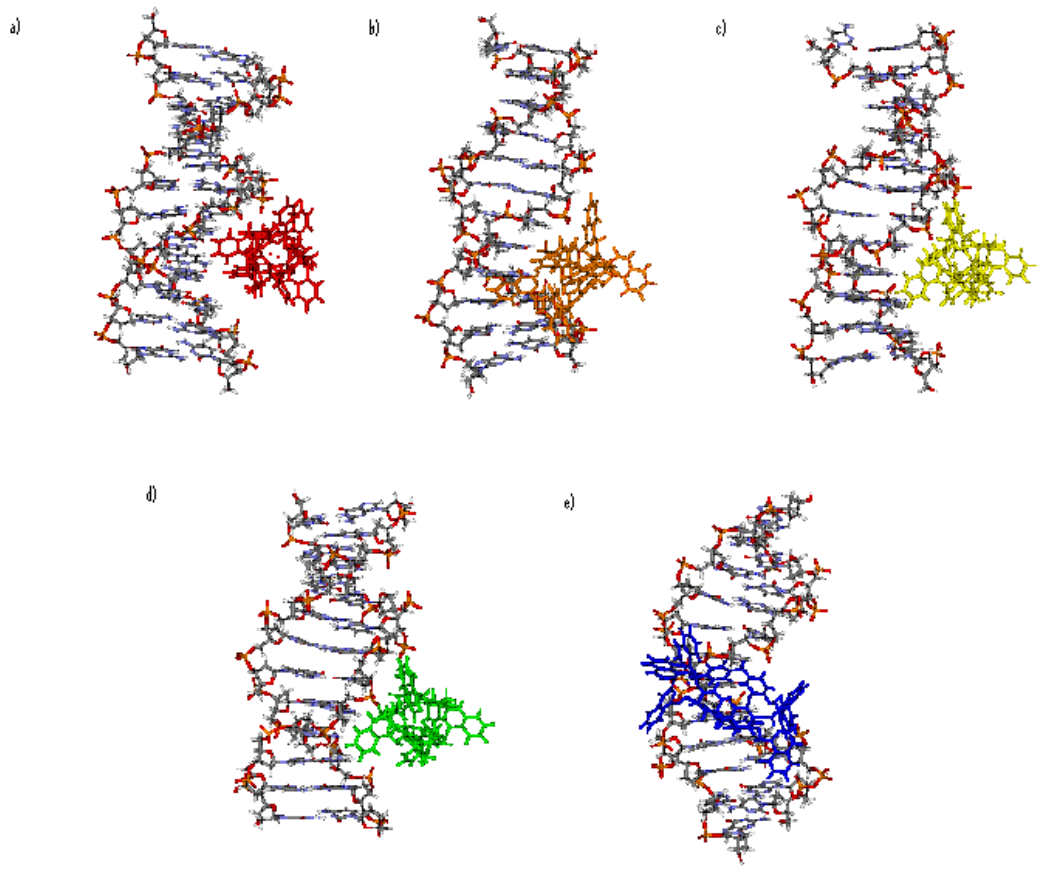


Figure 5-34 The migration of C_{Fe}^{4+} from an initial binding site centrally located in the major groove to a position where the end of the cylinder is sampling the spine of the DNA. (a) starting configuration, red (b) 1ns, orange (c) 2ns, yellow (d) 3ns, green (e) 4ns, blue.

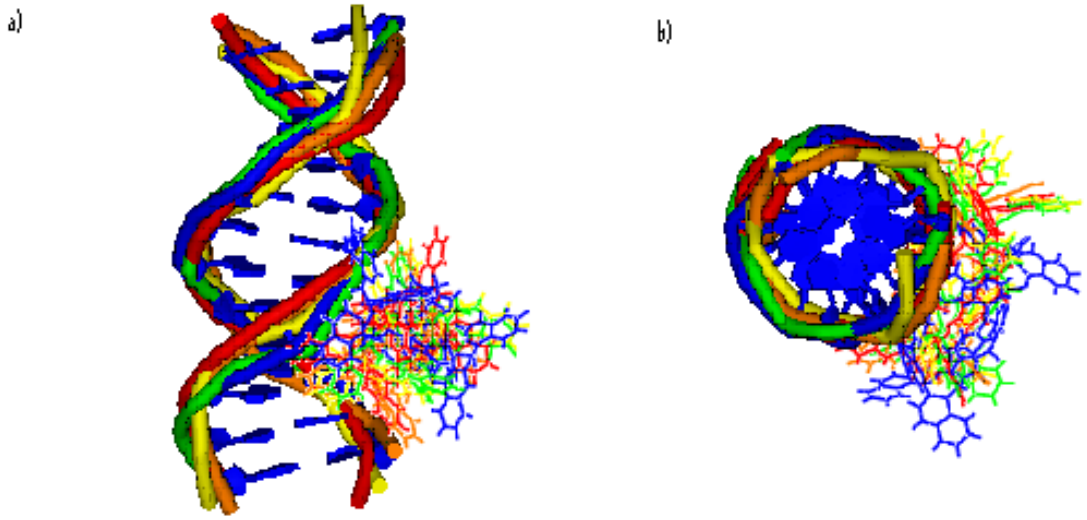


Figure 5-35 A schematic diagram of the five extracts superimposed to show the observed similarities of the DNA over 4ns of simulation, a) side view b) end view.

With the five extracted configurations superimposed the DNA sequence in Figure 5-35 can be seen to maintain the canonical B-form of the starting structure with deviations within the fluctuations due to flexibility.

Conformational and helicoidal parameters were calculated in 0.25ns portions for the trajectory. The conformational parameters for 0.75 – 1.00ns and 3.75 – 4.00ns are shown in Figure 5-36 and Figure 5-37. The general trend for the conformational parameters throughout this 4ns simulation is for the DNA to fluctuate about the values characteristic of B-DNA. While fluctuations can be seen in the conformational analysis these are observed to be short lived often lasting no longer than a 0.25ns period of the simulation. This behaviour is consistent with the expected stability for the $d(CGCGCGCGCGCG)_2$.

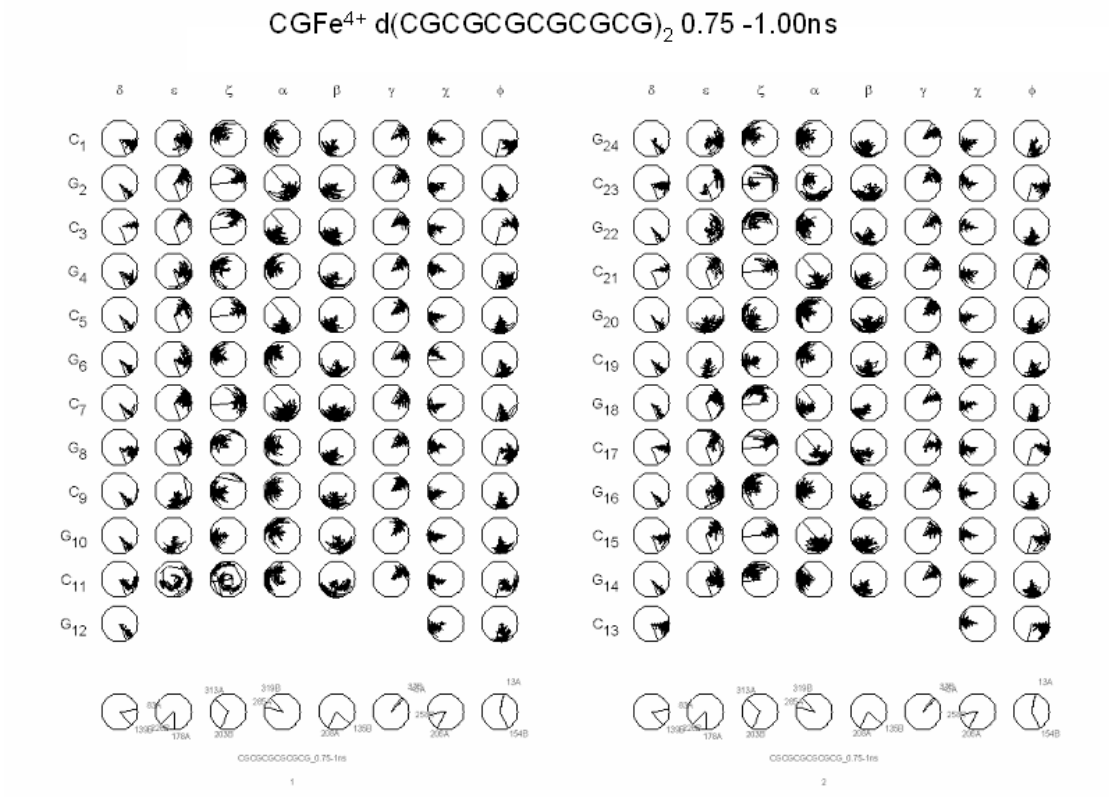


Figure 5-36 Conformational parameters for 0.75 – 1.00ns of C^{4+} $d(CGCGCGCGCGCG)_2$.

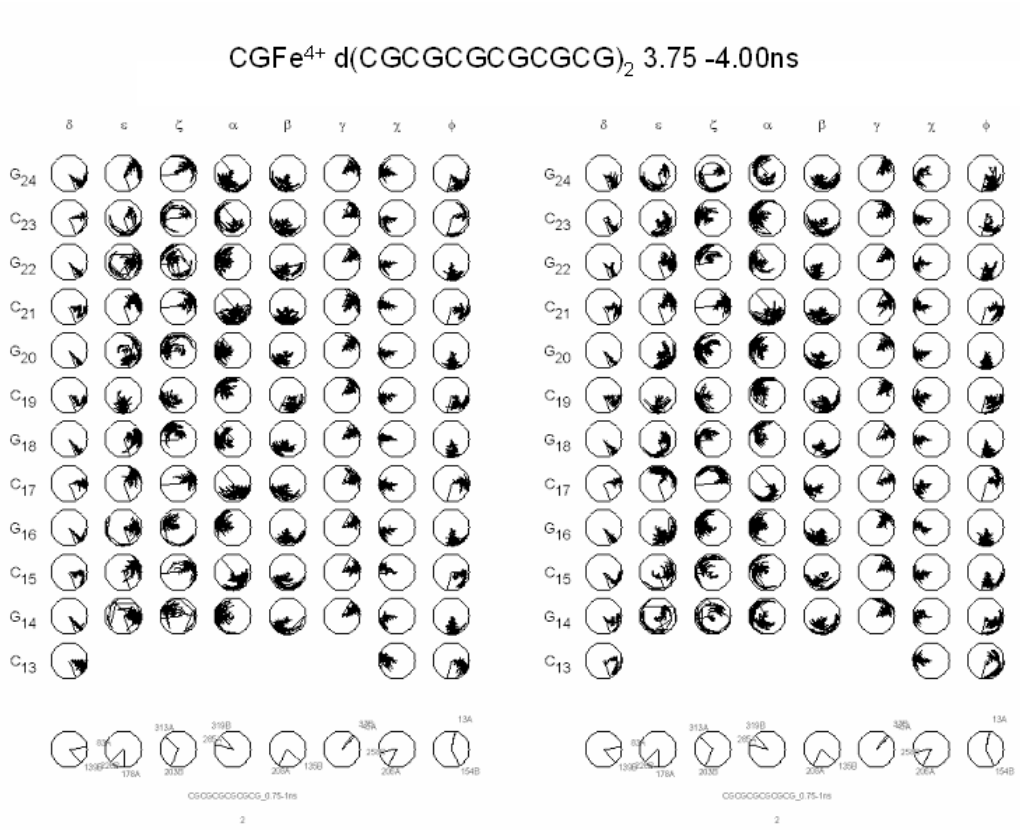


Figure 5-37 Conformational parameters for 0.75 – 1.00ns of C^{4+} d(CGCGCGCGCGCG)$_2$.

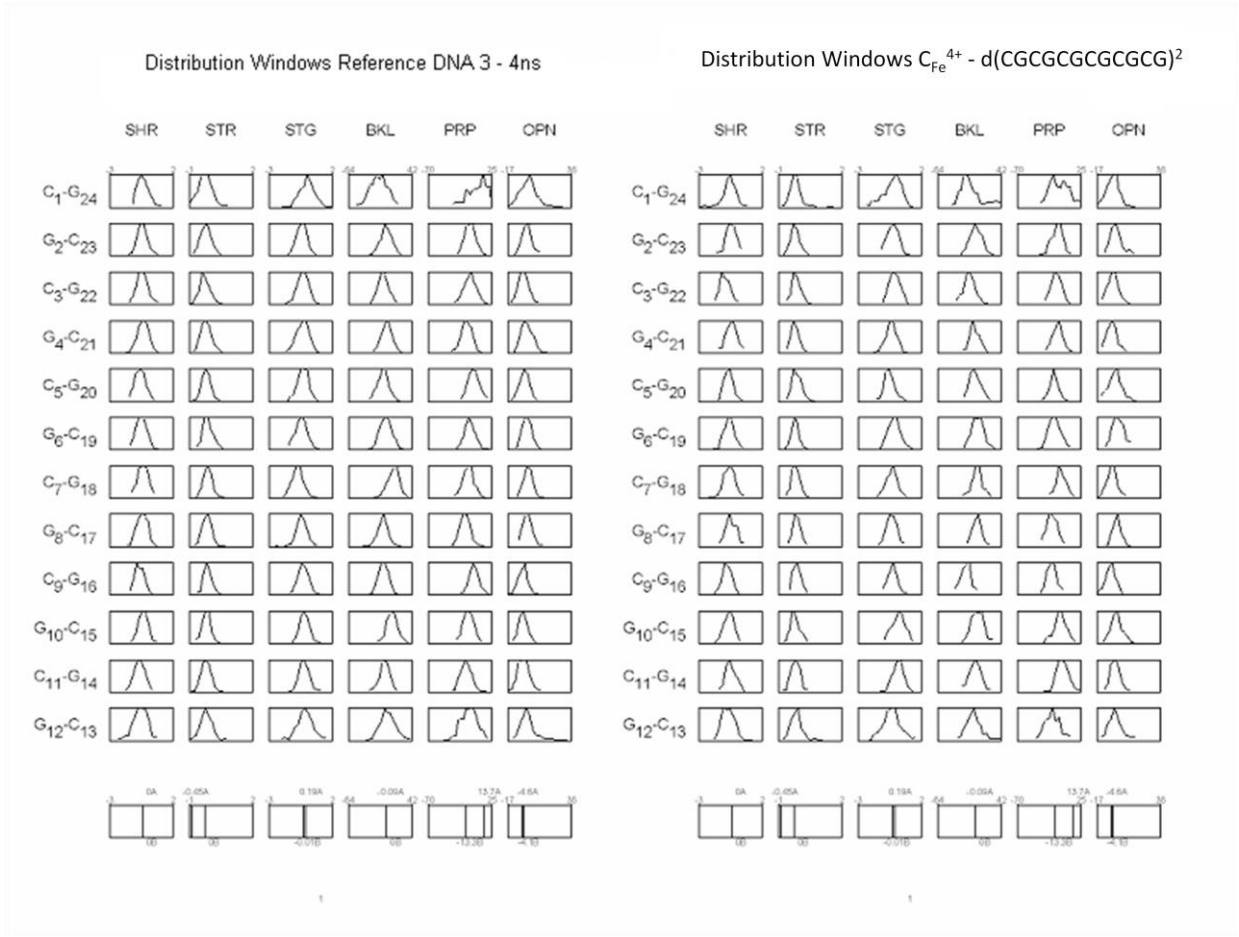


Figure 5-38 Distribution of conformational parameters for the reference $d(CGCGCGCGCGCG)_2$ DNA system and the C_{Fe}^{4+} system.

The stability of the DNA both with and without C_{Fe}^{4+} present is also confirmed by the smooth distribution curves of the conformational parameters. During this simulation it isn't obvious that the increased charge on C_{Fe}^{4+} compared with C_{Cu}^{2+} has any significant effect on the overall DNA conformation. The extra stability of the DNA demonstrated from this set of simulations, compared with those carried out using CHARMM22, makes any correlation between Cylinder charge and DNA conformation less obvious.

5.8 Conclusions

The aim of this work has been to study the interaction of DNA with groove binding ligands. From experiment the selected synthetic cylinders are seen to have an unprecedented ability to induce dramatic conformational changes in DNA. To study these systems methodology has

been developed to follow on from work by S. Khalid during her PhD. Development of the protocol to include CHARMM27 and the investigation of different cylinders and DNA sequences for longer simulations have added to our understanding of the systems being simulated and the methods used.

Before writing the conclusion it is important to understand the statistical significance of any observations that may have been made from viewing and analyzing the simulations . During this research, in an effort to begin simulations of a variety of systems, the attempt to sample a variety of different DNA-binding systems has come at the price of not having several trajectories for each system and therefore reduced sampling. To ensure that the conclusions are taken in context a section on statistical significance follows prior to specific conclusions.

