A STRUCTURAL STUDY OF M-DNA

A Thesis Submitted to the College of
Graduate Studies and Research
in Partial Fulfillment of the Requirements
for the Degree of Master of Science
in the Department of Biochemistry
University of Saskatchewan
Saskatoon

By

Angela G. Hoffort

PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professors who supervised my thesis work, or in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any materials in my thesis.

Request for permission to copy or make other use of material in this thesis in whole of part should be addressed to:

Head of the Department of Biochemistry
University of Saskatchewan
Saskatoon, Saskatchewan, S7N 0W0

ABSTRACT

In alkaline conditions, a complex called M-DNA is formed between a divalent metal ion, cobalt, nickel or zinc, and duplex DNA. The rate of formation and stability of M-DNA is dependent on many factors, including but not limited to temperature, pH, DNA sequence, and metal or DNA concentrations. It has been hypothesized that the divalent metal ions intercalate into the helix and replace the imino protons involved in the hydrogen bonding of both G-C and A-T base pairs. The complex is thought to have a double helical structure that is similar to B-DNA. The presence of the divalent metal ions and a more compact structure may contribute to M-DNA's remarkable ability to behave as a molecular wire. Because M-DNA is so similar to B-DNA, it adheres to the same rules with regards to self-assembly. The ability of DNA to self-assemble and the electronic conduction of M-DNA are ideal properties for nanotechnology of the future. M-DNA may eventually be used to detect the presence of biologically important small molecules and DNA binding proteins that block the flow of electrons. However, before M-DNA will be widely accepted, it is necessary to obtain an accurate 3-dimensional structure by X-ray crystallography and modelling.

In this work interactions between divalent cobalt, nickel or zinc with duplex DNA were studied using two different experimental methods; namely, X-ray crystallography and extended X-ray absorption fine structure spectroscopy. First, crystals of the sequence d[GA(5FU)(5FU)AA(5FU)C] and d[CG(5FU)G(5FU)GCACACG] complexed with divalent metals were grown in M-DNA favouring conditions. Both of the sequences gave crystals that provided diffraction data that were analyzed by molecular replacement using B-DNA models. Unfortunately, the quality of the diffraction was not high enough with either sequence to locate metal binding or to determine a model for M-DNA. Second, X-ray absorption spectroscopy data were analyzed for the Ni^{2+} *K*-edge of both Ni^{2+} M and B-DNA. Several differences between the M and the B-DNA data were noticed and some final bond distances were established.

ACKNOWLEDGMENTS

I would like to begin by first acknowledging my supervisors, Drs Jeremy Lee and Louis Delbaere for providing me with such an interesting project as well as their continued guidance and support. Special thanks go out to Shaun Labiuk for teaching me most of what I needed to know to complete the crystallography portion of this project. Any gaps left by Shaun were filled in by the following Dr. Stan Moore, Yvonne Leduc, Fumie Imabayashi, Julien Cotelesage, Dr. Lata Prasad, Dr. Sanjukta Aich, Kurt Nienaber, Jenny Puttick, Michael Dinsmore and Dr. David Wood. I would also like to thank Dr. Ramaswami Sammynaiken for his guidance with the X-ray absorption spectroscopy portion of the project, as well as Ryan Skinner for data collection. For their constant encouragement and feedback, I would like to thank my friends and family; namely, my parents, sisters, grandparents and my husband, Marc Hoffort.

For financial support, I would like to thank the Saskatchewan Synchrotron Sciences Institute and the College of Medicine. Research that took place at the Advanced Photon Source in Chicago was supported by the United States Department of Energy, Basic Energy Sciences, Office of Science, under Contract No. W-31-109-Eng-38. Use of BioCars Sector 14 was supported by the National Institute of Health, National Center for Research Resources.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

A	adenine
APS	Advanced Photon Source
5BrU	5-bromo-deoxyuridine
C	cytosine
CNS	Crystallography and NMR System
d	deoxyribo prefix
DNA	deoxyribonucleic acid
EXAFS	extended X-ray absorption fine-structure spectroscopy
5FU	5-fluorouracil
G	guanine
I	inosine
LLG	log-likelihood gain
MAD	multiple wavelength anomalous diffraction
MPD	2-methyl-2,4-pentanediol
NDB	Nucleic Acid Database
NiPC	nickel (II) phthalocyanine
pH	negative logarithm of the hydrogen ion concentration
pK_a	negative logarithm of the dissociation constant of an acid
rmsd	root mean square deviation
RNA	ribonucleic acid
s^4T	4-thiothymidine
T	thymine
TAPS	N-tris-[hydroxymethyl]methyl-3-aminopropanesulfonic acid
U	uracil
XAFS	X-ray absorption fine structure
XANES	X-ray absorption near edge spectroscopy

1.0 INTRODUCTION

Nucleotides are the most versatile of the major molecular living cell constituents. They are notable for their involvement in reactions essential to the propagation and maintenance of life. More specifically, nucleotides are necessary for energy transfer, storage and the decoding of genetic information as well as having a structural and catalytic role in the cell. Deoxyribonucleic acid (DNA) is composed of a polymer of nucleotides. In order to fully comprehend the function of DNA and exploit it to its fullest, it is important to have a complete understanding of its primary and secondary structure. Having an appreciation for DNA structure will not only help us understand its biological role, but will aid in the attempt to use DNA as a constituent in nanotechnology and genetic engineering. The DNA molecule has many appealing features for use in nanotechnology, including its miniscule size and consistent diameter, its short structural repeat and its stiffness (Seeman, 2003). The better we understand DNA's three dimensional structure, the better we will understand its biological role and its potential as a tool for genetic engineering and nanotechnology.

DNA was first discovered by Friedrich Miescher in 1869 (Portugal and Cohen, 1977). Miescher termed the find "nuclein" and distinguished it from protein by its phosphorous content. He incorrectly suggested the function of nuclein as the storage of phosphorous in the cell. DNA wasn't considered genetic material until 1944 when Oswald Avery discovered its role in inheritance within bacteria (Avery, 1944; McCarty, 1946). In the early 1900's, Albert Kossel played a key role in discovering the five bases found in DNA and ribonucleic acid (RNA): guanine (G), adenine (A), thymine (T), cytosine (C) and uracil (U). The correct primary structure for nucleotides binding together to form a long chain was proposed first by Phoebus Levene (Levene and Tipson, 1935). Levene is also known for discovering the difference between the sugars of RNA and DNA, that being DNA has deoxyribose sugar in its backbone, whereas RNA has a ribose sugar (Portugal and Cohen, 1977). The primary structure of DNA did not lend itself well to explaining how DNA functioned as genetic material. It was going to take a

better understanding of its three dimensional structure to answer these questions. In 1953, the first correct three dimensional model for DNA was proposed by James Watson and Francis Crick (Watson and Crick, 1953). Their conclusions come from a collaboration surrounded by controversy and credit must also be noted for Rosalind Franklin and Maurice Wilkins. It was Franklin and Wilkins, among others who are responsible for the first fibre diffraction X-ray studies on DNA that allowed Watson and Crick to make some of their conclusions (Franklin and Gosling, 1953; Wilkins *et al.*, 1953). It is often said that the determination of the DNA structure by Watson and Crick marks the beginning of modern molecular biology.

Since the discoveries of 1953, the biological function of DNA has completely opened up, and a new era of research has begun. The entire human genome has now been sequenced as well as the genomes of several other species. Using X-ray diffraction it is now possible to study in more detail the structure of short oligonucleotides as well as DNA complexes with proteins, drugs or metals among other molecules. Besides the enormous amount of information we have gained about the biological function of DNA, this understanding has also opened the doors for genetic engineering and nanotechnology.

The work presented in this thesis is an attempt to improve the understanding of DNA-metal interactions. It will begin with a thorough introduction to DNA structure followed by what is known about DNA-metal interactions. The work presented focuses on X-ray crystallography and X-ray Absorption Spectroscopy. Therefore, there will be a brief introduction of these two methods.

1.1 Structure of Deoxyribonucleic Acid

1.1.1 Primary Structure

DNA is a polymer of nucleotides. Each nucleotide consists of a nitrogenous base linked to a sugar to which at least one phosphate group is attached. The nitrogenous bases that make up nucleotides are aromatic, planar, heterocyclic molecules structurally derived from either a purine or a pyrimidine. The most common bases derived from purine are G and A. The pyrimidine derivatives are T, C and U (Figure 1.1). For clarity, the bases are labelled with the number of the atom and the pentose are labelled