5.8.1 Statistical significance

Having referenced literature, spoken to colleagues, built models and run simulations it is important to know how significant the results and subsequent conclusions are and how much confidence we have to be able to answer the original hypothesis.

The usefulness of DNA MD simulations rests on our ability to compare the predictions of such simulations with experiments. If the predictions regarding some properties of interest are in agreement with the experiments, we may hypothesize that other dynamic properties, in particular ones that may not be readily accessible experimentally, are also correctly predicted by the computational model. This may provide new insight into the system of interest and may lead to novel hypothesis and, ultimately, to the testing of the hypothesis with new experiments. Central to driving this cycle of knowledge is the ability to quantitatively compare MD simulations with experiments. The comparison of MD simulations carried out under different conditions is also very important, in particular for the advancement of the simulation

methodology (i.e., comparison of simulations based on different force-fields) or for comparative studies of similar systems.

As MD simulations of complex systems continue to develop the large biological molecules continue to be challenging. For large systems repeated from slightly different but equally plausible initial conditions, MD simulations of protein equilibrium dynamics predict different values for the same dynamic property of interest[283,284,285]. These variations occur because MD simulations fall short of properly sampling the complete conformational space for these large systems, an effect known as the "sampling problem" [286,287,288]. In 1995, when 1-nsec MD simulations of fully solvated proteins were hardly attainable it was speculated[289]that the sampling problem may be alleviated with a simulation time of 100 nsec. More recently, several 40-nsec MD simulations of HPr and T4-lysozyme were analysed for convergence in sampling[288]. Not only did these simulations fail to provide a complete picture of the proteins conformational space, they also suggested that this goal will remain challenging for the foreseeable future.

The central question concerning the reproducibility of MD simulations can be restated as follows: How do we know that some property observed in an MD simulation, which may lend itself to some important biological or physical interpretation, is not merely an "accident" of the particular simulation? For example, in two 4-nsec MD simulations, one of Ca^{2+}-loaded and one of Ca^{2+}-free calmodulin (CaM), it was observed that the central helix remained straight in the Ca^{2+}-loaded simulation but was bent in the Ca^{2+}-free simulation[290]. The research suggested that this indicated a change in CaM conformation induced by Ca^{2+} ions. In another, unrelated study, 15 independent 1-nsec MD simulations of Ca^{2+}-loaded CaM were performed[291]. In these simulations, the central helix remained straight in some MD runs but was bent in others, suggesting that bending of the central helix in any single simulation is a random event,

occurring with a certain probability given the simulation time. While bending of the CaM central helix may well be influenced by the bound Ca^{2+} ions, it seems unwarranted to draw such a conclusion based on only two observations, i.e., one Ca^{2+}-free and one Ca^{2+}-loaded simulation. This is a poignant example and word of caution with regards to the results presented in this thesis. While the simulations presented are extensive relative to some studies there is only one trajectory presented for each system simulated and in the overall timescale of a number of biological mechanism the longest simulation of 5ns is still very short. For this reason to extend the study documented within this thesis multiple production runs should be undertaken from slightly different starting configurations to enable better sampling of the conformational space.

The above example warns that predictions derived from MD simulations behave as a sample drawn from a certain parent population. Thus, to understand predictions of MD simulations in quantitative terms, one needs to understand the central tendency (such as the population mean) and the variability (such as the population standard deviation) of the parent population from which the predictions are "drawn." It follows that given some dynamic property of interest, understanding of the parent distribution of the prediction is of central importance for quantitative comparison of simulations with experiments.

The approach taken in this research has been limited in using the computer resource available to perform single MD simulations for each system for as long as possible. However, to continue this work and ensure that the simulations develop into statistically meaningful representations of the system being investigated it would be important to change the structure of the simulations. While it would be an improvement to simply take various related low energy starting configurations and generate more independent simulations to improve the statistical validity of the findings great benefits can be achieved by running many shorter simulations that

after analysing by statistical hypothesis testing may be combined and considered as a more complete image of conformational space.

MD simulations are stochastic in nature, and therefore, very infrequent, so-called events may be observed in any single MD run. In the context of multiple MD simulations, the term "rare event" refers to any process observed infrequently in only one MD run or a very few MD runs within the collected sample. The probability of such a process cannot be predicted reliably, no matter how large the sample is. Furthermore, a rare event may result in values that are outside of the likely range (i.e., outliers), which in turn may skew the overall picture obtained from the sample of MD runs. Therefore, it is highly desirable to detect such events early in the analysis. In principle, this could be done by finding outliers in data; however, finding outliers in a sample drawn from an unknown probability distribution is a nontrivial task. Literature suggests that analogs of the z-score based on outlier resistant estimators such as median absolute deviation perform well[292].

Developing the protocol to run simulations in this way merits further investigation; however, it is clear that a sample of MD simulations provides a much better opportunity to deal with rare events compared with the situation when only one MD run is collected.

5.8.2 CHARMM force-field comparison

In conclusion to hypothesis one (Section 1.6) that is proven to be true, solvated simulations with comparable DNA strands performed under the same simulation conditions showed that DNA strands simulated with the CHARMM22 force-field have a tendency for the conformation to change from a starting conformation of B-DNA to A-DNA even when there is no interacting cylinder present. The CHARMM27 control experiment of the same DNA strand maintained its form for the duration of the simulation.

Solvated simulations of DNA and cylinder, comparable to those modeled by S. Khalid, with C_{Fe}^{4+} and d(CGCGCATATACG)$_2$ didn't show any overall DNA response to the cylinder when modeled using the CHARMM27 force field. This was in contrast to the dramatic coiling of the DNA around the cylinder that was observed using CHARMM22. Based on these findings it appears that the earlier CHARMM22 force-field has some differences compared with CHARMM27. For example when the highly charged C_{Fe}^{4+} is bound to the DNA major groove CHARMM22 predicted that C_{Fe}^{4+} tended to disrupt A-T base pairs adjacent to the cylinder binding site.

In addition to the comments made earlier on force field dependence, section 5.5, further detail is included here to complete the comparison between these two force fields and their influence on the DNA response in their respective simulations. Previous work with related systems predicted a neutralized ligand to coil DNA at least as strongly as the tetra cationic ligand. The reason for this was shown to be that regions of low dielectric constant enhance the phosphate repulsions in DNA and cause it to bend away from the low dielectric region.[293] Although the neutralized ligands employed did generate electric field gradients, its enhanced negative surface charge should also have helped accentuate the phosphate repulsions and induce the DNA to bend away from the ligand, or at least coil less tightly around it. The finds were not as expected with the coiling being enhanced by the neutralized ligand implying that other forces – the van der Waals forces – drive the coiling.

Ligand charge does have some affect on the DNA response, but its principle manifestation is different in the two force fields. The contrast helps give insight into the binding mechanism. In CHARMM27, the +4 *e* charge has been seen to attract the negatively charged DNA backbone, causing the major groove to contract onto the cylinder. With CHARMM22 its effect is to strain the hydrogen bonding between base pairs in the duplex. The induction of such strain is

reasonable given that most force fields describe hydrogen bonding primarily through electrostatic interactions. Orbital interactions, such as charge transfer and π-electron polarization, do contribute strongly to the hydrogen bonding between base pairs,[294 295] but these are also likely to be strongly perturbed by the proximity of a tetracationic ligand. In reality, both the strain to the base pair hydrogen bonds and the contraction of the DNA backbone onto the ligand are likely to be present, and getting the balance right will be an important validation of the force field. In the context of the work by Khalid *et al*.[296] the CHARMM22 result that duplex base-pairing will be disrupted by C^{4+} provides an interesting prediction that awaits the measurement of definitive X-ray or NMR structures.

If, as indicated above, the disruption of base pairs in a DNA duplex with CHARMM22 is due to a competition between inter-base pair hydrogen bonding and ligand base electrostatics, then it is of interest to identify how strong the ligand electrostatics need to be in order to effect the base pair mismatches. Analogous simulations[296] to those reported in the last section were performed with CHARMM22 and with a variant of the cylinder in which overall charge was set to +3 *e* (C^{3+}), again achieved by a constant shift of the atomic charges of all C and non-polar H atoms. Within 0.25ns this system also showed disruption of the base pairing within the double helix. Similar calculations with a +2*e* variant of the cylinder gave results that were intermediate between C^{3+} and C^{0}, with frequent bifurcated pairings in which an A was found to bridge two T bases in the opposite strand, but these did not lead to irreversible breakdown in the base pairing on a nanosecond timescale. It is concluded that some disruption of the duplex base pairs is a robust feature of the CHARMM22 model, and is not especially sensitive to the parameterization of the cylinder charge distribution.

5.8.3 Reference DNA systems

All reference DNA systems were seen to be stable after 2ns and so forth as the simulations have been extended.

5.8.4 C_{Fe}^{4+}

When C_{Fe}^{4+} was docked with $d(CGCGCATATACG)_2$ the docking calculations indicated a major groove-binding mode. The DNA remained stable during the subsequent solvated simulation of 3ns during which the cylinder moved from the major groove to a position on the DNA backbone. Throughout this simulation the DNA was stable with the exception of one terminal base pair breaking at 2.0 – 3.0ns. While the effect of the cylinder on the DNA was not apparent as previously seen with CHARMM22 there is evidence of changes to base pair helical parameters that are near to the points of contact of the cylinder.

Following the docking calculation C_{Fe}^{4+} was simulated with $d(ATATATATATAT)_2$ from a starting configuration in the major groove. As in the previous system the cylinder is seen to migrate out of the major groove to a position on the backbone where it stays between 2.0-4.0ns. While there is no bending to a large extent in the DNA a finer detail of alternating base pairs changing to the A-DNA form is observed before the system develops to a point where the majority of the base pairs are in the form of A-DNA.

The simulation of C_{Fe}^{4+} shows the DNA sequence $d(CGCGCGCGCGCG)_2$ to be unaffected and not display the characteristics seen in the A-T system due to the reduced flexibility of a C-G sequence.

In conclusion to hypothesis number two (section 1.6) that is seen to be true. The greater size of C_{Fe}^{4+} compared with ${}^{P}C_{Fe}^{4+}$ and C_{Cu}^{2+} has prevented it from finding a favorable binding site in the major grove within the timescale of these simulations. The simulation in which C_{Cu}^{2+} was

bound to d(ATATATATATAT)$_2$ also supports this being the only simulation in which the cylinder maintained it position in the major groove for the full duration.

5.8.5 C_{Cu}^{2+}

A 5ns simulation of C_{Cu}^{2+} and d(CGCGCATATACG)$_2$ was performed from the major groove starting position identified from a high temperature docking calculation. During this simulation the cylinder left the major groove within the first nanosecond and spent the remainder of the simulation in a variety of configurations bound to the backbone of the DNA. During the simulation the inter-base pairs display an alternating shift in form towards A-DNA. This pattern through the base pair steps is well defined by 5ns and is seen to cause obvious kinks in the backbone of the DNA.

When simulated with d(ATATATATATAT)$_2$ for four nanoseconds the C_{Cu}^{2+} cylinder on this occasion is seen to start in the major groove and remain in the major groove for the length of the simulation. While there are no major changes to the DNA conformation on this timescale toward the end of the last nanosecond stage of the simulation a definite feature can be seen in the DNA backbone. During the simulation the DNA has accommodated the cylinder and shadowed its shape. This feature is not repeated in the C-G DNA strand and could indicate a combination of DNA sequence and cylinder that complement each other. It is not surprising that the DNA strand that this feature is observe in is the more flexible A-T strand with an overall greater flexibility. In conclusion to hypothesis four this appears to be confirmation that with all other controlled variables being the same the d(ATATATATATAT)$_2$ sequence is more able to respond to the presence of a metalosupramolecular cylinder.

In conclusion to hypothesis number three (section 1.6) the observed changes in the DNA backbone described above are some of the clearest indications within this research of a cylinder inducing conformational change. As C_{Fe}^{4+} didn't bind in the major groove to allow a

comparison of charges further work would be required to fairly answer the hypothesis and confirm the importance of charge in causing DNA conformational changes. This hypothesis could be investigated further in a theoretical experiment where the charge on C_{Cu}^{2+} is artificially increase to C_{cu}^{4+} within the model to then be able to rule out shape differences and have a more conclusive answer.

For the system C_{Cu}^{2+} with $d(CGCGCGCGCGCG)_2$ the cylinder is observed to sample a number of configurations during the 3ns of simulations. Within this system helicoidal parameters are seen to vary randomly and generally within the range defined as B-DNA. However, closer inspection of the helicoidal parameters shows a number of instances where a base may change its conformation and after a short period in one state it will return to its original B form with no intermediate configuration. On a few occasions these transitions are seen to take place local to the cylinder and so could be an effect of the cylinder on the DNA.

To conclude there is only one system for which the cylinder has been seen to reside for a length of time greater than 2ns in the major groove. As neither the C_{Cu}^{2+} nor C_{Fe}^{4+} has demonstrated a consistent binding mode it is hard to deduce the affect of charge on the DNA cylinder complex.

5.8.6 Timescales

Using the 5ns C_{Cu}^{2+} $d(CGCGCATATACG)_2$ as an example it is clear to see that had the simulation been any shorter whole portions of a varied simulation would have been missed and our ability to understand what appears to be a very mobile cylinder would be reduced. For this reason I think further simulation is required before we can identify if this cylinder has a preference for the major groove, minor groove or backbone, all of which it has sampled during the first 5ns.

In the context of achieving long timescale simulations the stability due to CHARMM27 could be useful as it enables long simulations in comparison to those with CHARMM22 in which the DNA can rapidly change form due to over stabilisation of the A form relative to the B form of DNA.

For similar reasons as the stability and timescale, to ensure what is being observed is reproducible, duplicate simulations will need to be run from like starting points. The duplicate starting points are readily available for extraction from the docking calculations already performed.

5.8.7 Loading

It is possible that the response of the DNA to the cylinder is related to the concentration of the cylinder present in the simulation cell. If this is the case and more than one cylinder needs to be bound to the DNA then it would be necessary to simulate a DNA strand composed of 24 base pairs to ensure they have room along the strand on which to bind. Also, if the DNA was showing fraying at the ends this would be a modification to consider. However keeping the system size to a minimum means a large saving in computational time that can be used to extend systems of short strands that much further.

5.8.8 Further Analysis

In addition to the parameters reported above calculation of the helix bending would improve the learning available from these simulations and also allow comparison with other experimental results (see 1.5.5). However, quantifying curvature in the DNA helix can be problematical. Any measure must be able to distinguish between the local oscillations in base-pair orientation found in, particularly, the A-form, and a real bend in the overall helix direction. This can be particularly difficult in short strands of DNA, such as the dodecamer studied here, since one

cannot use the behaviour of the helix beyond the binding site to confirm the persistence of any bend. To measure the extent of bending the 3DNA definition of the local helix direction can be used. This uses the geometry of any two adjacent base pairs to define a unit vector, U_i, along the helix axis at that step. Bending of the DNA can then be monitored by defining an angle, $v_i = \cos^{-1}(U_{ref} \cdot U_i)$, which describes the total bend in the DNA between some reference step (can be taken to be the first stable step) and the *i*th base-pair step. Most importantly, this definition gives perfectly aligned U_i(v_i=0 for every *i*) in both the canonical A and B forms, so that deviations from 0 can be interpreted as bending of the DNA.

Having defined the DNA helix the angle of the cylinder to the DNA can be calculated as being the tangent to the DNA helix that passes through the two metal centres of the cylinder.
V_i values for the, averaged, DNA structures following completion of the DNA response give a good indication of both the global bend induced in the DNA by the ligands, and how it is distributed along the dodecamer.

Another common descriptor for helix curvature is the normal vector plot (NVPs). NVPs are projections of the unit vector normal to the plane that is normal to the plane of each base pair onto a plane that is normal to the helix axis. Linear segments of the DNA can then be identified as clusters of neighbouring points, while coiling is seen as a steady change of orientation across a number of adjacent base pairs.

When considering DNA bending is observed it can also be useful to ask how the size of major and minor grooves is affected. These can be monitored by measuring the distance between appropriately displaced phosphate groups, with the P atom representing the position of the phosphate. The width of the major groove can be defined as the distance between the P atom in the *n*th nucleotide, counting from the 5' end of one strand, and the P atom of the $n+3^{rd}$

nucleotide from the 3’ end of the complementary strand; this numbering assumes 5’ nucleotides. Similarly, for the minor groove one can use the distance between the P atoms in the *n*-4th nucleotide from the 3’-end of the complementary strand. The distribution of these inter-phosphate distances could be determined by analysing every configuration saved from the final 0.25ns of each simulation. To avoid artefacts from end-fraying, those phosphates adjacent to the first and last base pairs would be omitted from the analysis. This method leaves five measurements of the major groove width, and six for the minor groove. It should be noted that for a dodecamer this means that the minor and major groove measurements were made on largely different regions of the DNA, with only two bases (6-19 and 7-18) being used to measure both the major and the minor groove widths.