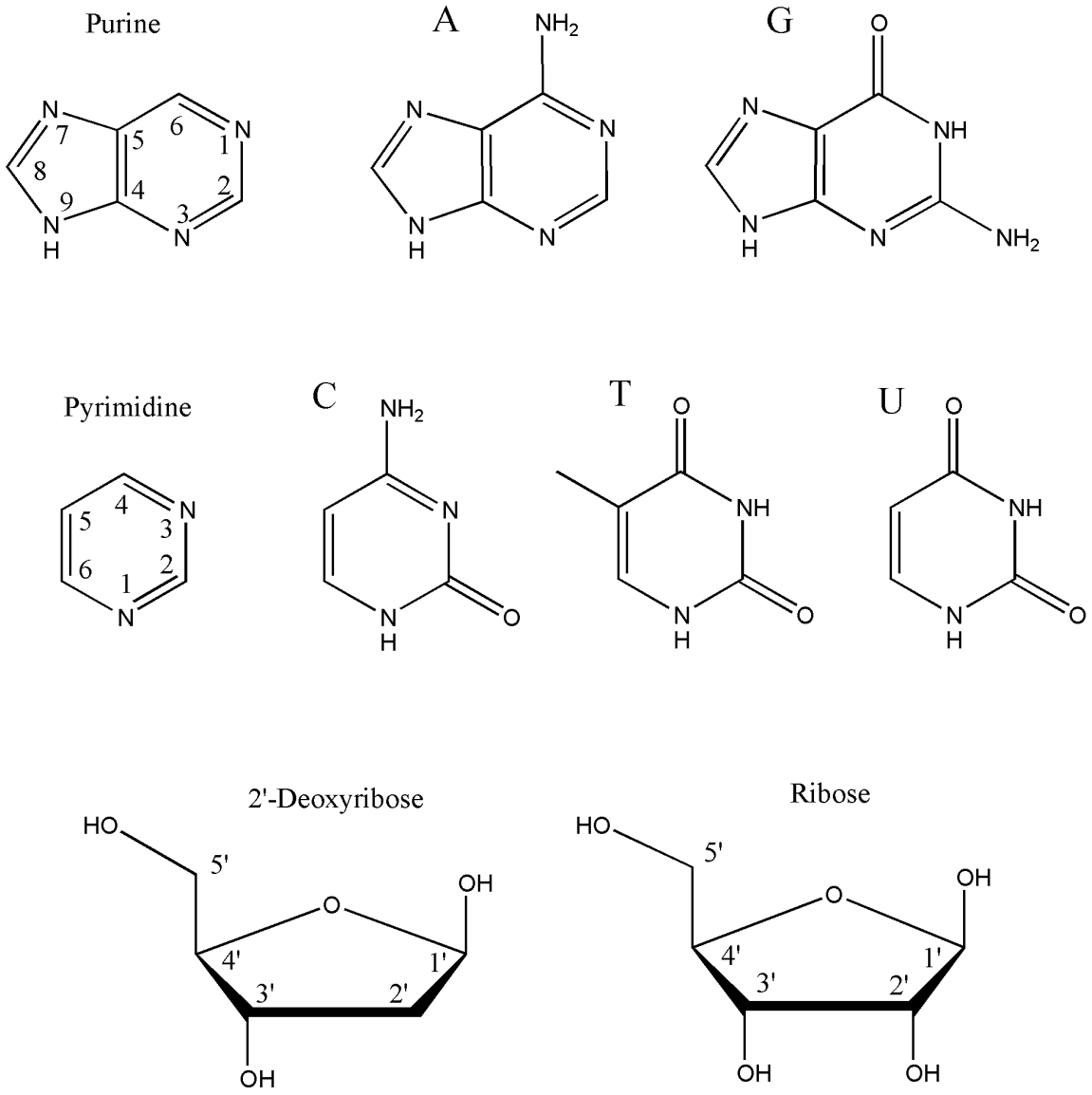

Figure 1.1: Components that make up nucleotides. The numbering schemes for the purine bases (A and G), the pyrimidine bases, (C, T and U, β-D-2'-deoxyribose and β-D-ribose) are also shown.

with prime numbers as shown in Figure 1.1. In nucleotides, the purine bases form bonds with the C1' of a furanoside-type sugar, *via* their N9 atoms, whereas pyrimidines bind through their N1 atom. DNA is made up of deoxyribonucleotides, or deoxynucleotides and ribonucleic acid (RNA) consists of ribonucleotides. In deoxyribonucleotides, the pentose is β-D-2'-deoxyribose and in ribonucleotides, the pentose is β-D-ribose (Figure 1.1). If the base is attached to a β-D-ribose, the molecule is termed a nucleoside. Therefore the five nucleosides are then called guanosine, adenosine, cytidine, uridine and thymidine. If they are attached to β-D-2'-deoxyribose, as in DNA, they are called deoxyguanosine (dG), deoxyadenosine (dA), deoxycytidine (dC), deoxyuridine (dU) and deoxythymidine (dT). If the base is attached to a β-D-ribose which is in turn attached to at least one phosphate, that is a complete nucleotide and they are labelled, guanylic acid, adenylic acid, cytidylic acid, uridylic acid and thymidylic acid. If the pentose is β-D-2'-deoxyribose, then the prefix deoxy is put before the names as it is for the nucleosides. A phosphate group can be bound to either the C3' or the C5' of the pentose to form the 3'-nucleotide or the 5'-nucleotide respectively.

The bulk of nucleotides found in any cell exist in a polymeric form, each connected through a phosphate bond between the C3' on one pentose and the C5' on the adjoining pentose (Figure 1.2). The linkage between the individual nucleotides is known as a phosphodiester bond. The tetradeoxynucleotide shown in Figure 1.2 is adenylyl-3',5'-thymidylyl-3',5'-guanylyl-3',5'-cytidine and can be abbreviated d(ApTpGpC) or d(ATGC). If the strand of DNA is a repeat of a smaller sequence, it is abbreviated with the prefix "poly" in front of the repeated sequence, for example poly[d(AG)]. The complementary strand to the poly[d(AG)] would be poly[d(CT)]. The duplex structure is written as poly[d(AG)]•poly[d(CT)]. It is by convention that a polynucleotide sequence is written from the 5' end to the 3' end. The nucleotide residues at the 3' and 5' ends of the sequence are termed the 3'end and the 5'end respectively. The next step in understanding DNA structure is to take a look at its secondary structure.

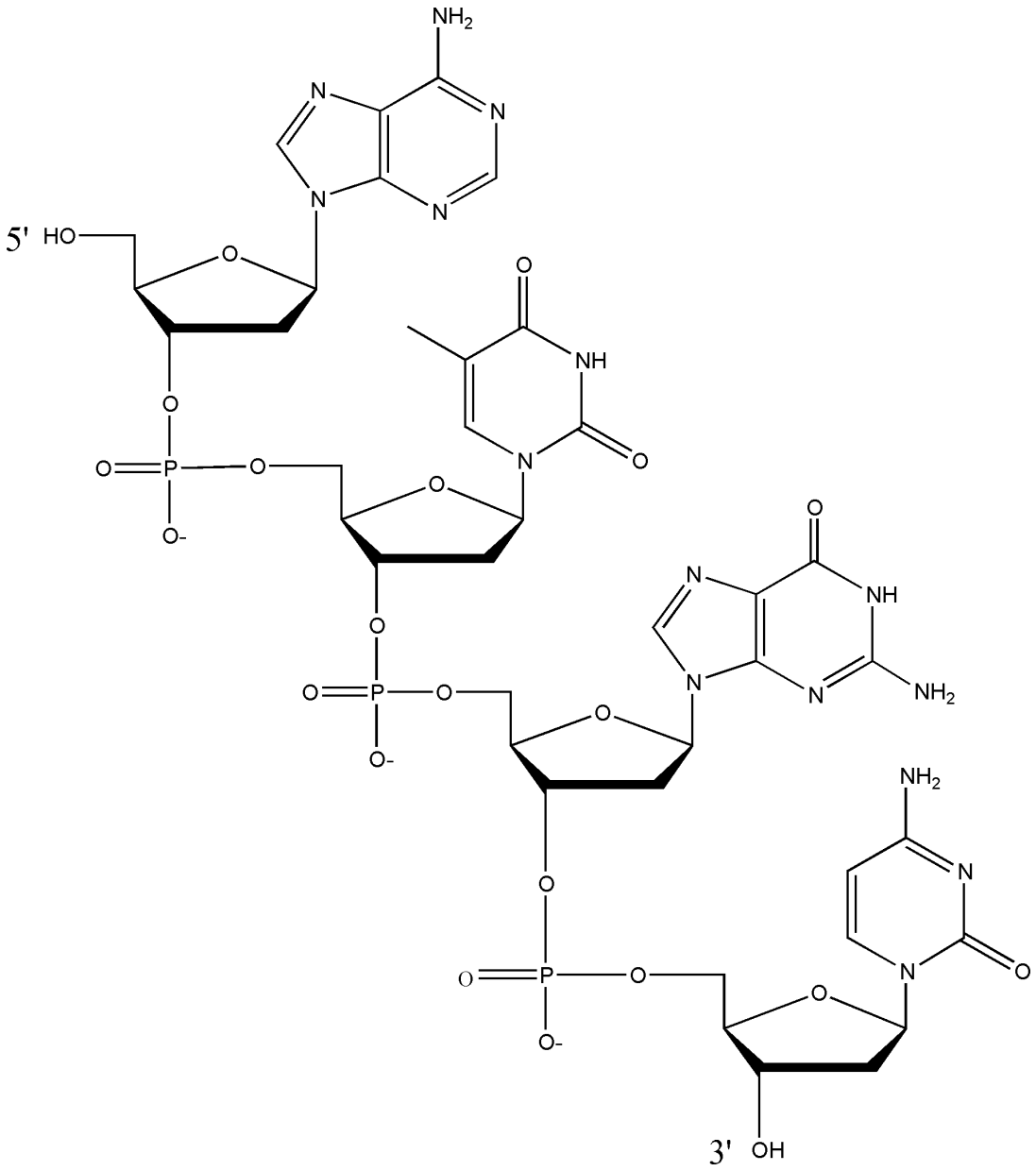

Figure 1.2: Primary structure for the tetradeoxynucleotide adenylyl-3',5'-thymidylyl-3',5'-guanylyl-3',5'-cytidine (d(ATGC)).