6 Related Computational Studies

While the structure of DNA is directly related to its solvent environment (section 1.5.2) there are many simpler molecules that can be observed and modelled to gain a greater insight into the competing properties that induce or disrupt structure in a variety of ways. The area of study looking at Hydrates and Hydrate inhibitors demonstrates techniques available for observing the properties of water and the interrelationships between a solute and its environment, some of which could be used to probe the water environment around DNA as future work to this project. The insight into water from these simulations is interesting and has been a way of developing an understanding of water on simpler systems without the complications of geometry and timescales which make solvation analysis around DNA difficult.[297]

Gas hydrates are a special type of inclusion compound. Inclusion compounds are compounds that comprise of two, or more, molecular species that order themselves such that one molecular species, the 'host', entraps the other molecular species, the 'guest'. In clathrates the host molecules are arranged to form hollow polyhedra, or 'cages'. For gas hydrates the host species is water which forms a crystalline lattice based on at least two types of polyhedra. These hollow polyhedra accommodate a variety of guest species. Gas hydrate formation is a concomitant process requiring the presence of both the host and guest molecular species. The water lattice is thermodynamically unstable by itself and so the stability of the gas hydrate is a direct result of the presence of the guest molecules in the hydrate cavities.[298]

Gas hydrates can cause a number of problems in a variety of contexts, for which reason there is a high demand for active and effective inhibitors. Inhibitors can be classified as either kinetic inhibitors or thermodynamic inhibitors. Kinetic inhibitors differ from thermodynamic inhibitors in that they do not shift the hydrate equilibrium temperature or pressure, but rather

prevent hydrate growth for a finite period of time. These inhibitors are typically low molecular weight polymers that are postulated to adsorb strongly to the hydrate surface. The kinetic inhibitors can also act as anti-agglomerates and therefore prevent the large hydrate masses from forming, which would in turn block the pipelines.[11] The aim of kinetic inhibition is to prolong the period prior to that in which catastrophic growth is experienced. Polyvinylpyrrolidone (PVP) and related compounds like Polyvinylcaprolactam (PVCap) have been assessed experimentally and found to delay the onset of hydrate formation;[21] thus they are a convenient standard by which to judge simulation results.

This chapter includes techniques, used to probe simple structures, some of which could be useful in understanding the solvent environment in the DNA binding work. The motivation for using this area of study as a starting point to understanding the structuring effects within solvents was the work by Carver *et al*[301] in his work he reported cooperative structuring between CO and CH_3 side groups, and an unusual dynamical effect that he attributed to a solvent cage making this an interesting test system on which to develop one's ideas for probing solvent structure.

6.1 Modelling Kinetic Hydrate Inhibitors

6.1.1 Molecular Dynamics Simulation

Setting up a MD simulation to investigate kinetic hydrate inhibitors (KHI) involves several steps. The first step is the creation of an equilibrated box of solvent, into which the solute of interest is added. The solute is added in such a way that any solvent molecules that create a "bad contact" with the solute are removed. Bad contacts are defined to occur when two atoms are found to be closer than $\xi\sigma$, where σ is the LJ parameter and ξ is a user-specified tolerance

(typically close to 1 for non hydrogen-bonding systems). Where bad contacts are identified, the associated solvent molecule is removed from the simulation. The whole system was then subjected to an energy refinement using an energy minimization algorithm. This energy refinement relieved local stresses due to bond-length and bond-angle stresses in the solute and solvent. The atoms were then assigned velocities from a Maxwell distribution at a temperature below the desired temperature, then the system heated to the equilibrium temperature before proceeding with an NVT simulation; usually, a temperature near zero is used as the starting temperature. The system is then equilibrated as an MD simulation while adjusting the temperature and density to appropriate values. As the insertion process necessarily created large overlaps, the resultant forces necessitated temperature scaling while the initial strain was relieved. The temperature is brought into the range of interest by incrementally increasing the velocities of all the atoms, either by reassignment from a Maxwell distribution at an increased temperature or by scaling all velocities. The instantaneous temperature, $T(t)$, at any given time t is defined in terms of the mean kinetic energy by,

$$T(t) = \frac{1}{(3N-n)k_B}\sum_{i=1}^{N} m_i \left| v_i \right|^2$$

Equation 6-1

Where $(3N - n)$ is the total number of unconstrained degrees of freedom in the system, v_i is the velocity of atoms i at time t, and k_B is the Boltzmann constant. From this expression, if the velocities are scaled by a factor of,

$$\sqrt{T'/T(t)}$$

Equation 6-2

then this will result in a mean kinetic energy corresponding to a temperature of T'. The aim of the heating and equilibration process is that it decreases the probability that localised fluctuations in the energy will be present at the beginning of the main simulation. It also

ensures that highly strained regions, that might otherwise lead to catastrophic numerical integration error, have been removed from the system. Once the equilibration process has been completed then the actual simulation can commence. Some details of the simulations run are given below.

Molecular Dynamics simulations (MD) were used to study the behaviour of four monomers of polymeric KHI immersed in water to see if they have unusual water structuring properties. The inhibitor molecules were modelled with all degrees of freedom and with bond lengths unconstrained. The CH_3 and CH_2 groups were represented using united atoms with the CHARMm force-field. CHARMm template charges, Fig 1, were used because of their known accurate description of crystal structures containing related pyrrolidone compounds.[299] It is also worth noting that the CHARMm force field has evolved over the years and so some slight differences between the results obtained from the MeP standard simulation here and that reported in the literature[4] may be expected. Water was modelled using the SPC potential,[300] in this model water is treated as a rigid body with partial charges on each atom, but with all van der Waals interactions associated with the oxygen atom.

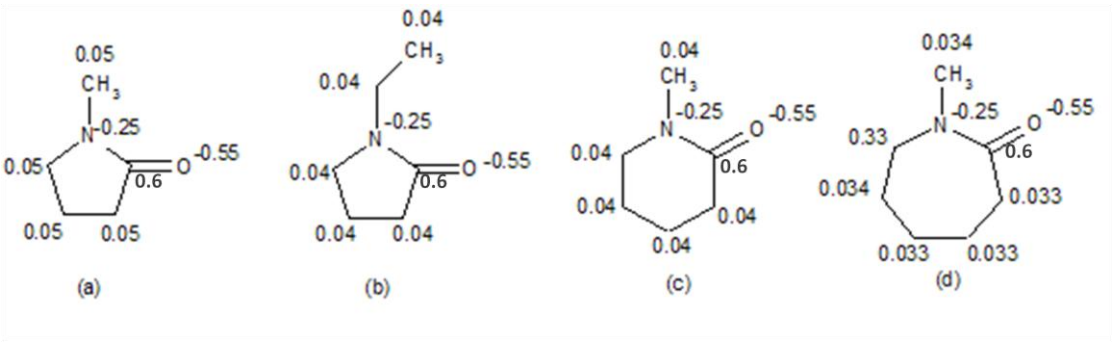


Figure 6-1 The structure of (a) MeP (b) EtP (c) MeVam and (d) MeCap. Numbers indicate the partial atomic charges for the CHARMm template charge model.

The major part of this study is based around four simulations of four potential hydrate inhibitor fragments with the aim of elucidating the effect of these monomers on the surrounding aqueous environment at about 292K. The first of the four simulations conducted was to validate the

methods being used against the work carried out by Carver *et al*.[301] on methyl pyrrolidone, MeP. The other three were conducted under the same conditions to allow direct comparison of our analysis of the structural variables. In each simulation a single inhibitor fragment was embedded in 400 water molecules, with cubic periodic boundary conditions used to mimic infinite systems. A simulation of pure water was also conducted to enable an evaluation of the inhibitors effectiveness relative to a system where it is absent. Details of the simulation conditions are given in Table 6-1.

Simulation	Solute	N_S	N_W	L/Å	t/ps	<T>/ K
1	MeP	1	400	23.000	2 x 400	292.0
2	EtP	1	400	23.000	2 x 400	292.0
3	MeVam	1	400	23.000	2 x 400	292.0
4	MeCap	1	400	23.000	2 x 400	292.0
Water	None	0	256	20.000	400	292.0

Table 6-1 Thermodynamic data for simulations (1)-(4) & Water, N_S and N_W are the number of solute and water molecules respectively; L is the cubic box length; t is the total simulation time; $\langle T \rangle$ is the average temperature.

Following Carver et al., we have assumed that a pressure of ~10^2 atm[301] does not markedly affect the behaviour of liquid water systems.

6.1.2 Simulation Details

All simulations were performed with DL_POLY[302] and in the canonical ensemble, (*N, V, T*); the temperature constraint was achieved using the Nosé-Hoover thermostat. Equations of motion were integrated using a 1fs time step, which was found to give stable simulations, and good energy conservation when tested in (*N, V, E*) simulations. Systems were allowed to equilibrate for 10 ps before accumulating two sets of 400 ps giving a total simulation time of 800 ps. Configurations were saved every 0.016 ps for later analysis. Cubic periodic boundaries were used throughout to mimic the infinite system. Bond constraints and bond angle constraints for water were implemented using the SHAKE algorithm. Electrostatic interactions were evaluated using the EWALD method with a convergence parameter $\alpha = 0.2437$ and reciprocal lattice truncation at $k = 14\pi/L$ in each of the *x, y, z* directions, where L is the length of the simulation

box. All other interactions were truncated when the inter-atomic separation exceeded 10.0 Å. Regarding the treatment of bond lengths Carver modelled MeP with fixed bond lengths but with all other degrees of freedom unconstrained while investigations for this project did not constrain the bond lengths.

6.2 Results and Discussion

6.2.1 Structural analysis

To investigate the water structuring influences of the monomers a number of methods have been used. The results relating to the RDFs will be presented first.

In the following discussion RDFs for the distribution of all atoms of type *X* around atoms of type *Y* will be denoted *g*(*XY*). The subscripts S and W will be used to distinguish between atoms belonging to the solute or to water molecules, respectively. The methyl group of each solute will be denoted as Me. A superscript 5, 5et, 6 or 7 will indicate the solute being considered as either methylpyrrolidone (MeP) with a five membered ring, ethylpyrrolidone (EtP) with a five membered ring, methylvalerolactam (MeVam) with a six membered ring, or methylcaprolactam (MeCap) with a seven membered ring.

6.2.1.1 MeP Analysis

RDFs for MeP were compared to results published by Carver[303] as confirmation of correct modelling techniques before expanding the investigation to the other three systems. This comparison was also essential to prevent misinterpretation when comparing with the original work. By running the MeP simulation, as a control simulation, the effect of several important factors which differ slightly from the original simulations, can be examined, so allowing a qualitative comparison with the original literature. By validating the results gained from the control simulation quantitative comparisons between the four hydrate inhibitor fragments being

modelled can be made. Slight differences between the MeP control simulation and literature is expected due to the CHARMm force field having evolved since the time of publication.[1] Considering these advances, variation between this project and literature will arise from the slightly different charges, which are part of the CHARMm potential, as well as modelling of the inhibitor fragment without constrained bond lengths rather than the semi-ridged entity modelled by Carver. This might give some differences when considering the RDFs of each system and should be considered with any of the following analytical methods.

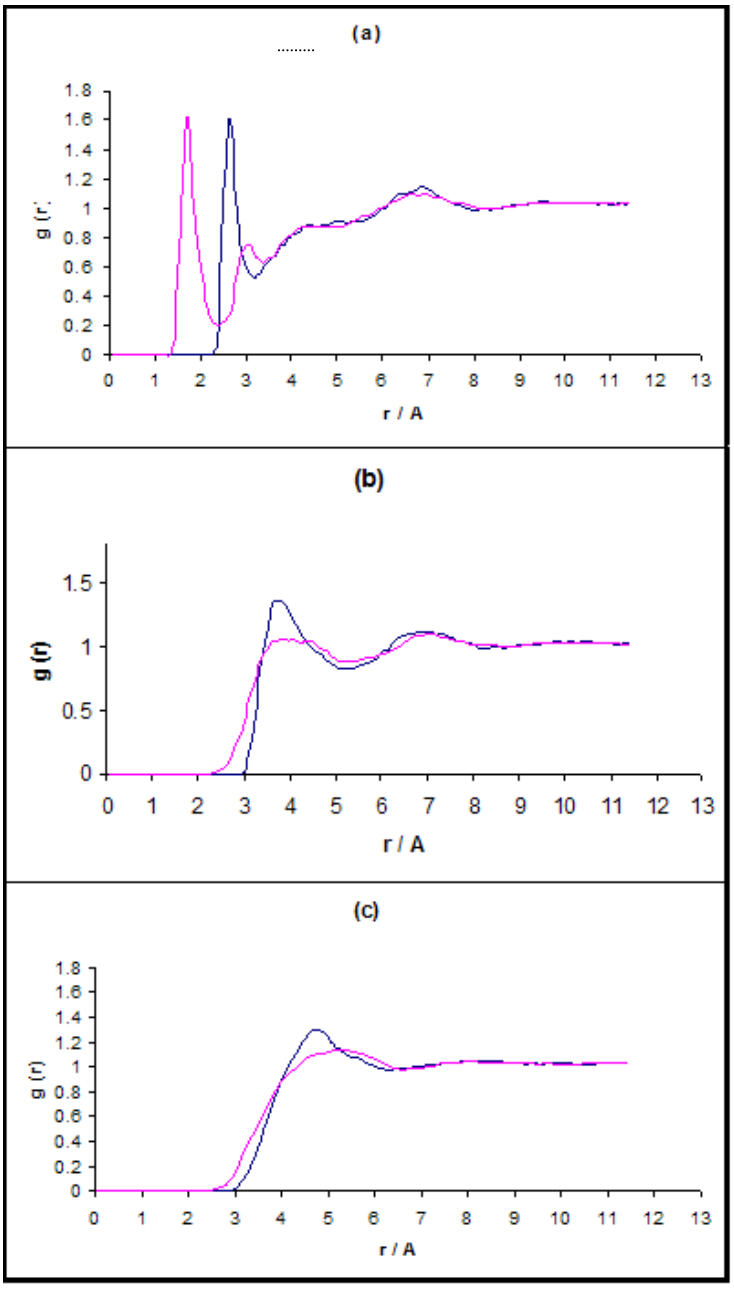


Figure 6-2 Radial distribution functions g(XY) solute atom Y is (a) oxygen; (b) the methyl group; and (c) the nitrogen. Solvent atom X is denoted by line colour: Pink = H_w; Blue = O_w. Data are taken from simulation (1) which is defined in Table 6-1.

RDFs for the distribution of water atoms around the O_S, Me_S and N_S sites are shown in Figure 6-2. Distributions characteristic of hydrogen bonding are seen around the amide oxygen (O_S). The first peak for $g^5(H_WO_S)$ is sharp and occurs at 1.67Å, indicating a strong hydrogen bond, and falls to a minimum at 2.47Å. Structure is also present in the RDF at longer distances in the range of 3-8 Å where three subsequent peaks are apparent. However, the excluded volume of the MeP molecule makes interpretation of this structure complicated:[304] peaks may represent second and subsequent solvation shells, or they may just represent first solvation shells that form around other parts of the molecule. Confirmation of strong hydrogen bonding behaviour is also seen in $g^5(O_WO_S)$, which displays a strong first peak at 2.59Å falling to a minimum at 3.28Å. The hydrogen bonding implied can be tested using the distribution of various hydrogen-bond related angles to gain a more detailed picture of the relative solute solvent orientations. These angular distribution functions (ADFs) may be calculated by defining a specific inter-atomic angle, and then constructing a frequency histogram of the values observed for this angle throughout the simulation.

As expected the methyl group demonstrates typical hydrophobic hydration, with broad peaks of $g^5(O_WMe_S)$ at 3.74Å and 7.1Å separated by a minimum at 5.12Å. The two peaks from $g^5(H_WMe_S)$, at 4.2Å and 7.2Å, with their broad nature, coincide with the values for O_w suggesting a hydrogen-bonding network forms between waters in the Me solvation shells, but the presence of only H atoms in the range 2.5–3.0Å indicates that the water O–H bonds do make significant incursions into the methyl cavity. If as anticipated from the RDFs the water tends to form hydrogen bonds around the Me group, then the favoured $Me\cdots O_w - H_w$ angles would be 180 and 70^o accounting for the H atoms in the range 2.5-3.0Å.

Structure around the MeP nitrogen is not well defined, with rounded peaks at 4.77Å for $g^5(O_WN_S)$ and 5.23Å for $g^5(H_WN_S)$. The two minima coincide, at around 6.7Å. The broad rounded peaks are characteristic of a solvation shell which is being occluded by other parts of the solute molecule.

From these RDFs it is possible to conclude that the solvation around the oxygen of MeP is hydrophilic while the solvation around the methyl is of a hydrophobic nature. The well-defined first peaks of the RDFs make it possible to assign the size of the first solvation shell. The radius of the aqueous solvation shells are 3.28Å for O_s^5 and 5.12Å around the Me_S^5 group. These results are in agreement with those of Carver *et al.*

Having confirmation that the results from the literature are repeatable, it remained for simulations **2**, **3**, and **4** to investigate the structure of the water around the O, N and CH_3 components of the potential inhibitor monomer fragments. The effect of changing the ring size was investigated with simulations **3** and **4** relative to **1**. While simulation **2** was used to observe the properties being affected by the alkyl group on the nitrogen. Also from this investigation the assumption that a single monomer fragment is representative of the polymer is tested.