1.1.2 Secondary Structure

1.1.2.1 Sugar-Phosphate Backbone and the Glycosyl Bond

The conformational flexibility of the nucleotide unit is limited by a variety of internal constraints that greatly restrict rotational freedom. The conformation of a molecule is usually described by bond lengths, bond angles and rotations of groups of atoms about bonds by torsion angles (Saenger, 1984). Specifically, there are seven torsion angles that are important for describing the secondary structure of a nucleotide, α, β, γ, δ, ε, ζ and χ (Figure 1.3). The first six describe the torsion angles of the sugar-phosphate backbone, the last torsion angle, χ, describes the rotation around the glycosyl bond. The rotation of a base around its glycosyl bond is greatly hindered by steric interactions, especially for the pyrimidine bases due the 2-keto group. Generally, the glycosyl bond is only found in one of two sterically permissible positions: *syn* and *anti*. For example, the two structures shown in Figure 1.3 are called *syn*-adenosine and *anti*-adenosine. In the *anti* conformation, most of the base is pointing away from the sugar, whereas in the *syn* conformation, most of the base is over toward the sugar giving rise to close interatomic contacts (Donohue and Trueblood, 1960; Haschemeyer and Rich, 1967). The amount of steric hindrance that exits in the *syn* conformation is affected by the conformation of the sugar or the sugar pucker.

1.1.2.2 Sugar Pucker

The sugar pucker describes the flexibility of the ribose ring which is generally nonplanar. The sugar may be puckered in an envelope form with four of the five atoms coplanar and the fifth atom being out of this plane by 0.5 Å; or in a twist form with two adjacent atoms displaced on opposite sides of the plane formed by the other three atoms (Hall, 1963; Saenger, 1984) (Figure 1.4). In the 2'-*endo* and 3'-*endo* conformations, it is the C2' or the C3', respectively, that is on the same side of the plane as the C5'. In the 2'-*exo*-3'*endo* conformations, the C2' is on the opposite side of the plane as C5' and the C3' is on the same side of the plane. In conformations with two atoms out of the plane, it is rare that the two will be displaced equally out of the plane. Therefore, the atom farthest from the plane has major puckering, whereas the other atom has minor

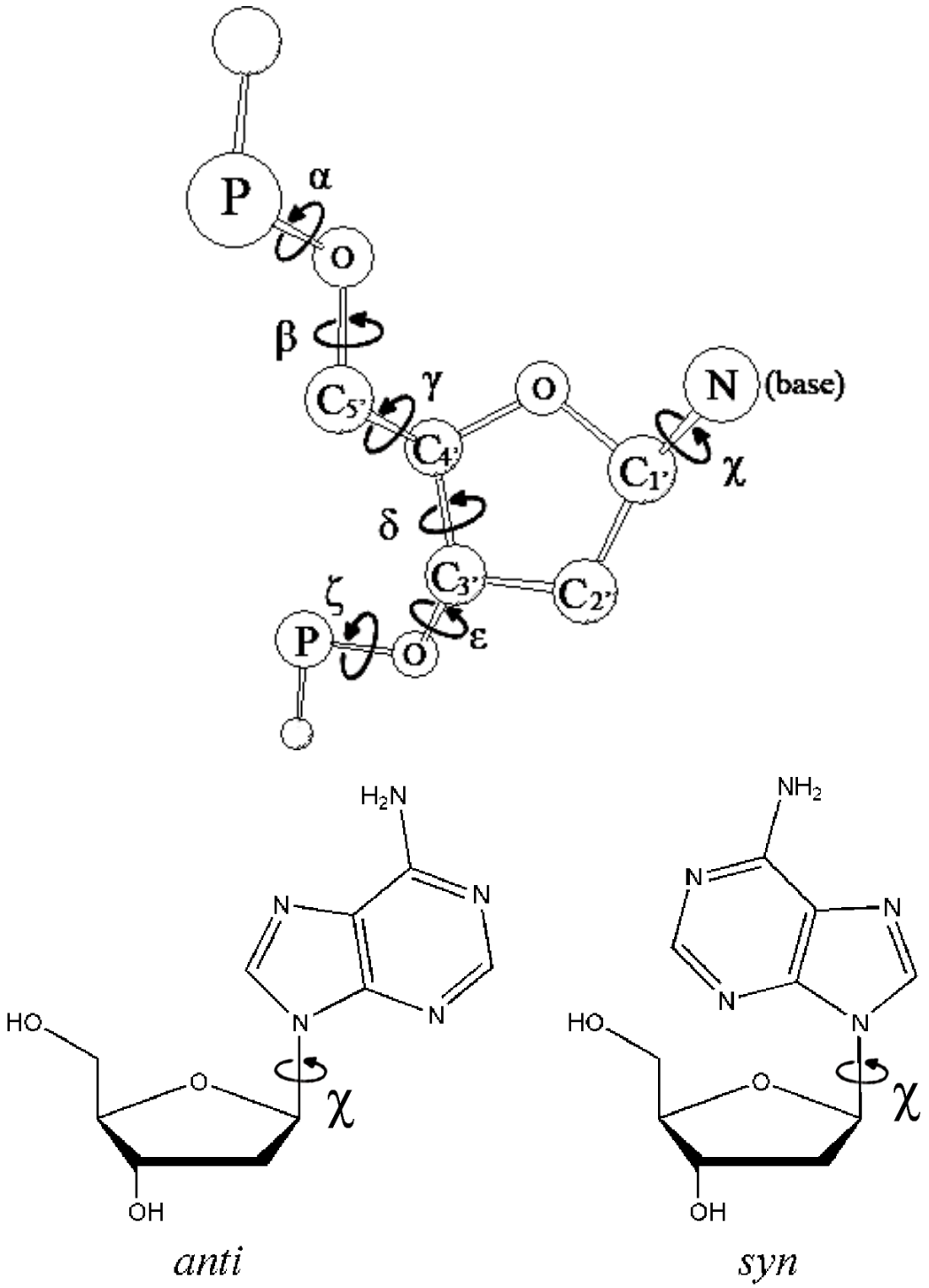

Figure 1.3: Diagram of torsion angles in a nucleotide along the sugar-phosphate backbone and the glycosyl bond (top). The glycosyl bonds can be either in the *syn* or *anti* conformations (bottom), *syn* being the conformation with the most steric hindrance.

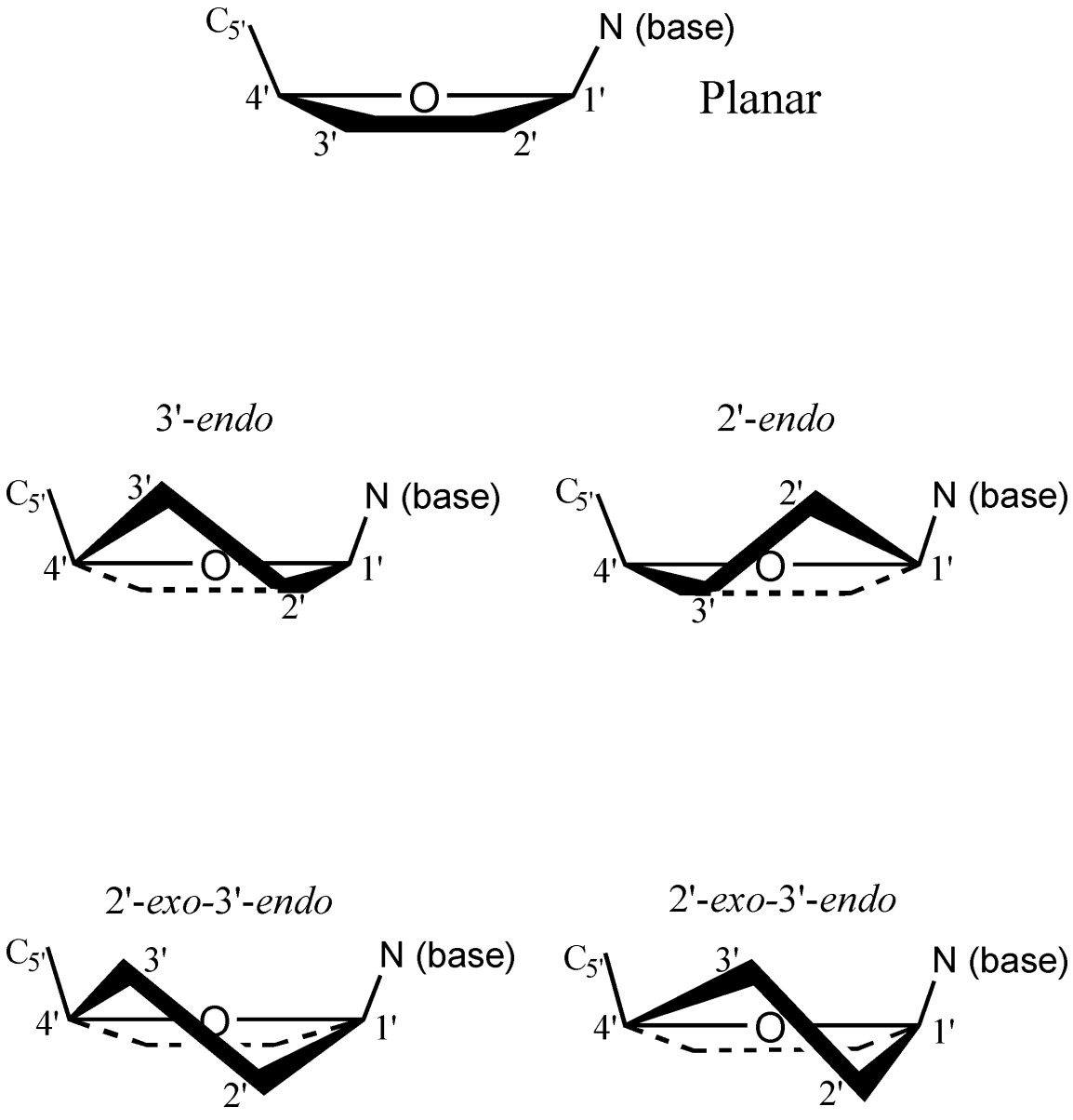

Figure 1.4: Different sterically permissible sugar pucker conformations including 3'-*endo*, 2'-*endo* and 2'-*exo*-3'-*endo*. The planar conformation is not sterically permissible, but rather an illustration to show the plane (top). Of the two 2'-*exo*-3'-*endo* conformations, the one on the left shows major 2'-*exo* with minor 3'-*endo* and the conformation on the right illustrates the opposite, minor 2'-*exo* with major 3'-*endo*. In DNA, the 2'-*endo* conformation is the most common.

puckering. Generally, in DNA, the 3'-*endo* and 2'-*endo* conformations are favoured with a larger portion of nucleotides being in the 2'-*endo* conformation. Now that we know about the sugar-phosphate backbone and the sugar pucker, it is important to understand how the polynucleotide chains bind in relation to one another. Here is where the helix and the studies of Watson and Crick come into play.

1.1.2.3 A, B and Z-DNA

In general, DNA is a double-stranded polymer of deoxynucleotides linked by phosphodiester bonds. However, there are exceptions to every rule and I will describe those later. The most common biological form of DNA, and the one described by Watson and Crick (1953a), is known as B-DNA. B-DNA exist as two anti-parallel polynucleotide strands winding in a right-handed fashion around a common axis. In B-DNA, the bases occupy the center of the helix, while the sugar-phosphate backbone occupies the outside, creating a narrow and deep major groove, with a wide and shallow minor groove (Figure 1.5) (Franklin and Gosling, 1953; Wilkins *et al.*, 1953). Watson and Crick also explained the pairing of the bases in DNA based on observations by Chargaff (Chargaff *et al.*, 1951). They proposed the bases pair with each other such that A always pairs with T and G always pairs with C (Figure 1.6) *via* two and three hydrogen bonds respectively. Since the discoveries of 1953, X-ray studies on single crystals of deoxynucleotides have revealed much more detailed information about the structure of B-DNA (Dickerson *et al.*, 1982; Dickerson, 1992). The ideal B-DNA duplex is about 20 Å in diameter, has ten base pairs per turn with a rise per base pair of approximately 3.4 Å. The plane of the base pairs tilts 6 ° with respect to the helical axis and has a helical twist of 36 °. In B-DNA, the favoured sugar pucker is 2'-*endo* and the glycosyl bond is in the *anti* conformation.

The structure of duplex DNA can assume several distinct forms depending on many factors including, but not limited to, solvent composition and base sequence. Other than B-DNA, the best known structures for duplex DNA are A-DNA and Z-DNA (Figure 1.5). A-DNA exits under dehydrating conditions and is a wider and flatter right-handed helix than B-DNA. It is approximately 26 Å in diameter and has 11 base pairs per turn. The helical twist per base pair is 33 ° with a rise per base pair of 2.6 Å. Most notably, A-

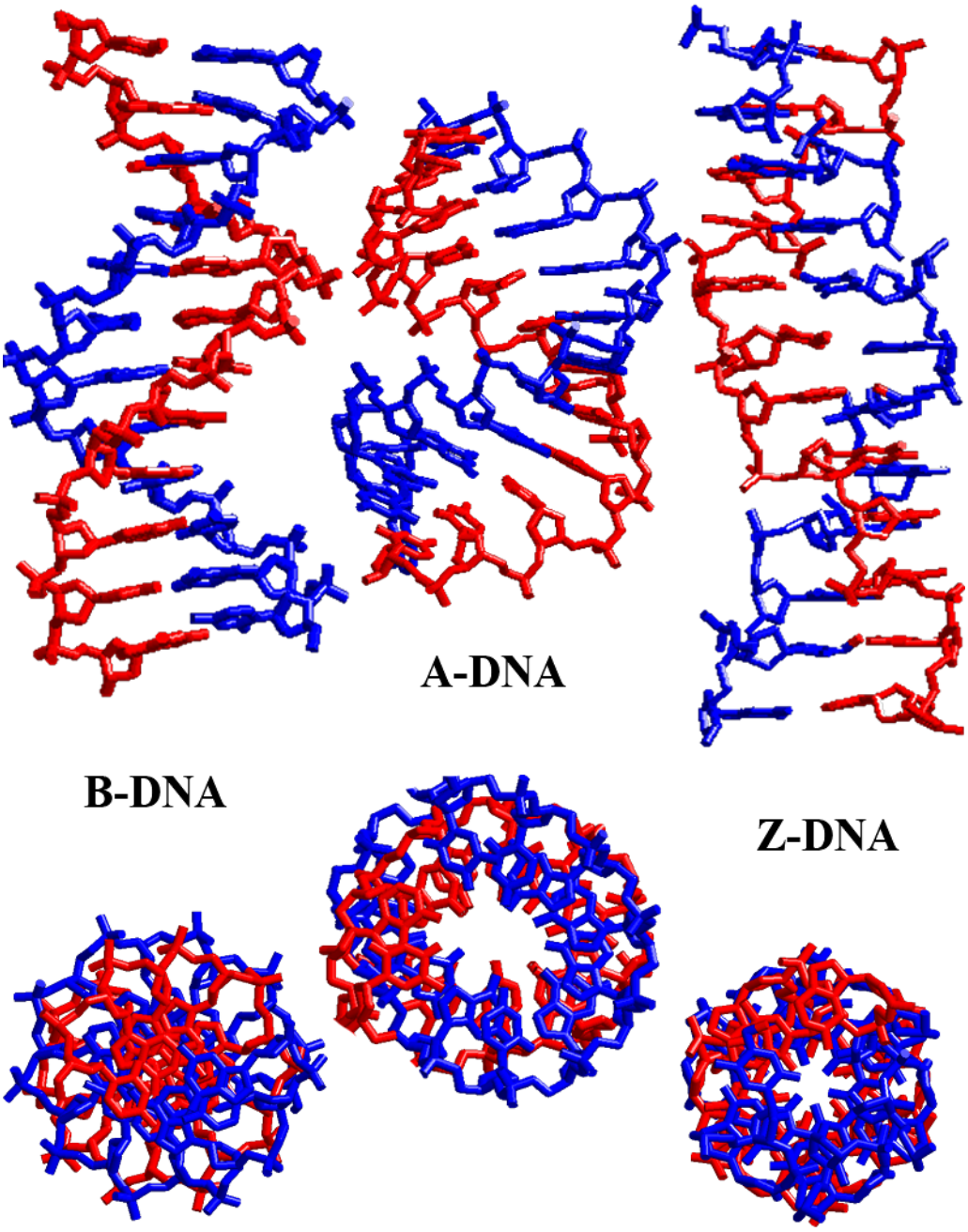

Figure 1.5: Structures of B, A and Z-DNA double helices. Both A and B-DNA are right-handed helices, whereas, Z-DNA is a left-handed helix. Adapted from Drew *et al.*, 1981; Brennan *et al.*, 1986; Dohm *et al.*, 2005.