RDFs for the distribution of Water oxygen and hydrogen atoms around the solute and hydrogen atoms around the solute are displayed in Figure 6-3 and Figure 6-4 respectively.

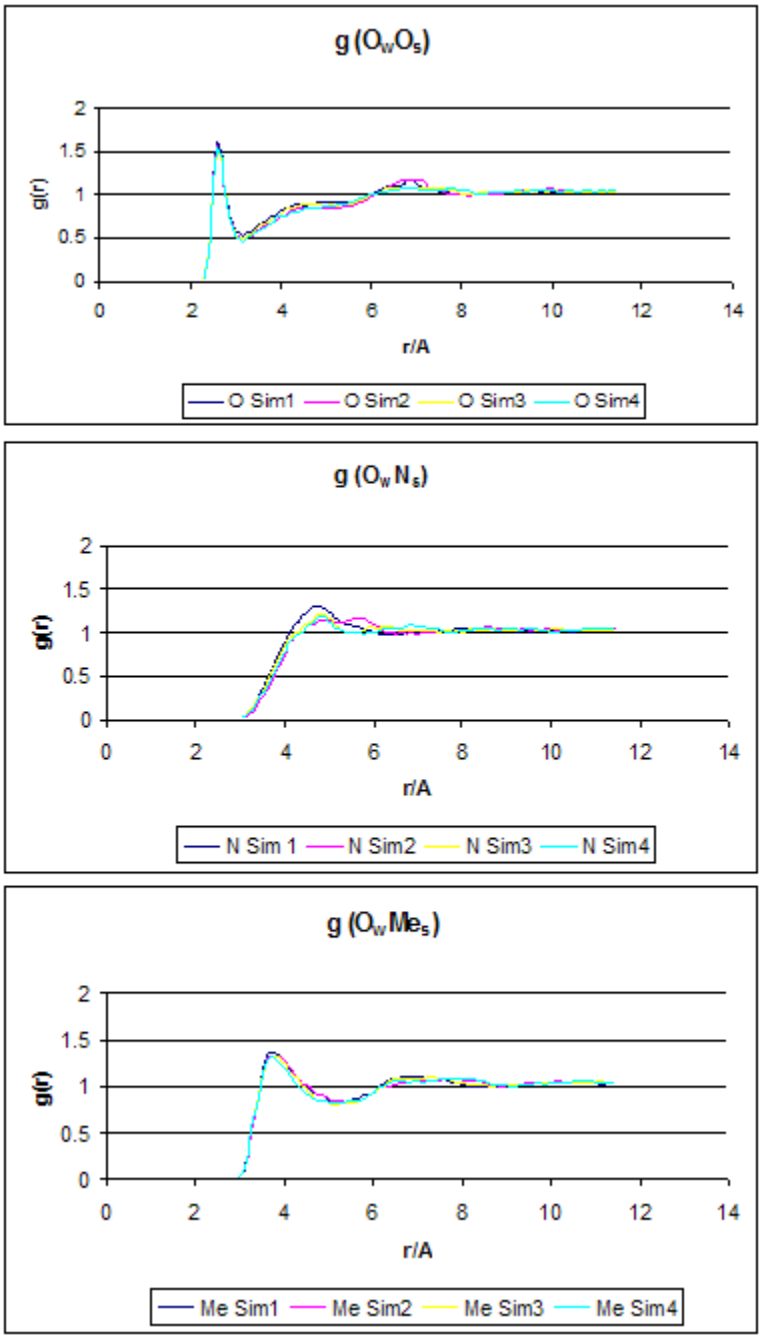


Figure 6-3 Radial distribution functions $g(O_wO_s)$ for simulations 1 - 4, $g(O_wN_s)$ for simulations 1 – 4, and $g(O_wMe_s)$ for simulations 1 – 4.

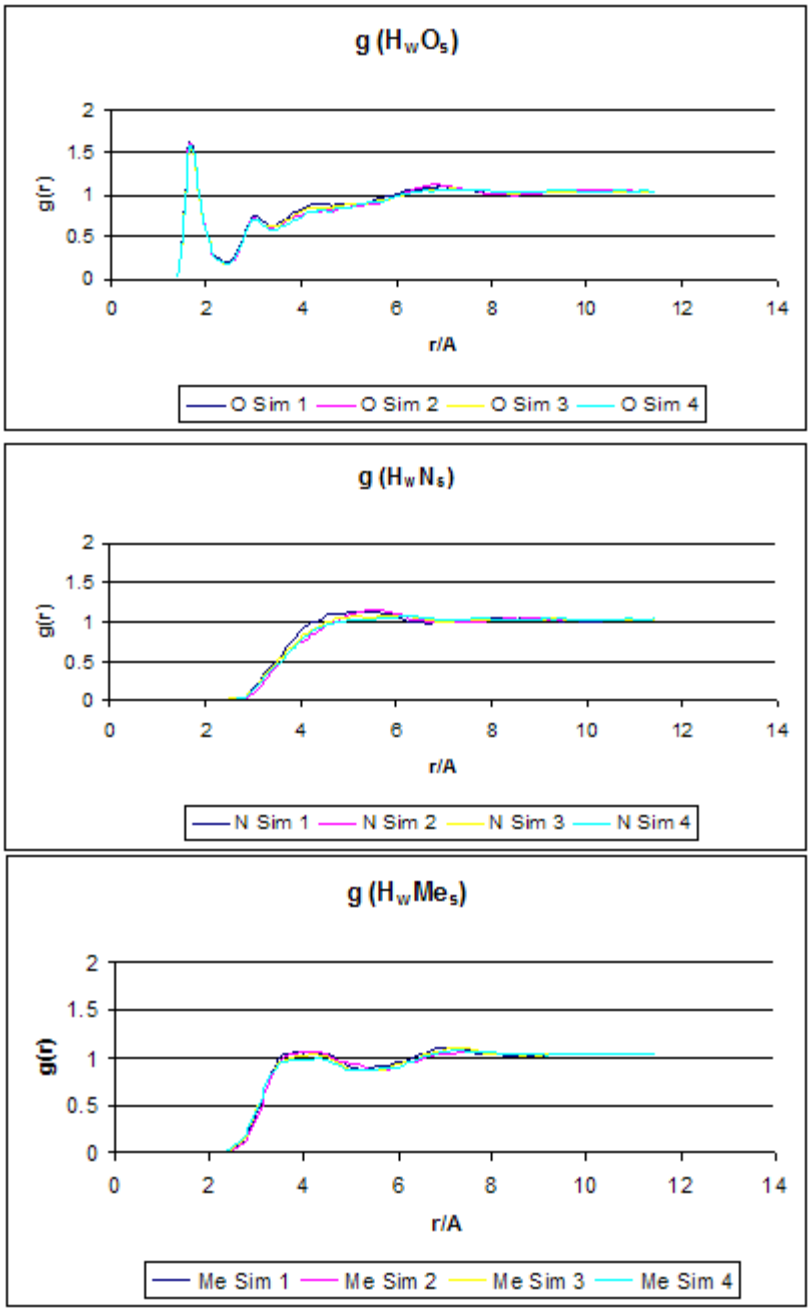


Figure 6-4 Radial distribution functions $g(H_wO_s)$ for simulations 1- 4, $g(H_wN_s)$ for simulations 1 – 4, $g(H_wMe_s)$ for simulations 1 – 4.

RDFs of water atoms around the amide oxygen, $g(O_wO_s)$ (Figure 6-3), show sharp well-defined first peaks with maxima at a radius of 2.59Å; these are indicative of hydrogen bonding. Although all four systems show characteristic hydrogen bonding behaviour it is difficult to see any significant difference in the intensities with the MeP intensity having the highest maximum, ~1.614, and the MeVam system the lowest value, ~1.555 it can't sure that this small difference is significant as it could be within the error introduced from plotting tabulated values and not

interpolating. The $H_w - O_s$ hydrogen distribution, $g(H_wO_s)$ reinforces the presence of hydrogen bonding with all systems having a sharp first peak coinciding at a maximum of about 1.67Å. Although when considering H_w they all look very similar with the differences between the four systems is not as great as for O_w. Also of interest when considering $g(H_wO_s)$ is the presence of secondary solvation structures. The well defined peaks at about 3.05 and much broader features at further distances, about 6 – 8Å, in $g(H_wO_s)$ indicate order present that is possibly attributable to a secondary solvation shell being generated by the electronegative N in the inhibitor molecule. This further structure is most prominent in the MeP system. When analysing $g(O_wO_s)$ structuring at around 7Å is evident with the peaks representing MeP and EtP having the narrower, more defined peaks, while the MeVam and MeCap systems are observed to level off with structure nearer to that of the bulk water.

The distribution of the water around the methyl group, as with the MeP, is typical of hydrophobic hydration and is similar for all systems. $g(O_wMe)$ exhibits maxima at about 3.73Å and 6.5 – 8Å at longer distances. MeP is again seen to generate the most intense peak and MeCap the least. For these relatively short simulations the error for these values is potentially large enough to reduce the significance of the differences observed. Extended simulations would improve the confidence in the significance of these values. The secondary structure shows more varied behaviour between the four systems. Both MeP and MeVam systems show the most structure while MeCap and EtP induce structure which is similar to that of bulk water. $g(H_wMe)$ displays two broad, rounded, peaks between 3.7 – 4.2 Å and 6.5 – 8 Å both of which coincide with the maxima seen in $g(O_wMe)$ and again suggest that a hydrogen bonding network forms between the waters in the Me solvation shells. When considering $g(H_wMe)$ the intensity of the peaks is much less than that observed around the hydrophilic amide oxygen with no

values going above 1.085. $g(H_wMe)^5$ causes the highest maxima for both peaks and has the most structured second peak.

The RDFs for the distribution of O_w and H_w around the MeP nitrogen are also shown in Figure 6-3 and Figure 6-4. There is no indication of hydrogen-bonding to the nitrogen group, with the first peaks for $g(O_wN_s)$ all being around 5Å with MeP once again showing the greatest structuring effect and MeCap causing similar behaviour to the MeVam fragment. However, the behaviour of the EtP is anomalous displaying two consecutive peaks at 4.88Å and 5.69Å. The $g(H_wN_s)$ for EtP is also different from the other three systems, with the maxima less rounded and at a greater distance. For MeP**,** the broad peak in $g(O_wN_s)$ at 5.23Å is indicative of the solvation shell being occluded by other parts of the solute molecule. For the larger rings (MeVam and MeCap) there seems to be a complete absence of solvation structure, with g(r) being equal to one from about 4.8Å onwards. With this understanding, the characteristics of $g(X_wN_s)$ for EtP in a shifting to longer distances could be explained with the larger ethyl group causing the solvation shell to be further away from the N.

From the study of the RDFs it is clear that the solvation around the oxygen of the fragments considered is in a hydrophilic region with MeP displaying the strongest structuring properties. The solvation around the methyl group, for all systems, was consistent with a hydrophobic region. MeP was seen to have significant structuring effects in both regions whereas other fragments, while able to cause ordering in the solvent were not consistent in both regions and showed differing hydrophilic and hydrophobic properties. This ability to structure both regions could be associated with the inhibiting power of the resultant polymeric inhibitor.

6.2.2 Hydrate distribution around the solutes.

When investigating the role of these molecules in the context of hydrate inhibitors it is helpful to look at the distribution of "hydrate like" waters around the solute. The "hydrate like" water is defined by a series of order parameters that have been developed for this purpose. This will give insight into the mode of action of the hydrate inhibitors by identifying generation of clathrate cages in regions around the solute molecule The presence of hydrate like water can be identified using a number of order parameters developed to investigate hydration structure around alcohols,[305] detailed examination of the hydration structure around solutes can be made. The order parameters employed measure the average angle between the water and oxygen from within a hydrogen-bonded triplet. The order parameter measures how the average angle deviates from the ideal tetrahedral angle and can be used to distinguish between solid and liquid phases. The resulting values can be used to determine if a well-defined tetrahedral hydrogen bonding network, as in gas hydrates, is present. Therefore using the order parameters allows analysis of the water around the solute to discover what hydrate structure, "hydrate-like waters" are present. The order parameters are also used to identify the local structure around a given water molecule as being hydrate-like, ice-like or liquid-like. These provide direct and quantitative information about the extent to which solvation leads to clathrate-like solvation cages.

All three order parameters[306] are derived from a two-level description of the water, in which a central water molecule A is designated and used to identify the set of water molecules $\{B_i\}$ in its solvation shell; the size of the solvation shell was defined using

Equation 6-3.

$$r_s = \sigma(2^{\frac{1}{6}} + 2^{\frac{1}{6}} - 1))$$

Equation 6-3

The location of A was then used to associate the order parameters with a specific hydration behaviour (i.e., solvating a specific group or bulk water). The first-order parameter is a three-body structural parameter, F_3, designed to probe deviations from the angles expected in a tetrahedral hydrogen bonding network. An angle θ may be defined between the oxygen atoms in a B_i-A-B_j triad and then used to calculate F_3 according to the equation

$$F_3 = \langle (\cos\theta|cos\theta| + cos^2\theta_t)^2 \rangle$$

Equation 6-4

where θ_t is the tetrahedral angle. The angular brackets denote an average over all unique pairs $\{B_i,B_j\}$ and over all the designated central atoms A. F_3 will be close to zero for tetrahedrally coordinated water and larger than zero otherwise. The other two order parameters are designed to probe the torsion angles within a hydrogen bonding network and were originally designed as a test for clathrate-like structure.[306] The torsion angles that characterize a hydrogen bonding network may be defined in terms of a quartet of hydrogen-bonded oxygens and in this sense represent a four-body order parameter. However, an analogous parameter can be defined from a water dimer by assuming that those O-H bonds not involved in the dimer hydrogen bond will point toward the next O in the network. This second definition has the computational advantage of requiring a search only through all water dimers, rather than through all water tetramers. Accordingly, the A-B_i dimers were used to define three normalized vectors: the vector between the two O atoms, r_{OO}; and for each of A and B*i*, the vector from the O atom to the H atom not involved in the hydrogen bond, r_{OHA} and r_{OHB} (see Figure 1).

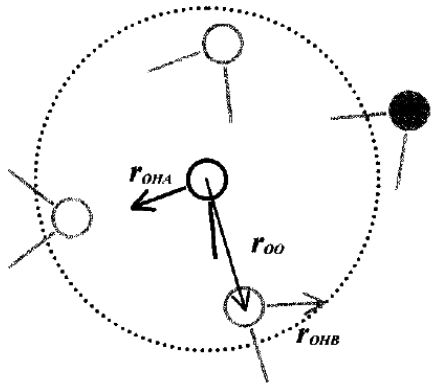


Figure 6-5 Schematic diagram showing how the water hydrogen-bonding network is defined. Water *A* is in black, and its solvation shell is depicted by the dotted circle. Three B_i waters are depicted in grey with open circles for the oxygen atoms. A fourth water molecule (grey with filled oxygen) is not considered as it is outside the A solvation shell. The vectors r_{OHA} and r_{OHB} lie along the OH bonds as shown. The third vector, r_{oo}, joins the two oxygen sites. For this example another two sets of three vectors would be used to define the F_4 parameters for water molecule A.

To provide a unique definition, the hydrogen furthest from the oxygen on the other molecule was always used to define the r_{OH} vectors. These vectors were then used to calculate order parameters based on the torsion angle, φ, and the vector triple product:

$$F_{4\varphi} = \langle \cos 3\varphi \rangle$$

Equation 6-5

$$F_{4\varphi} = \langle (r_{OHA} \times r_{oo} \cdot r_{OHB})^2 \rangle$$

Equation 6-6

Note that the use of cos 3φ for $F_{4\varphi}$ arises from the 3-fold rotational symmetry expected for torsion angles in a tetrahedral network. $F_{4\varphi}$ is about 0.7 for a clathrate-like structure (which is dominated by flat five- and six-membered rings), close to 0 for liquid-water,[306] and about -0.3 for ice-like structure.[307]

Analysis was carried out by initially observing the make up of hydrate waters around the MeP fragment, and then comparing the four systems, simulations 1-4.