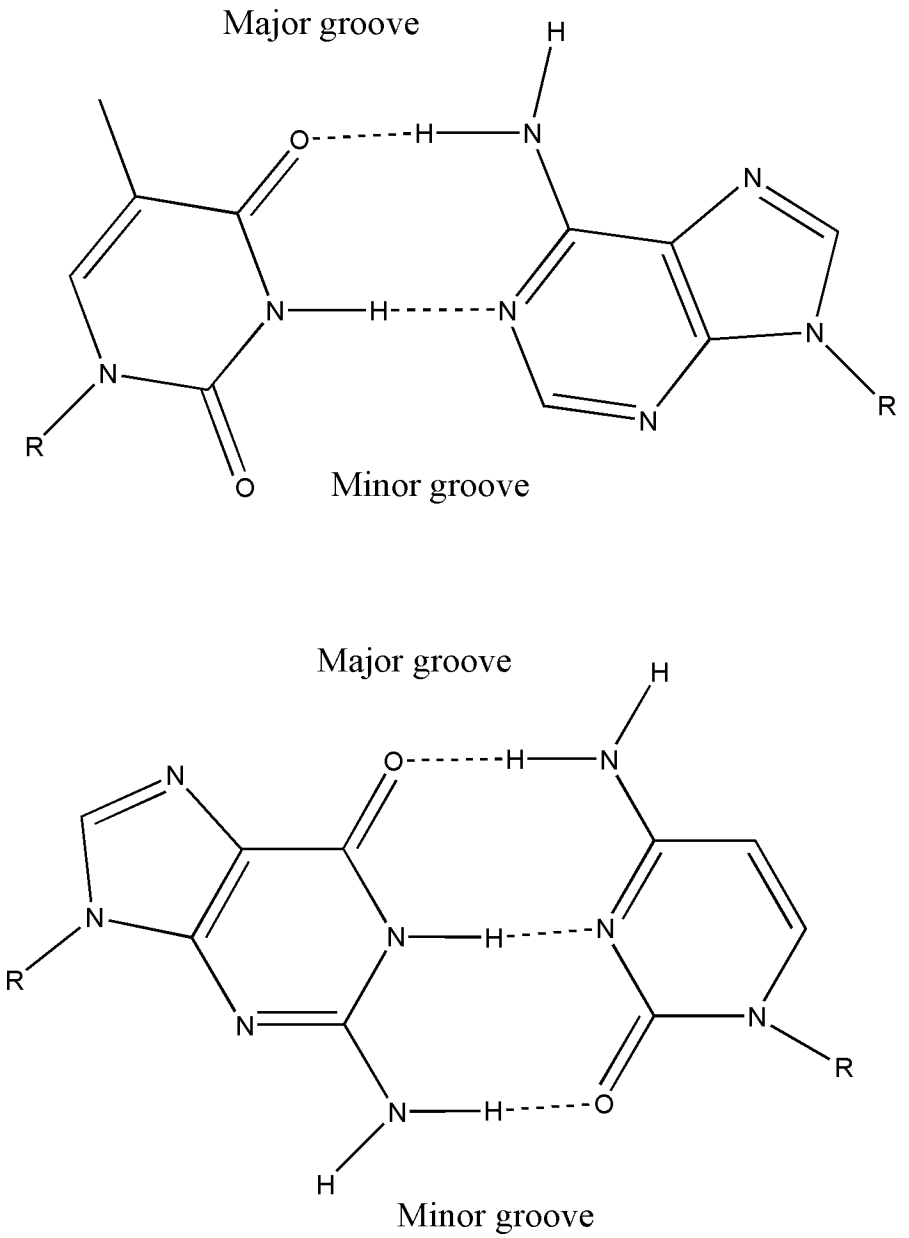

Figure 1.6: Ideal Watson and Crick base pair geometry in duplex DNA. T pairs with A (top) *via* two hydrogen bonds and G pairs with C (bottom) *via* three. In both pairs, the R represents 2'-deoxyribose.

DNA's base pairs are tilted 20 ° normal to the helix axis. The favoured sugar pucker in A-DNA is 3'-*endo* and the glycosyl bond is in the *anti* conformation. In contrast, Z-DNA is a left-handed helix that is about 18 Å in diameter and has 12 base pairs per turn. Its rise per base pair is 3.7 Å and the base tilt is 7 ° normal to the helix axis. Z-DNA has a deep minor groove and no apparent major groove (Figure 1.5). Z-DNA exits at high salt concentrations with a sequence of alternating purines and pyrimidines. The favoured sugar pucker is 2'-*endo* for pyrimidines and 3'-*endo* for purines. The glycosyl bond is in the *anti* conformation for pyrimidines and *syn* for purines. Although the A, B and Z-DNA conformations are the most commonly known DNA conformations, there are more than twenty other DNA conformations that have been discovered to date (Ghosh and Bansal, 2003; Egli, 2004).

1.1.2.4 Other DNA Structures

Changes in base sequence and environment can have an astounding effect on DNA structure. By varying the features mentioned so far, such as sugar pucker, sugar-phosphate backbone geometry and the glycosyl bond, completely different DNA structures are born. The different DNA structures may contain small deviations from the local parameters of Watson and Crick's model, or they be may be completely different, even in their essential features such as handedness, base-pairing or their number of strands (Ghosh and Bansal, 2003).

B'-DNA is a variant of B-DNA, thought to be formed by poly[d(A)]•poly[d(T)] sequences. It is characterized by a large propeller twist and a narrow minor groove (Chandrasekaran and Radha, 1992). C-form DNA has 9.3 residues per turn and is observed for calf thymus DNA. D-DNA has eight residues per turn and is observed for poly[d(AT)]•poly[d(AT)] (Ghosh and Bansal, 2003). Both C- and D-DNA are double stranded duplexes in the right-handed direction. H-DNA is a triple helical structure formed under low pH with a sequence of poly[d(purine)]•poly[d(pyrimidine)] and a Holliday junction is formed when one strand from each of two duplexes exchange to form a four-way junction. Metal cations can stabilize different DNA structures, such as triplex DNA or quadruplex DNA, which are helical structures of DNA composed of three or four single strands wound around the helical axis, respectively. Notably there is

a large variety of different DNA structures. An entire thesis could be dedicated completely to reviewing all the published DNA structures. A little later I will return to some particular structures that are important when understanding the role divalent metals play in stabilizing DNA structure.

1.1.3 Base pair Geometry

Duplex DNA structure indicates little about the local base pair geometries, which may vary significantly within an individual oligonucleotide. Different duplex DNA structures tend to have different average base-pair geometries. However, there are often significant local variations that may not be adequately represented in the average. Therefore, it is important to have a common reference to describe the three-dimensional arrangement of bases and base pair geometry in nucleic acids (Dickerson *et al.*, 1989; Olson *et al.*, 2001). Relative to the ideal Watson-Crick base pair, the coordinate frame used to describe base pair geometry follows established qualitative guidelines (Figure 1.7). If a line is drawn between the C1' of the purine with the C1' of its complimentary pyrimidine in an ideal Watson-Crick base pair, the bisector of this line is known as the pseudo-dyad axis. The *x*-axis points in the direction of the major groove with its origin being at the pseudo-dyad. The *y*-axis runs along the long axis of the ideal base pair, is parallel to the C1'-C1' vector and intersects the *x*-axis on the pseudo-dyad. The *z*-axis also originates at the pseudo-dyad and points in the 5'- to 3'-direction of the sequence strand.

It is important to note that the axes chosen must be independent of the global helical axis to make it possible to analyze bent DNA structures. The local helical parameters are used to describe the geometry of the base pair's *z*-axis relative to the global helix axis (Figure 1.7 (lower left)). Translation of the base pair along the *x*- and *y*-axes are described as x-displacement and y-displacement, respectively. Rotations of the base pair's *z*-axis about the *y*- and *x*-axes are known as inclination and tip, respectively.

As illustrated in Figure 1.7, the base pair parameters and dimer step parameters describe translation and rotation around the different axes. *Shear*, *stretch* and *stagger* describe the translation of the two bases along the *x*-axis, the *y*-axis and the *z*-axis, respectively. *Buckle*, *propeller* and *opening* refer to the rotation of the bases around the

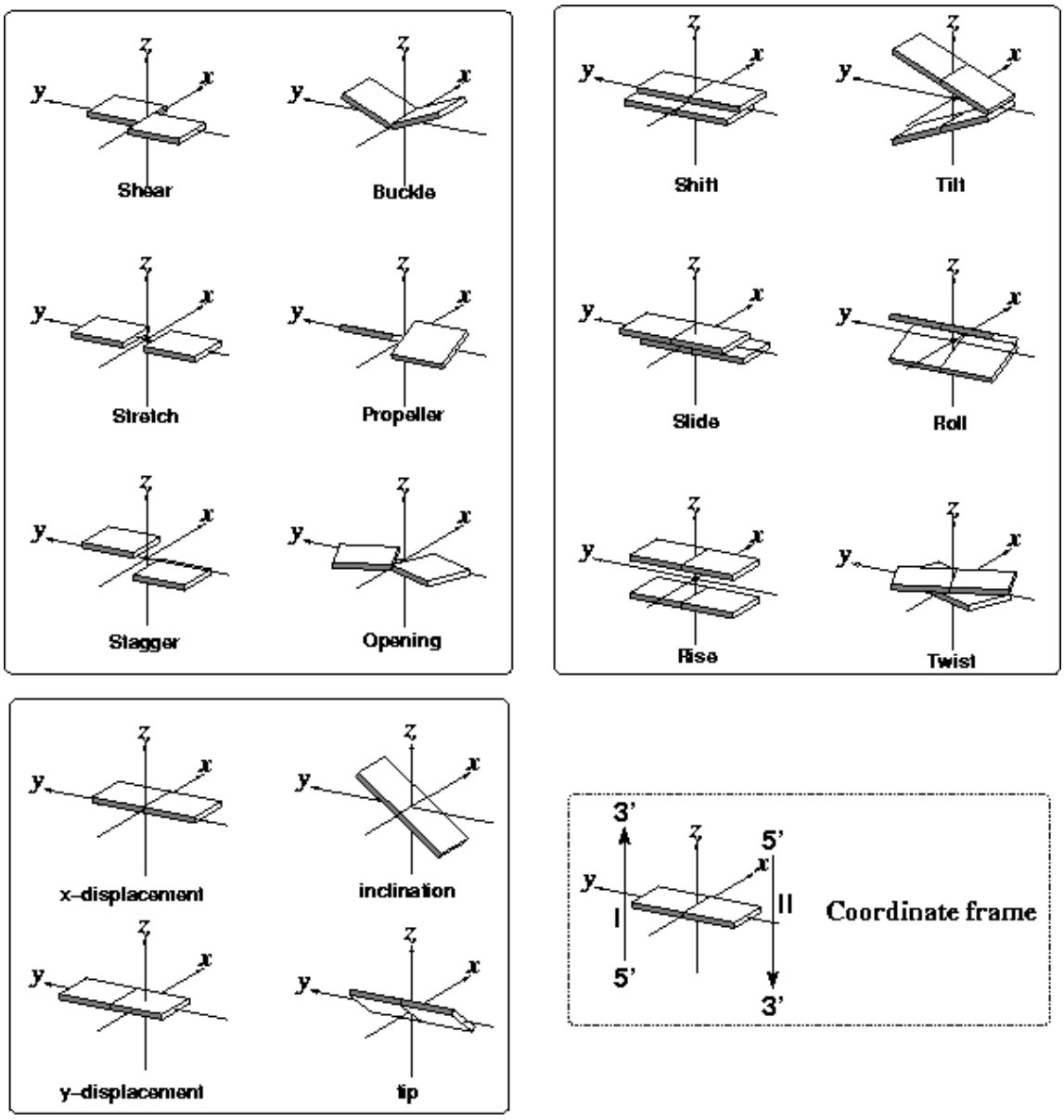

Figure 1.7: Illustration of positive values for the local base pair parameters (upper left), local helical parameters (lower left) and dimer step parameters (upper right) for duplex DNA. The coordinate frame defining the relevant axes is also illustrated (lower right). Adapted from Lu, 2001.

x-axis, the *y*-axis and the *z*-axis, respectively. Shift, slide and rise describe the translation of one base relative to another along the *x*-axis, the *y*-axis and the *z*-axis, respectively and finally, tilt, roll and twist refer to the rotation of one base relative to another along the *x*-axis, the *y*-axis and the *z*-axis, respectively.