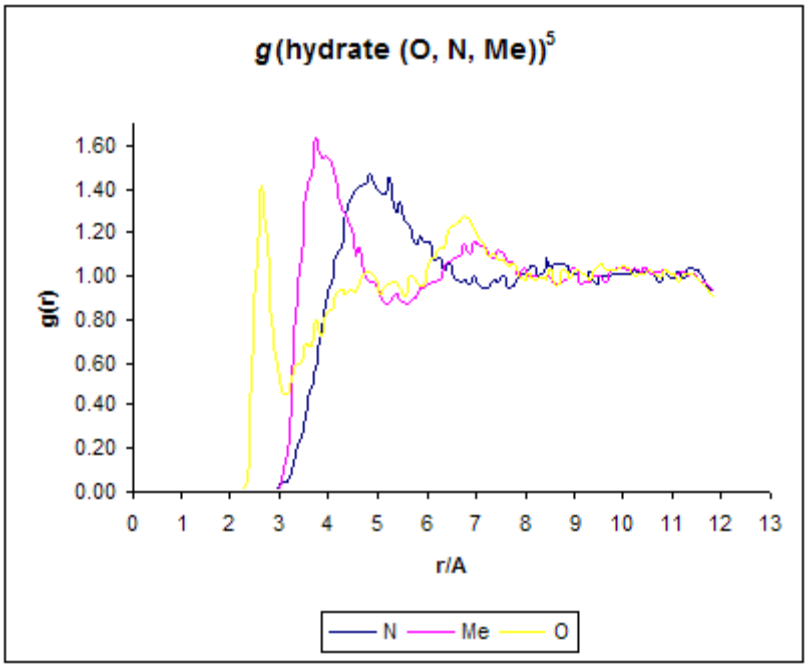


Figure 6-6 Hydrate like waters around the MeP solute

Before describing the observed structure it is worth mentioning that the resolution of the data used for this analysis is much lower due to there being many fewer hydrate like water molecules than the total number of waters. In order to get the resolution seen it was essential to improve the statistics and this was achieved by doubling the length of simulation. Even so there is considerably more noise in the resulting RDFs than was observed in Figure 6-3 and Figure 6-4.

When considering $g(O_{hydrate}\ O_S)^5$ the two peaks observed coincide well with those obtained for $g(O_wO_s)^5$ at 2.65Å and 6.75Å. However, there is a reduction in the intensity of the most local peak and increase for the peak representing structure further from the O. This demonstrates the order within the hydrate system which is greater than that of bulk water. The hydrate waters around the methyl group are also at the same distance but much more intense than the standard RDFs value. Unlike with the O, the secondary structure remains about the same due to the H-structure, showing that it is a secondary solvation shell being caused by the N. Hydrates around the N behave in much the same way as seen in $g(O_wO_s)^5$ but again are more ordered than the

average waters distribution with a greater value of $g(r)$. From this evidence it seems that the distribution of hydrate-like waters around the solute is highly ordered. The remaining three simulations describe the effect that structural variables have on this order.

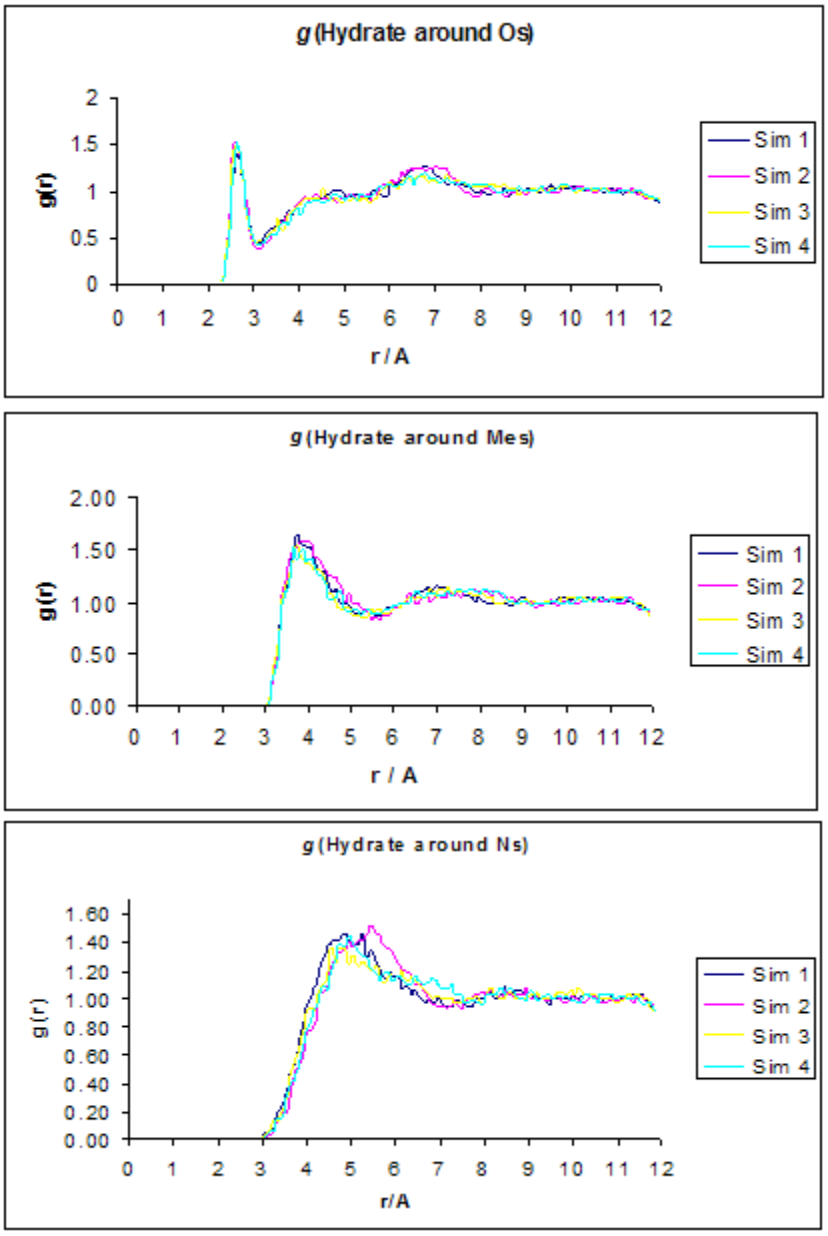


Figure 6-7 RDF plots of hydrate like waters for all inhibitors A,B,C and D around components of the inhibitor fragments.

Structural analysis of g(Hydrate O_s) for all systems, Figure 6-7 **RDF plots of hydrate like waters for all inhibitors A,B,C and D around components of the inhibitor fragments.**, shows the hydrate waters to be present in the solvation shells seen earlier in $g(O_wO_s)$. It does

however highlight the difference between the contributions to order in the first solvation shell caused by the different solutes. MeP exhibits a lower value for $g(r)$ of 1.41 compared with 1.58 in $g(O_wO_s)^5$ at the first maxima, the difference must be accounted for by O_w of different types being present in this solvation shell. The structure at about 7Å is consistent with $g(O_wO_s)^5$ once again showing the MeP and EtP to have long range influences.

When considering $g(O_{hydrate}Me_s)$ the trends and peaks coincide well, with each system having the same relative effect as before. The main difference is that all systems have greater intensity values for $g(r)$. That is the first peak changes for example from $g(r)$ 1.37 to 1.55 for MeP showing a preference of hydrate in this solvation shell.

The hydrate around the nitrogen is, as before, the least consistent between each of the systems. The structure attributable to the EtP system is at a greater radius than the others due to the greater excluded volume. Nitrogen being within the ring of the molecule is the most vulnerable of the three groups being studied to changes in molecular properties that may affect characteristics such as excluded volume, ring strain, and ring flexibility. It is for these reasons that studies of the nitrogen are less obvious and less consistent than for the oxygen and methyl in the fragment molecules.

The differences in the first peaks noted above can be seen more clearly by counting the number of waters that contribute to the first solvation shell. In particular the ratio of the areas in the first peak of $g(O_{hydrate},X)$ / g (O_w,X) gives the mole fraction of hydrate-like waters in the solvation shell of X. This mole fraction can be compared with a similar calculation for bulk water. The difference between these two mole fractions, $X_{solv} - X_{bulk}$, is a measure of the enhancement or disruption of water structure in the solvation shell. The results of these calculations are given in Figure 6-8.

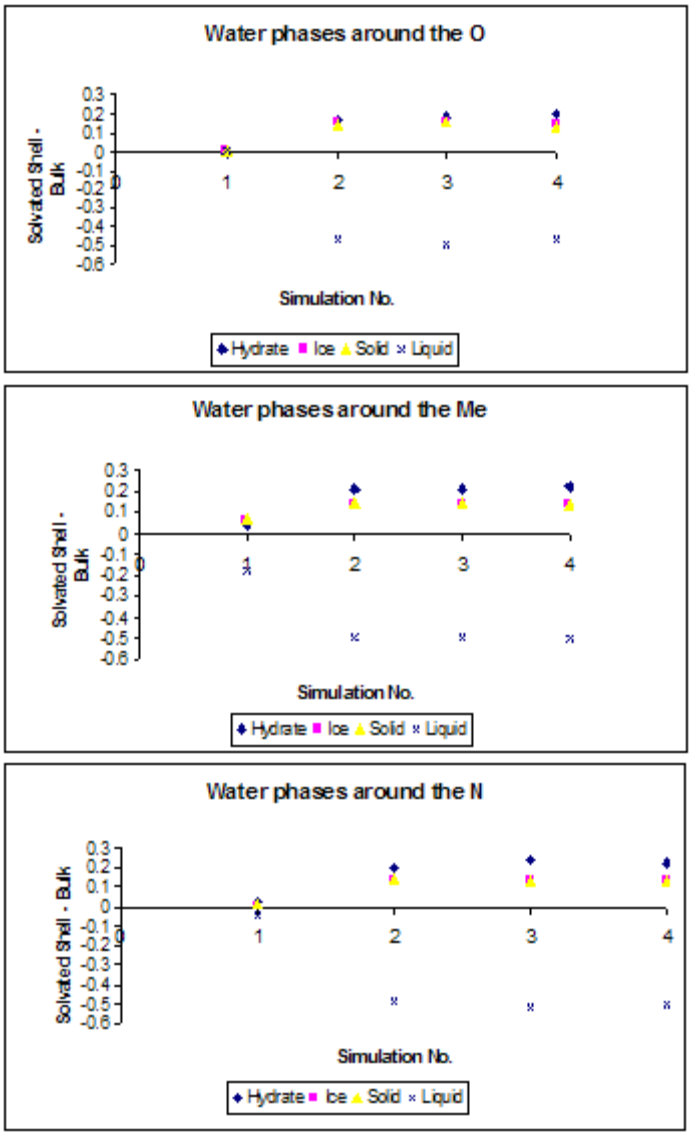


Figure 6-8 The difference between the population of water types within the solvation shell and the bulk simulation for O, Me and N.

For the solvation shell around the MeP O_1 there was very little difference between solvated and bulk water demonstrating a slight increase in hydrate and a slight decrease in ice, solid, and liquid, is seen for the four systems compared with that seen in the bulk of the experiments. However, simulations 2, 3 and 4 show considerably lower values for liquid waters around the oxygen and an increase in the other phases, relative to the bulk value, Figure 6-8. The methyl group exhibits an increase in structured waters, relative to the bulk, with all four systems showing in most cases at least double the difference between solvation shell and bulk for the

amount of ice water in the solvation shell. As much as five times the hydrate like water for simulations 2, 3 and 4 is observed compared with the bulk values. MeP is anomalous in showing only twice the bulk value for the hydrate water in its solvation shell. Observations about the nitrogen in all four systems again show deviations from the bulk values. The nitrogen belonging in the MeP molecule showed much less of a difference between its solvation shell and the bulk water than simulations two, three and four. The EtP and CAP simulations both produced very similar values for the number of waters of each type within solvation shells. The variation for the six member system is a demonstration of its different structuring properties and could be related to it not being a recognized hydrate inhibitor.

6.2.3 Spectroscopic properties: the power spectra.

Power spectra for all four systems were obtained from Fourier transform of the velocity autocorrelation functions for the MeP, N, and O atoms, and are presented in Figure 6-9. These power spectra would correlate with the IR spectrum if they were available. As with infrared, where a spectrum is collected by passing a beam of infrared through the samples, the power spectra is generated showing at with wavelengths the sample absorbs. Analysis of these absorption characteristics reveals details about the molecular structure of the sample. The power spectra is useful in showing likenesses in the effect of different inhibitors on water-water interactions which when ordered into such structures as a “solvent cages” will then dictate the frequency of the oscillatory motion rather than the inertia of the solute.

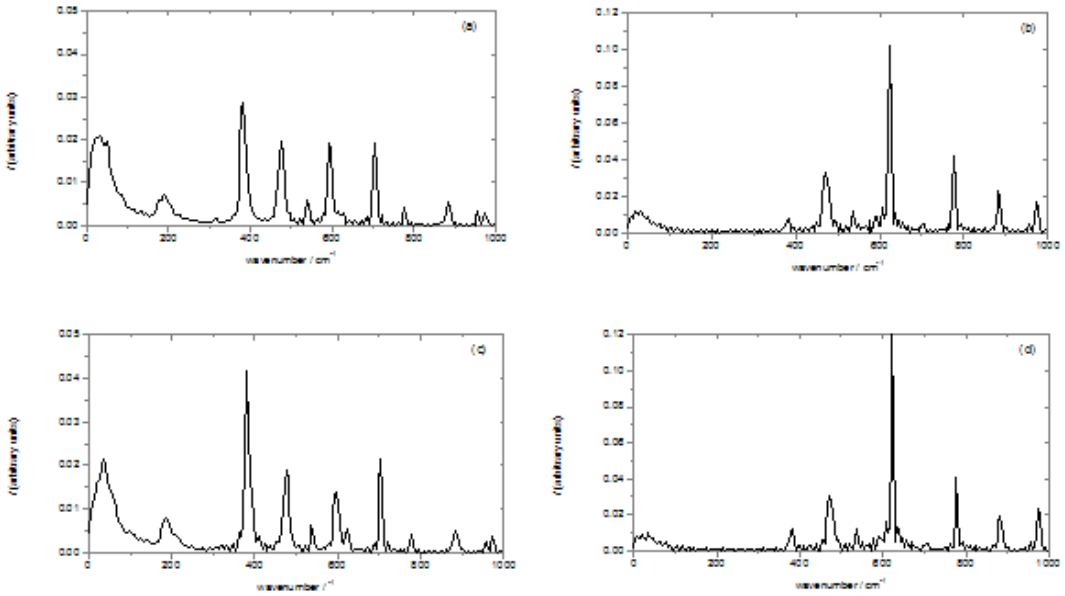


Figure 6-9 (above) Vibrational density of states for (a) the MeP oxygen (b) the MeP nitrogen (c) the MeP oxygen, analysis of part simulation and (d) the MeP nitrogen, analysis of part simulation. Where part analysis is described the spectra displays the data for just the final 400ps of the 800ps simulation.

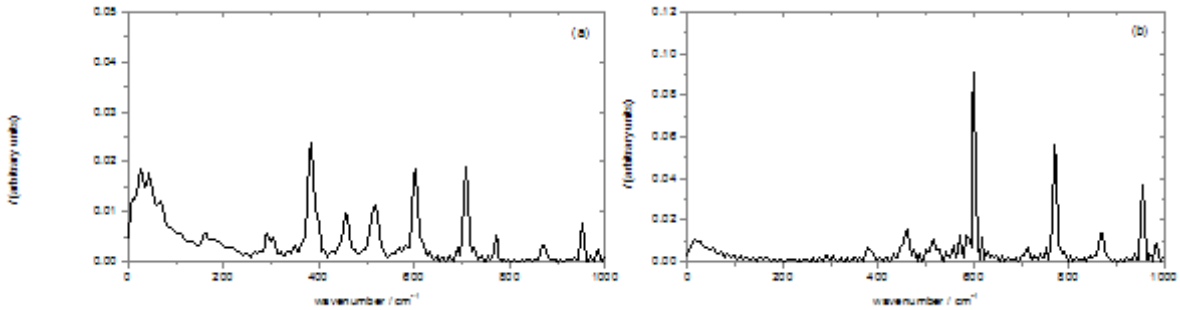


Figure 6-10 (above) Vibrational density of states for (a) EtP oxygen and (b) the EtP nitrogen atoms.

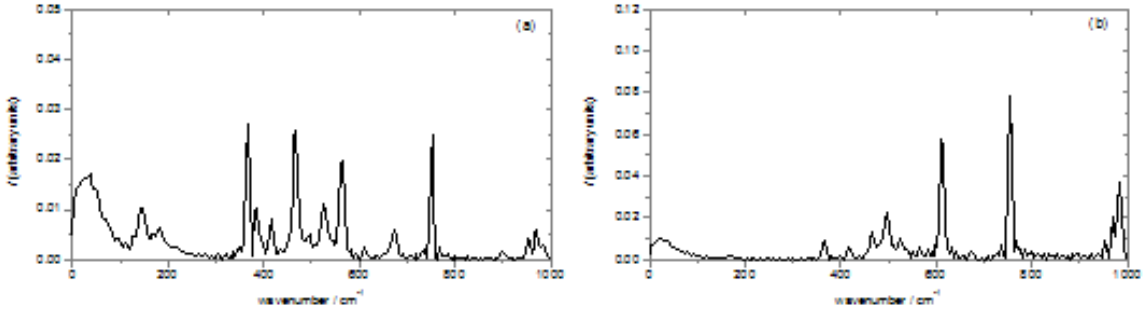


Figure 6-11 (above) Vibrational density of states for (a) the MeVam oxygen and (b) the MeVam nitrogen.