Helical conformation can affect some of the base pair geometry parameters more than others. For example, *roll*, *twist* and *slide* are three parameters that change significantly from A-DNA to B-DNA. In particular, the transformation from B- to A-DNA tends to decrease *twist* and *slide*, while increasing *roll* (Olson *et al.*, 2001). The most prominent of these three changes is *slide*. Typical values of *slide* for A-DNA are below -0.8 Å, whereas for B-DNA, typical values are above -0.8 Å. *Roll* and *twist* show more overlap between the two duplex structures. *Twist* angles between 20 ° and 40 ° and *roll* angles between 0 ° and 15 ° can be found in both A- and B-DNA structures (Olson *et al.*, 2001). *Shift*, *rise* and *tilt* appear to be unaffected by helical conformation and are found in both A- and B-DNA structures with a broad range of values. *Opening*, *stretch*, *buckle* and *propeller* show little variation, presumably because of hydrogen bonding and the planar geometry in all double helical structures. With regards to helical parameters, x-displacement is the best parameter to distinguish between A-DNA and B-DNA. Large variations in the local base pair parameters can also help distinguish non-Watson-Crick base pairs and local structural polymorphisms within an oligonucleotide.

1.1.4 Hydrogen Bonding and Base Stacking

There are two different types of non-covalent interactions that play a vital role in maintaining duplex stability. The first interaction is within the plane of the bases and is known as hydrogen bonding. Hydrogen bonding occurs when a hydrogen atom bridges two electronegative atoms; it is mostly governed by electrostatics. The second interaction, known as base stacking interactions or π-π interactions, is perpendicular to the plane of the bases and is stabilized by London dispersion forces and hydrophobic effects (Saenger, 1984). The hydrophobic effect refers to the tendency to minimize the surface area of the bases exposed to water. London dispersion involves induced dipole moments and electrostatic interactions. Base pair hydrogen bonding depends on composition while base stacking is dependent on composition and sequence. In nonpolar

solvents, hydrogen bonding is more prominent and base stacking effects are suppressed. In water, base stacking effects are dominant and hydrogen bonding interactions are negligible due to the competition for binding sites with the water molecules. Base stacking and hydrogen bonding have been recently discovered to influence each other. For example, the hydrogen bonding ability of a stacked base, depends on the hardness and the orientation of the stacking molecule (Mignon *et al.*, 2005).

Between like and different bases, there are 28 different ways the bases can pair together, all forming at least two hydrogen bonds. Why then do we not see all 28 of these different pairings in duplex DNA? Electronic complementarity may be the answer to this question. The stability of a base pair is not simply defined by the number of hydrogen bonds, but also the intrinsic electronic structure of the associated bases (Saenger, 1984). Total interactions energies for the Watson-Crick complementary base pairs indicate that they are considerably more stable than non Watson-Crick base pairs.

Base stacking is the additive stabilization of weak interactions. Purine-purine stacks are the most stable, while pyrimidine-pyrimidine stacks are the least stable. The energy of base stacking interactions is affected by four principle contributors (Hunter, 1993; Hunter *et al.*, 2001); van der Waals interactions, electrostatic interactions between partial atomic charges, electrostatic interactions between the charge distributions associated with the out-of-plane π-electron density and finally electrostatic interactions between the charge distributions associated with out-of-plane π-electron density and the partial atomic charges. Van der Waals interactions are highly attractive and constrain the base pairs in contact, but if the bases are too close, van der Waals interactions will act to repel them. The second contributor involves the non-uniform distribution among the atoms in an aromatic compound. The electronegative atoms in the base, such as nitrogen and oxygen polarize the electron density, hence creating an uneven charge density. Bases will tend to maximize these attractive forces when they stack and this is often the largest single electrostatic interaction. However, this force is dependent on the distance between the base pairs. The greater the distance between the base pairs, the less of an effect partial atomic charges have.

In aromatic compounds, such as the bases in DNA, the nuclei form a positively-charged framework that is “sandwiched” between two negatively-charged regions of π-

electron density. This is the third contributor to the total energy of base stacking interactions. If the bases are stacked in a face-to-face geometry, the electrostatic interactions are unfavourable because the negatively-charged regions are interacting. In an offset geometry, the electrostatic interactions are more favourable because the positively-charged regions are now interacting with the negatively-charged electron density. The final electrostatic interaction to be considered involves the out-of-plane π-electron density and its interaction with the partial charge distribution of the base. This type of interaction plays a crucial role in determining the sequence-dependent effects in DNA. Environmental factors such as solvent and metal ions also play a crucial role in the three-dimensional DNA structure.

1.1.5 Water and Hydration of DNA

Water does more than just dissolve solutes. It is also an active participator in stabilizing the secondary and tertiary structures in macromolecules. For nucleic acids in particular, water and hydrated counterions help to neutralize phosphate-phosphate electrostatic repulsions (Saenger, 1984). One of the major players in determining duplex DNA structure is hydrophobic forces, the arrangement of bases such that exposure to water is minimized. B-DNA is formed in environments with high relative humidity while reduced humidity often leads to the transition from B- to A-DNA. It is obvious that water plays an active role in DNA structure determination. Therefore it is necessary to devote some space in this thesis to understanding the hydration patterns of DNA.

DNA double helices are heavily hydrated having two discrete layers representing primary and secondary hydration shells (Saenger, 1984). The primary hydration shell is impermeable to cations and it is not ice-like (upon exposure to temperatures below 0 °C, the water molecules do not freeze in an ice-like manner). The secondary hydration shell is often indistinguishable from bulk water as far as its permeability to ions and its ice-like characteristics are concerned.

Using molecular dynamic simulations, Auffinger and Westhof (2001) came to some interesting conclusions regarding the primary shell of hydration around duplex DNA. If the entire nucleotide is considered, including the ribose and the phosphate, surrounding the G•C and A•T base pairs in the primary hydration shell, there are 20.1 and 19.8

solvent molecules, respectively (Auffinger and Westhof, 2001). Thus, G•C base pairs are more hydrated than the A•T base pairs. The minor groove hydration pattern is strongly sequence dependent. Water bridges are a structural feature long thought to aid in the stabilization of nucleic acids. Auffinger and Westhof have characterized two long-lived water bridges linking the G N3 and C O2 with the O4' atoms of adjacent nucleotides for C•G, but not for T•A. Therefore, sequences with alternating A•T and T•A base pairs do not favour the formation of a minor groove spine of hydration which had been previously predicted (Saenger, 1984). Based on their conclusions the following thermodynamic order of stability has been reported for the deoxynucleotides pairs: G•C > C•G > A•T > T•A. However, this correlation must not be taken too quantitavely. It is clear that duplex stability is linked to a larger number of factors, such as intra-helical hydrogen bonds, solvent interactions that go beyond the water count in the primary hydration shell and also the interactions with different types of ions.

1.2 DNA-Metal Interactions

Metal ions interact with DNA in such a way that they can either stabilize or destabilize DNA tertiary structure. They have an important influence on the overall structure of duplex DNA. Therefore, in order to study DNA structure, it is important to understand metal-DNA interactions. In DNA, there are four potential sites for metal ion binding and these include the ribose hydroxyls, the base nitrogens, the exocyclic base keto groups and the negatively-charged phosphate oxygens. Different metal cations will prefer a unique set of possible binding locations in DNA.

1.2.1 Interaction Sites and Duplex Stability

Ligands and metal ions are classified as either Class A or Class B according to their preferential bonding. Class A or hard metal cations include the alkali metals, the alkaline earth metals and lighter transition metals in higher oxidation states such as Ti^{4+}, Cr^{3+}, Fe^{3+} and Co^{3+}. Class B, or soft metal cations include the heavier transition metals in lower oxidation states such as Cu^{+}, Ag^{+}, Hg^{+}, Hg^{2+}, Pd^{2+} and Pt^{2+}(Ahrland *et al.*, 1958; Pearson, 1963). The hard metal cations tend to have a relatively small radius, and be only slightly polarizable. In comparison, the soft metal cations have a larger radius and

are more polarizable. Due to the high ratios of positive charge to size, hard metal cations favour ionic bonds with highly electronegative ligands, also known as hard donor atoms. Soft metal cations have relatively low ratios of positive charge to size and, therefore, favour more covalent interactions with less electronegative ligands, or soft ligands. The hard and soft ion classification scheme is relative and there is another group that falls between them. These are the borderline cations and included among them are many of the metal cations from the first transition series, i.e. Fe^{2+}, Co^{2+}, Ni^{2+}, Cu^{2+} and Zn^{2+}. The borderline metal ions will coordinate with either soft or hard ligands (Huheey *et al.*, 1993).

In general, all of the metals in either Class A or Class B will bind non-specifically to the negatively-charged phosphate oxygens. However, the phosphate oxygen atoms are hard ligands and are more often bound to Class A metal ions such as Na^{+} and Mg^{2+}. The metal ion may bind directly to the phosphate oxygen atoms or they may bind through bridging water molecules (Kleveckis and Grisham, 1996). When a cation binds the phosphate group on duplex DNA, it neutralizes the phosphates negative charge and stabilizes its structure. Another hard ligand in duplex RNA is the ribose hydroxyl, which will exclusively bind metals from Class A. However, due to steric hindrance in duplex RNA, it is uncommon to see a metal ion bound to a ribose hydroxyl. The nitrogen atoms in the bases are considered soft ligands and will bind Class B metal cations. The first row transition metal cations that are considered borderline will interact with both phosphate oxygen atoms and the bases.