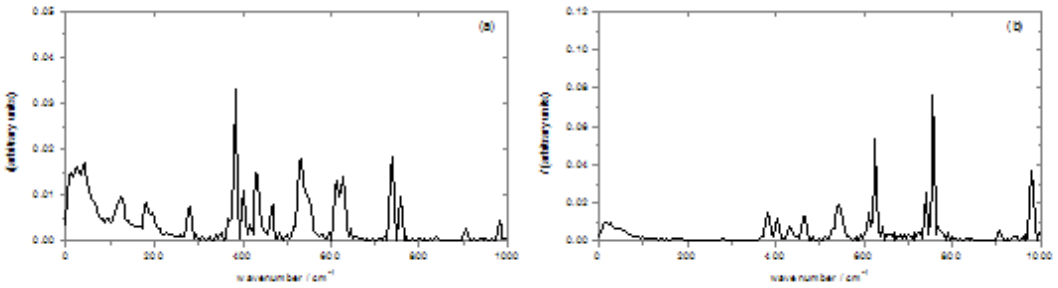


Figure 6-12 (above) Vibrational density of states for (a) the MeCap oxygen and (b) the MeCap nitrogen.

The pyrrolidone ring is very nearly rigid and so the modes seen in the N and O power spectra should identify the molecular translation and rotation modes. Some additional structure due to C=O stretches and bond angle bends will be present in the O power spectrum. Figure 6-9 shows the power spectra for the MeP oxygen (a) and nitrogen (b) and also demonstrates a similarity between peaks for both parts of the molecule. However for the simulations in this project the peaks do not coincide to the same degree as those in the literature[303].

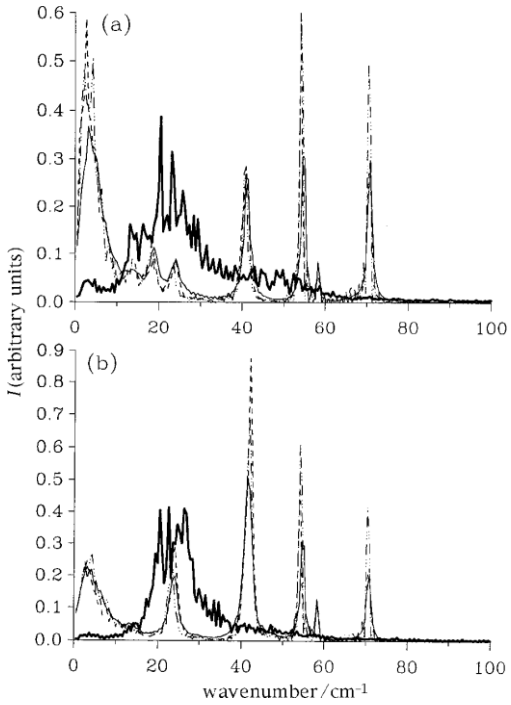


Figure 6-13 Vibrational density of states for (a) the MeP oxygen and (b) the MeP nitrogen atoms. Spectra were calculated from simulations using different solute-solvent interactions, as definded in Table 1[303] : View series (1) _____ for comparison with Figure 6-9 (above) Vibrational density of states for (a) the MeP oxygen (b) the MeP nitrogen (c) the MeP oxygen, analysis of part simulation and (d) the MeP nitrogen, analysis of part simulation. Where part analysis is described the spectra displays the data for just the final 400ps of the 800ps simulation.

For these values Carver *et al.* assigned them to oscillations in a water cage. For this study the extra peaks that are seen can be assigned to the internal modes of vibration, so they were not present when the bonds were constrained. The main reason for this difference is that none of the bonds within the MeP, or other fragments, were constrained in this study and so the oxygen and nitrogen power spectra will also contain intramolecular virbrational modes. The MeP oxygen spectrum above 200 cm^{-1} is dominated by four distinct peaks at 382, 476, 594 and 703 $cm^{-1}$with a peak at 883cm^{-1} with much lower intensity. There are also four peaks at 472, 623, 777 and 887 cm^{-1} for the MeP nitrogen. The sharp nature of these peaks means that they must correspond to well-defined oscillatory motions of the MeP.

Investigation of the variation from the literature of intensity and the shift of the peaks were carried out by analysing just the final 400ps of the simulation. The results for this shortened analysis, Figure 6-8 (c) and (d) show peaks represented by the same wavenumber but showing a definite change in the relative intensities of the peak. The change in intensities was much more for the spectra of the oxygens than that of the nitrogens. This demonstrates that there is some noise in the peak intensities and longer simulations are needed to remove any anomalous behaviour that is not representative of the system being simulated. Other reasons for variation in the results could be investigated by seeing the effect that conditions such as pressure, the water model used, and the size of system may have. Although variation is present the ratio of the wavenumbers for the flexible MeP, 1 : 1.32 : 1.88 are comparable to the ridged literature MeP of, 1 : 1.34 : 1.73 respectively, and allow the conclusion that the presence of these peaks is at least partly due to solvent effects, and may indicate association between solute and solvent or a highly structured solvation shell that creates an anisotropic cage in which the solute motion occurs.

When considering EtP oxygen, Figure 6-10 (a), three peaks at 382, 602 and 708 cm^{-1} are noticeably of a higher intensity with the middle peak having shifted compared with Carvers results. The ratios of the wave numbers are 1 : 1.57 : 1.85 and show a similar trend to MeP. The EtP nitrogen, Figure 6-10(b) has three intense peaks at 602, 769 and 952 cm^{-1} with two lower intensity peaks at 460 and 952 cm^{-1} the latter of which coincides with a peak on the EtP oxygen. Relative to the MeP analysis the sharp peaks seen for EtP coincide well indicating that the properties of the key functional groups are comparable and so with regards to their ability to induce structure in the solvent it is appropriate to consider MeP as a reliable fragment by which to model the MeP polymer.

The MeVam system (Figure 6-11) produces a spectrum with the main features being comparable with those of the MeP spectra. For both the nitrogen and oxygen spectra there is considerably more noise than for the MeP. With the increase in ring size allowing a greater range of intramolecular vibrations the loss of definition is not unexpected, but could be removed by using rigid, or semi-rigid solute in the simulations. The peaks at 366 and 753 cm^{-1} are present on both the oxygen and nitrogen spectra. With other well defined peaks such as that at 565 cm^{-1} present on the oxygen spectra not being observed on the nitrogen spectra. The difference in spectra demonstrates that as the ring size is increased the nitrogen and oxygen behaviour become less similar.

In keeping with the above observation analysis of spectra for CAP, Fig 10 (a) and (b), it is expected that CAP, having a larger ring with increased flexibility, will display less well defined oscillatory motions. The spectra do display more interference and less well defined peaks than those obtained from the analysis of MeP. However, five peaks can be seen to coincide very well 382, 533, 627, 741 and 757 cm^{-1}. This demonstrates that CAP, although large and flexible, does have a relationship between the nitrogen and oxygen oscillations. With a maximum of six

vibrations describing oscillatory motions the presence of many significant peaks for both CAP and the six membered ring could be describing effects of a different nature.

6.3 Conclusion

In this section we present a detailed analysis of simulations of methylpyrrolidone (MeP) ethylpyrrolidone (EtP) methylvalerolactam(MeVam) and methylcaprolactam (MeCap) in water. These systems are particularly interesting because pyrrolidone and caprolactam are common constituents of clathrate hydrate inhibitors, and are believed to have significant structuring effects on water. Simulations of MeP and MeCap gave conclusive results in agreement with those published with Carver. The MeVam, which is inactive as a hydrate inhibitor, was simulated to highlight the characteristics of the known hydrate inhibitors MeCap and MeP. EtP was simulated and confirmed that MeP is a suitable fragment to model as a representative of PVP.

The structural analyses support hypotheses that the pyrrolidone and caprolactam moieties are structure making solutes. The simulations provide evidence for well-defined solvation shells around all four solutes. Throughout the analysis particular attention has been paid to the expected main centres of hydrophobic solvation (the methyl substituent) and the hydrophilic hydration (the solute O). Well-defined sharp peaks for water around O_s are indicative of hydrogen bonds between the solute and first solvation shell. The water distribution around the Me group suggests it is surrounded by a hydrate like complex with hydrogen bonds between molecules. All four solutes were observed to share the same characteristics around the O and N sites with the difference in the solutes structuring ability being demonstrated by the intensity of the RDF plots.

The four systems are seen to differ in their ability to induce long range structuring effects as well as the structuring resulting from the N. The effects from the N are inconclusive due to the difference in excluded volume between the four solute molecules.

Structural analysis of hydrate like waters using order parameters displayed behaviours characteristic of hydrophilic solvation around the O and hydrophobic solvation around the Me. Comparisons of the mole fractions of hydrate-like waters in the first solvation shell with that of bulk water showed the enhancement or disruption of water structure in the solvation shell. When observing these values MeP was obviously different to the other three systems.

Evidence for the enhanced water structure was also found from the power spectrum calculated for the four systems. The spectra all showed sharp peaks due to the solvent effect but were seen to suffer an increase in noise, with the larger solute rings, as there was an increase in intramolecular vibrational modes.

We conclude that the increased structure of the solvated water is due to association between neighbouring water molecules around the inhibitor solutes. The MeP and MeCap were seen to demonstrate similar characteristics to MeVam but of a different magnitude. The MeP and EtP simulations were similar enough to conclude that the MeP is a good representative of the PVP inhibitor.

6.4 Future Work

Following this section of study an increase in awareness and understanding of hydrates, hydrate inhibitors and computer simulation has caused me to recognise areas which, with more time could be expanded and in turn increase understanding of these systems.

Increasing the length of simulations would improve the level of statistics and remove noise from analysis such as that carried out to calculate the structure parameters. This would improve the resolution at important features like maxima which are representative of the solvation shell, making final conclusions more accurate.

7 References

[1] J.D. Watson, F.H.C. Crick, (1953) *Nature*, **171**, 737-738

[2] U. Nagaswamy, N. Voss, Z. Zhang, G.E. Fox, (2000) *Nuc. Acids. Res.*, **28**, 375-76

[3] R. Ahmad, H. Arakawa, H.A. Tajmir-Riahi, (2003) *Biophysical Journal*, **84**, 2460-2466

[4] K. Bryson, R.J. Greenall, (2000) *J. Biomol Struct Dyn*, **18**, 393-412

[5] W. Saenger (1984) *Principles of Nucleic Acid Structure*

[6] Stryer et al. (2002), *Biochemistry*, 5th Edition. New York: WH Freeman and Company, Section 21.7

[7] R.H.T. Bates, G. A. McKinnon, R.P. Millane, G.A. Rodley, (1980) *J. Phys*., **14**, 233–252

[8] G.A. Rodley, R.P. Millane, G.A. McKinnon, R.H.T. Bates, (1984) *J. R. Soc. N. Z*., **14**, 223–231

[9] A.R. Stokes, (1955), *Prog. Biophys. Biophys. Chem*., **5**, 140–167

[10] T. Banerjee, R. Mukhopadhyay, (2008), *Bioochem. and Biophys. Res. Comm.,* **374**, 264-268

[11] H. Wackerbarth, M. Grubb, J. Wengel, I. Chorkendorff, J. Ulstrup, (2006), *Surface Sci.,* **600**, 122-127

[12] C. Delmonte, (1997) *Advances in AFM and STM Applied to the Nucleic Acids*, Invention Exploitation Ltd., UK

[13] T.C. Marsh, J. Vesenka, E. Henderson, (1995), *Nucleic Acids Res*., **23**, 696–700

[14] C.S. Delmonte, l.R.B.Mann, (2003) *Current Science*, **85,** 1565-1570

[15] Y.J. Liu, W.J. Mei, L. Zu, H.J. Zhao, H. LX, W. FH, (2008) *J. Coord. Chem.*, **61** 3213-24

[16] M. Hannon (2007) *Chem. Soc. Rev*, **36**, 280 - 295

[17] A. Papakyriakou, G. Malandrinos, A. Garoufis. (2006) *J. Inorg. Biol,.* **100,** 1842-1848

[18] A.Rodger, et al. (2006) *Phys. Chem. Chem. Phys*, **8**, 3161-3171

[19] R.E. Dickerson, et al. (1982) *Science*, **216**, 475-485

[20] D. Ussery et al. (2002) *Computers & Chemistry*, **26**, 531-541

[21] P. Christoph Champ, T. Tim, Binnewies, N. Nielsen, G. Zinman, K. Wu, J. Bohlin, D.W.Ussery (2006) *Microbiolgy*, **152**, 579-583

[22] A.H.J. Wang, G.J. Quigley, F.J. Kolpak, J.L. Crawford, J.H. Van Boom, G. Van der Marel, A. Rich A (1979) *Nature*, **282**, 680-686

[23] D. Ho, Y. Kim, A. Rich (2002) *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 16666-71

[24] R.L. Ornstein, R. Rein, D.L. Breen, R.D. MacElroy, (1978) *Biopolymers*, **17**, 2341-2360

[25] Kabsch, Sander, Trifonov, (1982) *Nucleic Acids Research*, **10**, 1097-1104

[26] J. Sponer, H.A. Gabb, H.A. Leszcynski, J. Hobza, (1997) *Biophysical Journal*, **73**, 76-87

[27] M. Feig, M. Pettitt, (1999) *J. Mol. Biol.*, **286**, 1075-1095

[28] M. Nagaoka, Y. Sugiura, (2000) *J. Inorg. Biochem..*, **82**, 57-63

[29] L.H. Hurley, (2002) *Nat. Rev. Cancer.*, **2**, 188-200

[30] B.S.P. Reddy, S. K. Sharma, J. W. Lown, (2001) *Curr. Med. Chem.*, **8**, 475-508

[31] R. Eckel, R. Ros, A. Ros, S. D. Wilking, N. Sewald, and D. Anselmetti, (2003) *Biophys. J*., **85**, 1968-1973

[32] J.E. Coury, L. McFail-Isom, L.D. Williams, L.A. Bottomley. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 12283-12286

[33] A. Gelasco, and S. J. Lippard, (1998) *Biochemistry*, **37**, 9230-9239

[34] T.E. Cheatham, M.A. Young, (2000) *Biopolymers*, **56**, 232-256

[35] D.L. Beveridge, G. Ravishanker, (1994) *Curr. Opin. Struc. Biol.*, **4**, 246-255

[36] Cheatham, T. E. (2004) *Curr. Opin. Struc. Biol*, **14**, 360-367.

[37] S. Khalid., P.M. Rodger. (2007) *J. Biol. Inorg. Chem.*, **12**, S7-S52

[38] L.S. Lerman, (1961) *J.Mol.Biol.*, **3**, 18

[39] M. Waring, (1970) *J.Mol.Biol.*, 2, 247

[40] C. Temperini *et al.* (2003) *Nucleic Acids Research*, **31**, 1464-1469

[41] A. Wang *et al.* (1987) *Biochemistry*, **26**, 1152–1163

[42] C. Temperini *et al*. (2003) *Nucleic Acids Research*, **31**, 1464-1469

[43] T.R. Krugh, (1994) *Curr. Opin. Struct. Biol.*, **4**, 351

[44] S. Neidle, (1997) *Biopolymers*, **44**, 105

[45] D.E. Wemmer, P.B. Dervan, (1997) *Curr. Opin. Struct. Biol.*,**7**, 335

[46] S. Neidle, (1994) '*DNA structure and recognition'*, Oxford University Press.