Inside the Watson-Crick base pairs, there are both electron donor and acceptor regions which are essential for the interaction of cations with DNA. The order of affinity of the different nitrogen atom to bind metal cations is as follows:

$$C\ (N3) > A\ (N1) > G\ (N7) > A\ (N7) > G\ (N1) \approx U\ (N3) > T\ (N3)$$

However, factors such as pH can alter the affinity of nitrogen-cation interactions. There are more potential cation binding sites other than the ones listed above, such as the N3 of G or A which are both not included due to steric hindrance with the sugar. Also not listed are primary nitrogens which are not available for cation binding due to the partial double bond character and charge delocalization within the aromatic ring system. It is

also important to remember the oxygen atoms within the bases which can also interact with metal cations, the most common being the O2 of C.

Coordination of a metal cation to the phosphate group of a polynucleotide tends to stabilize the helix, whereas, coordination of the metal cation to the nitrogenous bases tends to destabilize the helix. Therefore, if we make a list of metal cations in the order of phosphate *vs*. base binding, $Mg^{2+} > Co^{2+} > Ni^{2+} > Mn^{2+} > Zn^{2+} > Cd^{2+} > Cu^{2+}$, the resulting list is also in the order of decreasing ability to stabilize the duplex (Eichhorn and Shin, 1968). The work presented in this thesis focuses mainly on the interactions of Co^{2+}, Ni^{2+} and Zn^{2+} with duplex DNA.

1.2.2 Co^{2+}, Ni^{2+} and Zn^{2+}

The structures of several complexes of Co^{2+} with purine monophosphates have been solved, all yielding the same results as far as where and how the Co^{2+} coordinates to the nucleotide (De Meester *et al.*, 1974b; Aoki, 1975; Gellert *et al.*, 1979; Poojary and Manohar, 1986). The Co^{2+} interacts through octahedral coordination with the N7 and five other water molecules. The metal also binds indirectly *via* the hydration sphere to the O6 and the phosphate oxygen atoms. In all cases, the glycosyl bond is in the *anti* conformation and the sugar has 3'-*endo* pucker. At least two different pyrimidine monophosphate-Co^{2+} complexes have been solved and this time there are some minor differences with regards to Co^{2+} coordination (Clark and Orbell, 1975; Cartwright *et al.*, 1977). In Co^{2+}-5'-CMP, the metal binds N3, O2, one water and two phosphate oxygens each from a different neighboring phosphate creating tetrahedral coordination. In Co^{2+}-5'-UMP, the metal binds N3, two water molecules and four phosphate groups giving the metal an octahedral coordination. Several Co^{2+} deoxyoligonucleotide complex structures have been solved and there appears to be a consensus as to where the Co^{2+} can and cannot bind within the duplex (Labiuk *et al.*, 2003; Valls *et al.*, 2004). In duplex DNA, the Co^{2+} has only been observed binding to the N7 of G. The structure of Z-DNA supports the binding of Co^{2+} to G N7 at both end terminal and within the helix (Gao *et al.*, 1993). B-DNA, however, does not support the binding of Co^{2+} within the helix. Of the Co^{2+}-B-DNA structures solved to date, the Co^{2+} consistently binds to only end terminal G residues. It shows octahedral coordination often with five surrounding water

molecules or a neighbouring end terminal G residue. Based on modelling studies, it has been suggested that Co^{2+} cannot bind within the helix due to steric hindrance. There isn't enough room for octahedral coordination around the metal ion within the helix.

Ni^{2+} and Zn^{2+} coordinate with nucleic acids in much the same way that Co^{2+} does. The Ni^{2+} ion binds five surrounding water molecules and the purine monophosphates N7, giving the metal ion octahedral geometry. The metal binds indirectly, through its hydration sphere, to the O6 as well as to two different phosphate oxygens (Aoki, 1975; Gellert *et al.*, 1979; Pezzano and Podo, 1980). When complexed with B-like deoxyoligonucleotides, Ni^{2+} will have octahedral coordination with five water molecules and the N7 of extra-helical or end terminal G residues. The metal ion is also seen bridging the end terminal G of one duplex with another, hence, stabilizing the unit cell (Abrescia *et al.*, 2002; Labiuk *et al.*, 2003). Zn^{2+}, however, is a little different than either Co^{2+} or Ni^{2+}. In the Zn^{2+}-purine complex, the Zn^{2+} coordinates in a distorted tetrahedral fashion to the N7 of the base and three neighbouring phosphate oxygens (De Meester *et al.*, 1974a; Pezzano and Podo, 1980). In the Zn^{2+}-pyrimidine complex, the Zn^{2+} binds the bases N3, O2, a water molecule and two phosphate oxygens from two neighbouring phosphate groups. The coordination is approximately tetrahedral (Aoki, 1976; Pezzano and Podo, 1980). The way Zn^{2+} coordinates with duplex B-DNA is the same as that of both Co^{2+} and Ni^{2+}, in that it will only interact with end terminal or extra-helical G residues (Soler-Lopez *et al.*, 2002; Labiuk *et al.*, 2003). Most of the deoxyoligonucleotide structures presented here were solved by X-ray crystallography. There are some limitations to X-ray crystallography that may effect where we can see metal ions within the structure. This will be discussed further later on in this thesis.

1.2.3 M-DNA

Metal cations can stabilize a variety of interesting DNA structures including triplex DNA, quadruplex DNA and, most important for this thesis, M-DNA. M-DNA is formed when the transition metals, Co^{2+}, Ni^{2+} or Zn^{2+} bind duplex DNA in alkaline conditions (Lee *et al.*, 1993). This particular type of DNA-metal duplex is of great interest due to its remarkable metallic-like conduction and the possible applications. In order to understand

these properties, we must first understand the distinct structure of M-DNA and how it differs from the more traditional B-DNA.

1.2.3.1 Structure of M-DNA

The currently accepted hypothesis explaining the structure of M-DNA reveals an arrangement similar to B-DNA. According to Eichhorn and Shin (1968) adding Zn^{2+} to duplex DNA at neutral pH will destabilize the DNA, yet at pHs above 8.0, the addition of Zn^{2+} stabilizes M-DNA (Lee *et al.*, 1993), indicating that there must be a conformational change between M- and B-DNA. M-DNA forms with any sequence, including sequences containing 6-methyladenine and 7-deazaadenine, two bases that do not allow Hoogsteen hydrogen bonding. Also, a solution of M-DNA can be rapidly converted back to B-DNA by lowering the pH or by the addition of ethylenediaminetetraacetic acid (EDTA) (Lee *et al.*, 1993). Therefore, to facilitate this conversion, A must still pair with T and G with C in a Watson-Crick fashion. The ultraviolet and circular dichroism spectra of M-DNA are not significantly different than that of B-DNA. From the mobility of linear and closed circular M-DNA compared to that of B-DNA, it can be concluded that M-DNA is right-handed with approximately 10 base pairs per turn (Lee *et al.*, 1993). All of these experiments consistently suggest a right-handed duplex with approximately 10 base pairs per turn and a Watson-Crick base pairing scheme (Aich *et al.*, 1999). Most of these experiments indicate similarities between B-DNA and M-DNA, however, there are also experiments suggesting differences, such as recent studies using capillary electrophoresis (Hartzell and McCord, 2005) and studies using atomic force microscopy (Moreno-Herrero *et al.*, 2003). Atomic force microscopy experiments have indicated a much more compact structure for M-DNA compared with B-DNA. In fact, their measurements show a five-fold shortening in the length of M-DNA as well as increase in the height of almost one order of magnitude. These findings are contradictory to previous studies on M-DNA structure and are most likely due to the precipitation of the DNA. Another difference between M- and B-DNA is the binding of ethidium bromide, namely, M-DNA does not bind ethidium, whereas B-DNA does (Lee *et al.*, 1993).

In alkaline conditions, the deprotonated forms of T N3 and G N1 are more available for metal cation interactions and nuclear magnetic resonance experiments indicate that both the A•T and G•C base pairs lose their imino proton upon addition of the divalent metal ions. Further evidence for the metal binding site is suggested by replacing the bases with modified bases. The pK_a of dT N3, dG N1 and dU N3 are 9.9, 9.4 and 9.3, respectively. Consistent with the pK_as, the U•A base pairs form M-DNA more readily than G•C base pairs, which form M-DNA more readily than A•T base pairs. Substituting T with 5-fluorouracil (5FU), which has a pKa value for the N3 of 7.8, causes M-DNA to form even at pH values below 8.0 (Wood *et al.*, 2002). Therefore, the metal ion appears to be coordinating on T N3 and G N1 within each base pair (Figure 1.8). In order to get a feel for the complete structure of an M-DNA oligonucleotide, molecular modelling was preformed on a 12 base pair helix (Figure 1.9) (Aich *et al.*, 1999). Compared with B-DNA, there is a slight change in the duplex geometry in order to accommodate the metal. The minor groove is opened up due to a 20 ° to 30 ° rotation of the bases and the coordination of the metal is distorted square planar with a water molecule providing the fourth ligand (Aich *et al.*, 1999). The distance between adjacent metal cations is about 4 Å, helping to explain the metallic-like conduction of M-DNA.