[47] E.F. Gale, E. Cundliffe, P.E. Reynolds, M.H. Richmond, M.J. Waring, (1981) '*The Molecular Basis of Antibiotic Action'*

[48] C. Zimmer, U. Wahnert, (1986) *Prog. Biophys. Mol. Biol*,.**47**, 1986, 31-112

[49] M.L. Kopka, C. Yoon, D. Goodsell, P. Pjura, R.E. Dickerson, *The molecular origin of DNA-drug specificity in netropsin and distamycin*

[50] Y.Z. Chen, E.W. Prohofsky, (1995) *Biopolymers*, **35**, 657

[51] C. Zimmer, N. Kakiuchu, W. Guschlbauer, (1982) *Nucleic Acids Res.*, **10**, 1721

[52] J. Lah, G. Vesnaver, (2000) *Biochemistry*, **39**, 9317

[53] F. Gago, W.G. Richards, (1989) *Mol. Pharmacol.*, **37**, 341

[54] R.E. Dickerson, H.R. Drew, (1981) *J. Mol. Biol.* **149**, 761

[55] H.R.Drew, R.E. Dickerson, (1981) *J. Mol. Biol.* **151**, 535

[56] R.Wing, H. Drew, T. Takano, C. Broka, S. Takana, K. Itakura, R.E. Dickerson, (1980), *Nature*, **287**, 755

[57] H.R. Drew, R.E. Dickerson,(1981) *J. Mol. Biol.*, **151**, 535

[58] X. Shui, L. McFail-Isom, G.G. Hu, L.D. Williams. (1998) *Biochemistry*, **37**, 8341

[59] D. Hamelberg, L. McFail-Isom, L.D Williams, W.D. Wilson, (2000) *J. Am. Chem. Soc.*, **122**, 10513

[60] V. Tereshko, G. Minasov, M. Egli, (1999) *J. Am. Chem. Soc.*, **121**, 3590

[61] M. Feig, B.M. Pettitt, (1999) *Biophys. J.*, **77**, 1769

[62] B. Wellenzohn, W. Flader, R.H. Winger, A. Hallbrucker, E. Mayer, K.R. Liedl, (2002) *Bioploymers*, **61**, 276

[63] T. Darden, D. York, L. Pederson, (1993) *J. Chem. Phys.*, **98**, 10089

[64] D. Hamelberg, L. McFail-Isom, L.D. Williams, W.D. Wilson, (2000) *J. Am. Chem. Soc.* **122,** 10513

[65] S.B. Singh, P.A. Kollman, (1999) *J. Am. Chem. Soc.*, **121**, 3267

[66] S. Miyamoto, P.A. Kollman, (1993) *P. Natl. Acad. Sci. USA*, **90**, 8402

[67] V.K. Misra, B. Honig, (1995), *P. NAtl. Acad. Sci. USA.*, **92**, 4691

[68] M.K. Teng, N. Usman, C.A. Frederick, A.H. Wang, (1988) *J. Nucleic Acids Res.*, **16**, 2671

[69] P.E. Pjura, K. Grzeskoniak, R.E. Dickerson, (1987) *J. Mol. Biol.*, **197**, 257

[70] M.S. Searle, K.J. Embrey, (1990) *Nucleic Acids Res.*, **18**, 3753

[71] S.R. Patel, L.K. Kvols, J. Rubin, M.J. O'Connell, J.H. Edmonson, M.M. Ames, J.S.Kovach, (1991) *Invest. New Drugs*, **9**, 53

[72] R.R. Tidwell, S.K. Jones, N.A. Naiman, L.C. Berger, W.B. Brake, C.C. Dykstra, J.E. Hall, (1993) *Antimicrobial Agents & Chemotherapy*, **37**, 1713

[73] P. Giri, G.S. Kumar, (2008) *Archives of BioChem. And BioPhys.*, 183-92

[74] M. Rahimian, Y. Miao, W.D. Wilson, (2008) *J. Phys. Chem.*, **29**, 8770 - 78

[75] C.R. Kissinger, B. Liu, E. Martin-Blanco, T.B. Kornberg, C.O. Pabo, (1990) *Cell*, **63**, 579

[76] D.A. Gilbert, J. Feignon, (1991) *J. Biochemistry*, **30**, 2483

[77] Y.H. Chen, Y. Yang, J.W. Lown, (1996) *Biochem. Biophys. Res. Commun.* **220**, 213

[78] S.A Harris, E. Gavathiotis, M.S. Searle, M. Orozco, C.A. Laughton, (2001) *J. Am. Chem. Soc.*, **123**, 12658

[79] J. Shlitter, (1993) *J. Chem. Phys. Lett.*, **215**, 617

[80] B. Wellenzohn, W. Flader, R.H. Winger, A. Hallbrucker, E. Mayer, K.R. Liedl, (2001) *Biophys. J.*, **81**, 1588

[81] J. Parkinson, S. Sadat-Ebrahimi, A. Wilton, J. McKie, J. Andrews, K.T. Douglas, (1995) *Biochemistry*, **34**, 16240

[82] C.E. Bostock-Smith, K.J. Embrey, M.S. Searle, (1997) *J. Chem. Soc. Chem. Commun.*, **1**, 121

[83] C.Bailly, G. Chessari, C. Carrasco, A. Joubert, J. Mann, W.D. Wilson, S. Neidle, (2003) *Nucleic Acids Res.*, **31** 1514

[84] A. Rodger, K.J. Sanders, M.J. Hannon, I. Meistermann, A. Parkinson, D.S. Vidler, I.S. Haworth, (2000) *Chirality*, **12**, 221

[85] E.M. Proudfoot, J.P. Mackay, P. Karuso, (2001) *Biochemistry* **40** 4867

[86] W.Li, C.Zhao, C. Xia, W.E. Antholine, D.H. Petering, (2001) *Biochemistry*, **40**, 7559

[87] K.E. Erkkila, D.T. Odom, J.K. Barton, (1999) **99**, 2777

[88] J.K. Barton, J.M. Goldberg, C.V. Kumar, N.J. Turro, (1986) *J. Am. Chem. Soc.*, **108**, 2081

[89] C. Hiort, B. Norden, A. Rodger,(1990) *J. Am. Chem. Soc.*, **112**, 1971

[90] I.S. Haworth, A.H. Elcock, J. Freeman, A. Rodger, W.G. Richards, (1991) **9**, 23

[91] P. Yang, D. Han, Z. Xiong, (2001) *J. Mol. Struct. (Theochem)*, **540**, 211

[92] B. Lippert, ed. (1999) Cisplatin, *Chemistry and Biochemistry of a Leading Anti-Cancer Drug*

[93] D.Z. Coggan, I.S. Jaworth, P. Bates, A. Robinson, A. Rodger, (1999) *Inorg. Chem.*, **38**, 4486-4497

[94] B. Onfelt, P. Lincoln, B. Norden, (1999) *J. Am. Chem. Soc.*, **121**, 10846-10847

[95] K.E. Erkkila, D.T. Odom, J.K. Barton, (1999) *Chem. Rev*., **99**, 2777-2795

[96] S.C. Harrison, A.K. Aggarwal, (1990) *Annu. Rev. Biochem.*, **59**, 933

[97] A.klug, D.Rhodes, (1987) *Trends Biochem. Sci.*, **12**, 464

[98] M. Elrod-Erickson, T. Benson, C. Pabo (1998) *Structure*, **6**, 451

[99] B. Rosenberg, L. Van Camp, J.E. Trosk, V.H. Mansour, (1969) *Nature*, **222**, 385

[100] P. Horacek, J. Drobnik, (1971) *J. Biochem. Biophys. Acta.*, **254** 341

[101] S.E. Sherman, A. Gibson, A.H. Wang, S.J. Lippard, (1988) *J. Am. Chem. Soc.* **110**, 7368

[102] C. Manzotti, G. Pratesi, E. Menta, R. Di Domenico, E. Cavalletti, H.H. Fiebig, L.R. Kelland, N. Farrell, D. Polizzi, R. Supino, G. Pezzoni, F. Zunino, (2000) *Clin. Cancer Res.*, **6**, 2626

[103] J.W. Trauger, E.E Baird, P.B. Dervan, (1996) *Nature*, **382**, 559

[104] K. Miaskiewicz, R. Osman, H. Weinstein, (1993) *Am. Chem. Soc*., **115**, 1526-1537

[105] E. Stofer, C. Chipot, R. Lavery (1999) *J. Am. Chem. Soc.,* **121**, 9503-9508

[106] J. Norberg, L.Nilsson, (1998) *Biophys. J.,* **74**, 394-402

[107] E. Giudice, P. Varnai, R. Lavery. (2001) *Chem. Phys. Chem.*, **2**, 673-677

[108] D. Bandyopadhyay, D. Bhattacharrya, (2000) *J. Biomol. Struct. Dyn.* **18**, 29-43

[109] N. Pastor, A. D. MacKerell, H. Weinstein (1999) **16**, 787-810

[110] S. Derreumaux, S. Fermandiian (2000) *Biophys. J*. **79**, 656-669

[111] T. E. Cheatham, J. Srinivassan, D. A. Case, P. A. Kollman (1998) *J. Biomol. Struct. Dyn.* **16**, 265-280

[112] L. Trantirek, R. Stefl, M. Vorlickova, J. Koca, V. Sklenar, J. Kypr (2000) *J. Mol. Biol.* **297**, 907-922

[113] R. Stefl, L. Trantirek, M. Volickova, J. Koca, V. Sklenar, J. Kypr (2001), *J. Mol. Biol.* **307**, 513-524

[114] R. H. Winger, K. R. Liedl, A. Pichler, A. Hallbrucker, E. Mayer, (2000) *J. Phys. Chem. B* **104**, 11349-11353

[115] R. Soliva, F. J. Luque, C. Alhambra, M. Orozco (1999) *J. Biomol. Struct. Dyn.* **17**, 89-99

[116] D. Sprous, M. A. Young, B. L. Beveridge (1998) *J. Phys. Chem. B.* **102**, 4658-4667

[117] T. E. Cheatham, J. Srinivasan, D. A. Case, P. A. Kollman (1998) *J. Biomol. Struct. Dyn.* **16**, 265-280

[118] B. Jayaram, D. Sprous, M. A. Young, D. L. Beveridge (1998) *J. Am. Chem. Soc.* **120**, 10629-10633

[119] Case, D. A., D. A. Pearlman, J. W. Caldwell, T. E. Cheatham III, W. S. Ross, C. Simmerling, T. Darden, K. M. Merz, R. V. Stanton, A. Cheng, J. J. Vincent, M. Crowley, D. M. Ferguson, R. Radmer, G. L. Seibel, U. C. Singh, P. Weiner, and P. Kollman. (1999). *AMBER: Version 6*. University of California, San Francisco

[120] W.D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. (1995) *J. Am. Chem. Soc.* **117**, 5179–5197

[121] H. Arthanari, K. J. McConnell, R. Beger, M. A. Young, D. L. Beveridge, P. H. Bolton. (2003). *Biopolymers.* **68**, 3–15

[122] D.R. Bevan, L. Li, L. G. Pedersen, T. A. Darden. (2000) *Biophys. J.* **78**, 668–682

[123] D.L. Beveridge, G. Barreiro, K. S. Byun, D. A. Case, T. E. Cheatham, S. B. Dixit, E. Giudice, F. Lankas, R. Lavery, J. H. Maddocks, R. Osman, E. Seibert, H. Sklenar, G. Stoll, K. M. Thayer, P. Varnai, and M. A. Young. (2004). *Biophys. J*. **87**, 3799–3813

[124] Beveridge, D. L., G. Barreiro, K. S. Byun, D. A. Case, T. E. Cheatham, S. B. Dixit, E. Giudice, F. Lankas, R. Lavery, J. H. Maddocks, R. Osman, E. Seibert, H. Sklenar, G. Stoll, K. M. Thayer, P. Varnai, and M. A. Young. (2005). *Biophys. J.*, **89**, 3721–3740

[125] H. R. Drew, R. E. Dickerson (1981) *J. Mol. Biol.*, **151**, 535-556

[126] E. Westhof (1988) *Ann. Rev. Biophy. Chem.*, **17**, 125-144

[127] L. A. Labeots, A. M. Weiss (1997) *J. Mol. Biol.*, **269**, 113-128

[128] N. C. Horton, J. J. Perona, (1998) *J. Mol. Biol.*, **273**, 21721-21729

[129] C. R. Robinson, S. G. Sligar (1993) *J. Mol. Biol.*, **234**, 302-306

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

[131] D. Kosztin, T. C. Bishop, K. Schulten (1997) *Biophys. J.*, **73**, 557-570

[132] M. Billeter, P. Guntert, K.Wutrich (1996) *Cell*, **85**, 1057-1065

[133] Y. Nishimura, C. Torigoe, M. Tsuboi (1986) *Nucleic Acids Res.*, **14**, 2737-2748

[134] V. I. Ivanov, L. E. Minchenkova, G. Burckhardt, E. Birch-Hirschfeld, H. Fritzche, C. Zimmer (1996), *Biophys.*, **71**, 3344-3349

[135] Y. Wang, G. A. Thomas, W. L. Peticolas, (1989) *Biomol Struct. Dynam.*, **6**, 1177-1187

[136] M. Feig, M. B. Pettitt, (1997) *J Phys. Chem.*, **101**, 7361-7363

[137] M. A. Young, B Jayaram, D. L. Beveridge (1998) *J. Phys. Chem.* ,**B 102**, 7666-7669

[138] Y. Duan, P. Wilkosz, M.Crowley, J.M.Rosenberg (1997) *J.Mol.Biol*, **272**, 553-572

[139] P. Auffinger, E. J. Westhof (1997) *J Mol. Biol.*, **269**, 326-342

[140] A. D. Mackerell, J. Wiorkiewicz-Juczera, M.Karplus (1995) *J. Am. Chem. Soc.*, **117**, 2564-2566

[141] S.Y. Reddy, F. Leclerc, M. Karplus (2003) *Biophysical Journal*, **84**, 1521-1449

[142] H. Savage, A. Wlodawer (1986) *Methods Enzymol* **127**, 162-183

[143] E. Liepinsh, G. Otting, K. Wuethrich (1992) *Nucleic Acids Res.* **20**, 6549-6553

[144] E. Liepinsh, W. Leupin, G. Otting (1994) *Nucleic Acids Res.* **22**, 2249-2254

[145] W. Saenger, (1987) *Ann. Rev. Biophy. Biophy. Chem.* **16**, 93-114

[146] M. Feig, B. M. Pettitt, (1999) *J. Mol. Biol.* **286**, 1075-1095

[147] B. Schneider, H. M. Berman, (1995) *Biophys J*, **69**, 2661-2669

[148] B. Schneider, K. Patel, H. M. Berman (1998) *Biophys J*, **75**, 2422-2434

[149] B. Schneider, D. Cohen, H. Berman, (1992) *Biopolymers*, **32**, 725-750

[150] B. Schneider, D. Cohen, L. Schleifer, A. R. Srinivasan, W. K. Olson, H. M. Berman (1993) *Biophys. J.*, **65**, 2291-2303

[151] B. Schneider, K. Patel, H. M. Berman, (1998) *Biophys. J.*, **75**, 2422-2434

[152] F. Vovelle, J. M. Goodfellow (1993) *In Water and Biological Molecules,* 244-265

[153] V.P. Denisov, G. Carlstrom, K. Venu and B. Halle, (1997) *J. Mol.Biol*., **268,** 118-136

[154] S.T. Lin, P.K. Maiti, W.A.Goddard, (2005) *J. Phys. Chem. B.*, **109**, 8663-8672

[155] V.P. Chuprina, U. Heinemann, A.A. Nurislamov, P. Zielenkiewicz, R.E. Dickerson, W. Saenger. (1991) *Proc. Natl. Acad. Sci. USA.,* **88**, 593 - 597

[156] T. Schlick, D.A. Beard, J. Huang, D.A. Strahs, X. Qian. (2000) *Comp. in Sci and Eng.,* **2**, 38-51

[157] C. Mura, J.A. McCammon (2008) *Nuc. Acids Res.*, **36**, 4941-55

[158] T. Cheatham et al. (1995) *J. Am. Chem. Soc.*, **11**, 4193-94

[159] P. Auffinger, E. Westhof (2001) *J. Mol. Biol*., **305**, 1057-1072

[160] H. Johannesson, B. Halle (1998) *J.Am. Chem. Soc*., **120** 6859-6870

[161] A. Jacobson, W. Leupin, E. Liepinsh, E, G. Otting (1996) *Nucl. Acids Res.*, **24**, 2911-2918

[162] E. Giudice, R Lavery (2002) *Acc. Chem. Res*., **35**, 350-357

[163] P. Auffinger. E. Westhof (2000) *J. Mol. Biol.*, **300**, 1113-1131

[164] D. Hamelberg, L. McFail-Isom, L.D. Williams, W.D. Wilson (2000) *J. Am. Chem. Soc.*, **122**, 10513-10520

[165] R. Stefl. J. Koca. (2000) *J. Am. Chem. Soc*., **122**, 5025-5033

[166] P. Auffinger. E. Westhof. (2000) *J. Mol. Biol.*, **300**, 1113-1131

[167] P. Auffinger. E. Westhof. (2001) *J. Mol. Biol.*, **305**, 1057-1072

[168] J. K. McConnell. D. L. Beveridge. (2000) *J. Mol. Biol*., **304**, 803-820

[169] M. Feig, B. M. Pettitt, (1999) *Biophys. J*., **77**, 1769 - 1781

[170] M. Feig, B. M. Pettitt, (1999) *J. Mol. Biol*., **286**, 1075-1095

[171] E. Guidice, R. Lavery, (2002) *Acc.Chem.Res*., **35**, 350

[172] G. Roxtröm, I. Velázquez, M. Paulino, O. Tapia, (1998) *J.Phys.Chem.B*., **102**, 1828

[173] X. Daura, B Olivia, E. Querol, F.X. Avilés, O. Tapia, (1996) *Proteins*, **25**, 89

[174] V. Tsui, I. Radhakrishnan, P.E. Wright, D.A. Case, (2000) *J.Mol.Biol*. **302**, 1101

[175] T.E. Cheatham, P. Cieplak, P.A. Kollman, (1999) *J.Biomol.Struct.Dyn*., **16**, 845

[176] W. Jorgensen, J. Chandrasekar, J. Madura, R. Impey, M. Klein, (1983) *J.Chem.Phys*., **79**, 926

[177] M.A. Eriksson, L. Nilsson, (1999) *Eur.Biophys.J*., **28**, 102

[178] Y. Tang, L. Nilsson, (1998) *Proteins: Struct., Funct. Genet*., **31**, 417

[179] Ch.K. Reddy, A. Das, B. Jayaram,(2001) *J.Mol.Biol*., **314**, 619

[180] S. Bandyopadhyay, M. Tarek, M.J Klein, (1999) *J.Phys.Chem.B.*, **103**, 10075

[181] N Foloppe, A.D MacKerell, (1999) *Biophys. J.*, **76**, 3206

[182] R.R. Sinden, (1994) *DNA Structure and Function* (Academic London)

[183] R.E. Dickerson, (1998) *Nucleic Acids Res.*, **26**, 1906

[184] C. Branden, J. Tooze, (1999) *Introduction to Protein Structure* (Garland, New York), 2nd Ed.

[185] B.C. Baguley, (1982) *Mol. Cell. Biochem.* **43**, 167

[186] W.A. Denny, (1989) *Anti-Cancer Drug Des.,* **4**, 241

[187] S.J. Lippard, J.M. Berg (1994) *Principles of Bioinorganic Chemistry* (University Science Books, Mill Valley, CA).

[188] D.Z Coggan, I.S Haworth, P.J. Bates, A. Robinson, A.Rodger (1999) *Inorg. Chem.* 4486

[189] B. Onfelt, P. Lincoln, B. Norden, (1999) *J. Am. Chem. Soc.*, **121**, 10846

[190] K.E. Erkkila, D.T. Odom, J.K. Barton, (1999) *Chem. Rev.*, 2777

[191] M.J. Hannon, C.L Painting, A. Jackson, J. Hamlin, W. Errington, (1997) *Chem. Commun*, 1807

[192] J.-M. Lehn, (1995) *Supramolecular Chemistry: Concepts and Perspectives*

[193] D. Philip, J.F. Stoddart, (1996) *Angew. Chem. Int. Ed. Engl*., **35**, 1155-1196

[194] B. Schoentjes, J.-M Lehn, (1995) *Helv. Chim. Acta.*, **79**, 1-12

[195] I. Meistermann, V. Moreno, J.M. Prieto, E. Moldrheim, E. Sletten, S. Khalid, P.M. Rodger, J. Peberdy. C.J. Issac. A. Rodger, M.J. Hannon, (2002) *Supramolecular Chemistry And Self-assembly Special Feature Chemistry*, **99**, 5069-5074

[196] J.H. Moon, S.K. Kim, U. Sehistedt, A. Rodger, (1996) *Biopolymers*, **38**, 593

[197] R. Lyng, A. Rodger, B. Norden (1991) *Biopolymer,* **31**, 1709

[198] R. Lyng, A. Rodger, B. Norden (1992) *Biopolymer,* **32**, 1201

[199] M. J. Hannon, V. Moreno, M. J. Prieto, E. Molderheim, E. Sletten, I. Meistermann, C. J. Isaac, K. J. Sanders, A. Rodger, (2001) *Angew. Chem. Int. Ed*., **40**, 880 –884

[200] M.J. Hannon. C.L. Painting. A. Jackson. J. Hamblin. W.Errington, (1997), *Chem. Commun*., 1807-1808

[201] M.J. Hannon, V. Moreno, M.J. Moreno, M.J. Prieto, E. Moldrheim, E.Sletton, I. Meistermann, C.J. Isaac, K.J. Sanders, A. Rodger, (2001) *Angew. Chem. Intl. Ed. Engl.*, **40**, 879

[202] M.J. Hannon, V. Moreno, M.J. Preto, E. Moldrheim, E., Sletten, I. Meistermann, C.J. Isaac, K.J. Sanders, A. Rodger (2001) *Angew. Chem., Intl. Ed. Engl*., **40**, 879-884

[203] L.J. Childs, M. Pascu, A.J. Clarke, M.W. Alcock, M.J. Hannon, (2004) **10**, 4291

[204] M. J. Hannon, C. L. Painting, N. W. Alcock, *Chem.* (1999) *Commun*., 2023 – 2024

[205] Lengths quoted are the distances between the centroids of the three pyridine C5 atoms at each end of the helix or the three pyridine H5 atoms. Radii are taken perpendicular to the helix axis.

[206] D.S. Sigman, T.W. Bruice, A. Mazumder, C.L. Sutton (1993) *Acc. Chem. Res.,* **26**, 98

[207] M. Pitie, B. Donnadieu, B. Meunier, (1998) *Inorg. Chem.,* **37**, 3486

[208] S.T. Frey, H.H.J. Sun, N.N. Murthy, K.D. Karlin, (1996) *Inorg. Chem. Acta.,* **242**, 329

[209] B. Schoentjes, J.-M. Lehn, (1995) *Helv. Chim. Acta.* **78**, 1

[210] L.J. Childs, J. Malina, B.E. Rolfsnes, M. Pascu, M.J. Perito, M.J. Broome, P.M. Rodger, E. Sletten, V. Moreno, A. Rodger, M.J. Hannon, (2006) *Chem. Eur. J.,* **12**, 4919

[211] M. Pitie, C. Boldron, H. Gornitzka, C. Hemmert, B. Donnadieu, B. Meunier, (2003) *Eur. J. Inorg. Chem.,* **3**, 528

[213] J.H. Moon, S.K. Kim, U. Sechlestedt, A. Rodger, (1996) *Biopolymers,* **38**, 593

[214] S. Khalid, (2004) *PhD Thesis*

[215] L. Verlet, (1967) *Phys.Rev.*, **159**, 98

[216] W.Smith, T.R. Forester (1994) *Comput. Phys. Commun.*, **79**, 63

[217] J. Wang, P. Cieplak, and P. A. Kollman, (2000) *J. Comput. Chem.*, **21**, 1049

[218] N. Foloppe and A. D. MacKerell, Jr., (2000) *J. Comp. Chem.*, **21**, 86

[219] E. Lindahl, B. Hess, and D. Van der Spoel, (2001) *Journal of Molecular Modeling 7*, **306**

[220] R. C. Rizzo and W. L. Jorgensen, (1999) *J. Am. Chem. Soc.*, **121**, 4827

[221] J. Wang, P. Cieplak, and P. A. Kollman, (2000) *J. Comput. Chem.*, **21**, 1049

[222] N. L. Allinger, M. T. T. andd M. A. Miller, and D. H. Wertz (1971) *J. Am. Chem. Soc.*, **93**, 1637

[223] N. L. Allinger (1977) *J. Am. Chem. Soc.*, **99**, 3279

[224] N. L. Allinger, Y. H. Yuh, and J.-H. Lii (1989) *J. Am. Chem. Soc.*, **111**, 8551

[225] J.-H. Lii and N. L. Allinger (1989) *J. Am. Chem. Soc.*, **111**, 8566

[226] J.-H. Lii and N. L. Allinger (1989) *J. Am. Chem. Soc.*, **111** 8576

[227] N. L. Allinger, K. Chen, and J.-H. Lii (1996) *J. Comput. Chem.*, **14**, 642

[228] N. L. Allinger, K. Chen, J. A. Katzeellenbogen, S. R. Wilson, and G. M. Anstead (1996) J. *Comput. Chem.*, **14**, 747

[229] N. L. Allinger, Y. H. Yuh, and J.-H. Lii (1989) *J. Am. Chem. Soc.*, **111**, 8551

[230] N. Nevins, K. Chen, and N. L. Allinger (1996) *J. Comput. Chem.*, **14**, 669

[231] S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Chio, G. Alagona, S. Profeta, and P. Weiner (1984) *J. Am. Chem. Soc.*, **106**

[232] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, T. Fox, J. W. Caldwell, and P. A. Kollman (1995) *J. Am. Chem. Soc.*, **117**, 5179

[233] B.R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus (1983) J. *Comput. Chem.*, **4**, 1234

[234] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, M.J. Karplus, (1983) *Comput Chem.*, **4**, 187–217

[235] A.D. MacKerell, B. Brooks, C.L. Brooks, L. Nilsson, B. Roux, Y. Won, M. Karplus, *In Encyclopedia of Computational Chemistry*; P. Schleyer, N.L. Allinger, T. Clark, J. Gasteiger, P. Kollman, H.F. Schaefer, P.R. Schreiner, Eds., (1998) *John Wiley & Sons: Chichester*, **1**, 271–277

[236] L. Nilsson, M. Karplus, (1986) *J Comp Chem.*, **7**, 591–616

[237] B.R. Gelin, M. Karplus, (1979) *Biochemistry.*, **18**, 1256–1268

[238] W. L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, (1983) *J Chem Phys.*, **79**, 926–935

[239] W. Saenger, *Principles of Nucleic Acid Structure*; Springer–Verlag: New York, (1984)

[240] A.D. MacKerell, (1997) *J Phys Chem B.*, **101**, 646–650

[241] L. Yang, B.M. Pettitt, (1996) *J Phys Chem.*, **100**, 2550–2566

[242] N. Foloppe, A.D. MacKerell, (2000) *J Comp Chem.* , **21**, 86–104

[243] D. Beglov, B.J. Roux, (1994) *J Chem Phys.*, **100**, 9050–9063

[244] N. Foloppe, A.D. MacKerell, (1998), *J Phys Chem B.*, **102**, 6669–6678

[245] N. Foloppe, A.D. MacKerell, (1999) *Biophys J.*, **76**, 3206–3218

[246] D.R. Langley, (1998) *J Biomol Struct Dynam.*, **16**, 487–509

[247] J. Khandogin, C. Brooks. (2005) *Biophysical Journal*, 89, 141

[248] J.E. Lennard-Jones, (1931) *Proceedings of the Physical Society*, **43**, 461-482

[249] N. L. Allinger, Y. H. Yuh, J.-H. Lii (1989) *J. Am. Chem. Soc.*, **111**, 8551

[250] A.J. Russell, A. R. Fersht (1987) *Nature*, **328**, 496-500

[251] M.W. Pantoliano, M. Whitlow, et al. (1988) *Biochemistry*, **27**, 8311-8317

[252] R.H. Stote, M. Karplus (1995). *Struct. Func. Gen.*, **23**, 12-31

[253] Allen, M. P. and D. J. Tildesley (1989). Computer Simulation of Liquids, Oxford University Press

[255] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W Impey, M.L. Klein, (1983) *J. Chem. Phys.*, **79**, 926

[256] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, (1983) *J. Chem. Phys.*, **79**, 926

[257] E. Stofer, R. Lavery (1994) *Biopolymers*, **34**, 337-346

[258] R.E. Dickerson, M. Bansal, C.R. Calladine, S. Diekmann, W. Hunter, O. Kennard, R. Lavery, H. Nelson, W. Olson, W. Saenger, Z Shakked, H. Sklenar, D.M. Soumpasis, C.S. Tung, E. Kitzing, A. Wang, V. Zhurkin, (1989) *J. Mol. Biol.* , **205**, 787–791

[259] G. Ravishanker D.L. Beveridge, (1991) *Molecular Dynamics Analysis Toolchest.* Department of Chemistry, Wesleyan University, Middletown, CT

[260] Lu et al., (1999) *J. Biomol. Struct. Dynam.,* **16**, 833

[261] Olson, W.K., X. J. L., (2000) *A standard reference frame for the description of nucleic acid base-pair geometry*

[262] G. Guzikevich-Guerstein, G., Shakked, Z., (1996) *Nature Struct. Biol.*, **3**, 32-37

[263] W. Kabsch, (1978) *Acta Cryst.* **A34**, 827

[264] R. Lavery, H. Sklenar, (1988) *J.Biomol.Str.Dyn*., **6**, 63

[265] G. Ravishanker, S. Swaminathan, D.L. Beveridge, R. Lavery, H. Sklenar, (1989), *J.Biomol.Str.Dyn.*, **6**, 669

[266] R.E. Dickerson, *et al.*, (1989) *J. M. Biol.,* **205**, 787

[267] G. Ravishanker, *Molecular Dynamics Analysis Toolchest*

[268] Jr. A.D. MacKerell, D. Bashford, M. Bellot, R.L. Dunbrack, J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, Ha S. Guo., D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnik, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, M.J. Karplus. (1998) *J. Phys. Chem. B.*, **102**, 3586

[269] P. Auffinger, E. Westhof, (1997) *Curr. Opin. Struct. Biol.,* **8**, 227

[270] M. Feig, B. Pettit, (1998) *Biophys.*, **75**, 134

[271] T.E. Cheatham, P.A. Kollman (1996) *J. Mol. Biol.*, **259**, 434

[272] L. Yang, B. M. Pettitt, (1996) *J. Phys. Chem.*, **100**, 2564

[273] N. Foloppe, A.D. MacKerell, (2000) *J. Comp. Chem.* **21**, 86

[274] R. Swarnalatha, F. Leclerc, M. Karplus, (2003) *Biophysical J.*, **84**, 1421

[275] A.D. MacKerell, N. Banavali, (2000) *J. Comp. Chem.*, **21**, 105

[276] T.E. Cheatham, P. Cicplak, P.A. Kollman, (1999) *J. Biomol. Struct. Dyn.* **16**, 845

[277] A. Perez, I. Marchan, D. Svozil, J. Sponer, T.E. Cheatham, C. Laughton, M. Orozco, (2007) *Biophys. J.*, **92**, 3817-29

[278] D.R. Langley (1998) *J. Biomol. Struct. Dyn.*, **16**, 487

[279] W. Smith, T.R. Forester, (1996) *J. Mol. Graphics*, **14**, 136

[280] M.P. Allen & D.J. Tildesley, (1987) *Computer Simulation of Liquids*, Oxford University Press, Oxford, UK

[281] H.R. Drew, R.M. Wind, T. Takano, C. Broka, S. Tanaka, K. Itakura, R.E. Disckerson, (1981) *Proc. Nat. Acad. Sci.*, USA, **78**, 773-782

[282] B. Wellenzohn, W. Flader, R.H. Winger, A. Hallbrucker, E. Mayer, K.R. Liedl, (2001) *Biopolymers*,**61**, 276

[283] A. Elofsson, L. Nilsson, (1993) *J. Mol. Biol.*, **233**, 766

[284] P. Auffinger, S. Louise-May, E. Westhof, (1995) *J. Am. Chem. Soc.*, **117**, 6720

[285] V.A. Likic, F.G. Prendergast (2001) *Proteins*, **43**, 65

[286] J. Straub, A.B. Rashkin, D. Thirumalai, (1994) *J. Am. Chem. Soc.*, **116**, 2049

[287] L.S.D. Caves, J.D. Evanseck, M. Karplus, (1998) *Protein Sci*., **7**, 649

[288] B. Hess, (2002) *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.*, **65**, 31910

[289] J.B. Clarage, T. Romo, B.K. Andrews, B.M. Pettitt, G.N. Phillips, (1995) *Proc. Natl. Acad. Sci.*, **92**, 3288

[290] Y. Komeiji, Y. Ueno, M. Uebayasi, (2002) *FEBS Lett.*, **521**, 133

[291] V.A. Likic, E.E. Strehler, P.R. Gooley, (2003) *Protein Sci.* **12**, 2215

[292] A.L Vladimir, P.R. Gooley, T. P. Speed, E.E. Strehler, (2005) *Protein Sci.*, **12**, 2955

[293] A.H. Elcock, A. McCammon, *J. Am. Chem. Soc.*, (1996) **118**, 3787

[294] C.F. Guerra, F.M. Bickelhaupt, (2002) *Angew. Chem.*, **41**, 2092

[295] C.F. Guerra, F.M. Bickelhaupt, (1999) *Angew. Chem. Int. Ed.*, **38**, 2942

[296] S. Khalid, M.J. Hannon, A.Rodger, P.M.Rodger (2006) *Chem. Eur. J.*, **12**, 3493

[297] B. Bagchi, (2003) *Annu. Rep. Prog. Chem. Sect. C*., **99**, 127-175

[298] P.M. Rodger, (1989) *J. Phys.Chem*., **93**, 6850

[299] T. J. Carver, M. G. B. Drew and P. M. Rodger, (1995) *J. Chem. Soc., Faraday Trans.*, **91**, 3449

[300] H. J. C. Berendsen, J. P. M. Postma, W. F. von Gunsteren, J. Hermens, (1981) *Intermolecular Forces*, 331 – 342

[301] T. J. Carver, M.G. B. Drew, P. M. Rodger, (1999) Phys. *Chem. Chem. Phys.*, **1**, 1807

[302] W. Smith, T.R. Forester, *DL_POLY is a package of molecular simulations routine*, (1995)

[303] T.J. Carver, G.B. Drew, P.M. Rodger, (1999) *Phys. Chem. Phys.*, **1**, 1807

[304] T. Astley, G. Birch, G. Drew, P.M. Rodger, (1998) *J. Comput. Chem*., **19**, 363

[305] J. Fidler, P.M. Rodger, (1999) *J. Phys. Chem. B*, **103,** 7695

[306] P.M. Rodger, T.R. Forrester, W. Smith (1996) *W. Fluid Phase Equilib.*, **116**, 326

[307] P. M. Rodger, (1999) *Ann. N. Y. Acad. Sci.*, **870**, 330