1.2.3.2 Unique Properties of M-DNA

M-DNA formation is dependent on many interconnected factors, including, but not limited to, pH, temperature, type of metal cation, DNA concentration, metal concentration and DNA sequence. Out of the divalent metal cations, Mg^{2+}, Co^{2+}, Ni^{2+}, Mn^{2+}, Zn^{2+}, Cu^{2+}, Ca^{2+} and Au^{+}, only Co^{2+}, Ni^{2+} and Zn^{2+} will form M-DNA (Lee *et al.*, 1993). Size may have something to do with the exclusion of the other possible metal cations listed. Cobalt, nickel and zinc all have ionic radaii of 0.75 Å or less, whereas Mn^{2+}, Cu^{2+}, Ca^{2+} and Au^{+} all have ionic radaii of 0.80 Å or more (Cotton and Wilkinson, 1980) and are perhaps too large to be included within the helix. Magnesium has an ionic radius of 0.65 Å, but does not form stable complexes with the nitrogen base. Cobalt, nickel and zinc are small in size and readily coordinate with both nitrogen and oxygen atoms, making them well suited for the interactions necessary to form M-DNA. The concentration of DNA and metal are both key to encouraging the conversion of B- to M-

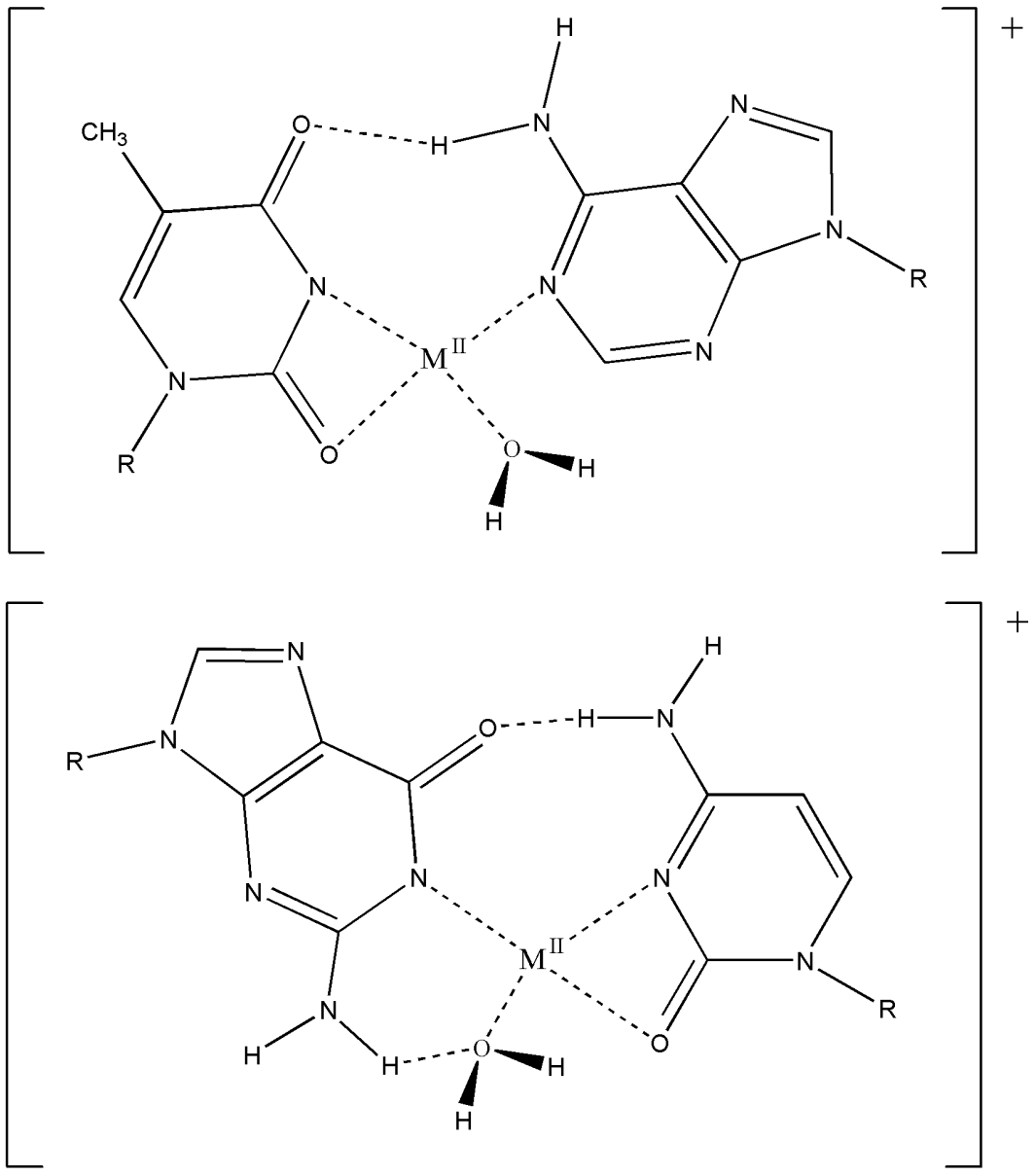

Figure 1.8: Structure of proposed M-DNA base pairs. Shown are the T•A base pair (top) and the G•C base pair (bottom). In M-DNA, the R represents 2'-deoxyribose and the M represents either Co, Ni or Zn.

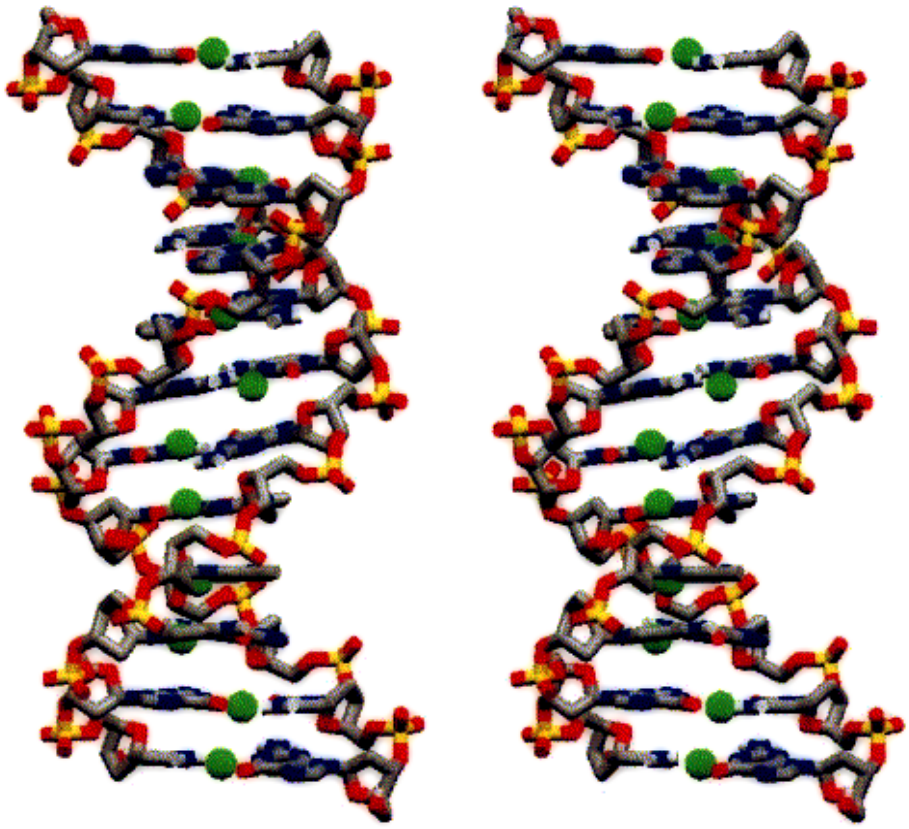

Figure 1.9: Stereo diagram of the M-DNA model calculated using the base pairs in Figure 1.8. The model consists of 12 base pairs with the Zn^{2+} cations coloured green, the oxygen atoms coloured red, the phosphorus atoms yellow, the nitrogen atoms blue and the carbon atoms grey. Adapted from Aich *et. al.*, 1999.

DNA. Increasing the DNA concentration will lead to an increase in the metal ion concentration needed to form M-DNA (Wood *et al.*, 2002). The pH and temperature are also two vital components. The lower the pK_a of the imino proton of the base, the lower the pH at which M-DNA will form (Wood *et al.*, 2002; Wood and Lee, 2005). At temperatures near 0 °C, M-DNA formation is relatively slow, but with temperatures increasing up to 37 °C, the rate of formation becomes quicker (Lee *et al.*, 1993). The extent of M-DNA formation will not vary with base composition; it will, however, vary with different synthetic sequences. The sequence that forms most readily is poly[d(TG)]•poly[d(CA)], followed by poly(dA)•poly(dT), poly[d(TCC)]•poly[d(GGA)], poly[d(TTC)]•poly[d(GAA)], poly[d(GC)] and finally, with the slowest forming sequence being poly[d(AT)] (Lee *et al.*, 1993).

Perhaps the most useful property of M-DNA is its electron transfer capabilities (Wettig *et al.*, 2003). Duplex DNA structures are very promising for nanoelectronic circuits because of their ability to self-assemble predictably based on sequence and their potential electron transfer properties. When DNA is synthesized with a fluorescein fluorophore at one end of the duplex and a rhodamine quencher on the other end, the fluorescein fluorophore is quenched only under M-DNA forming conditions, implying electronic communication through M- and not B-DNA (Aich *et al.*, 1999; Aich *et al.*, 2002; Wettig *et al.*, 2003). Although conductivity experiments on B-DNA are often contradictory, there is now direct evidence that M-DNA has metallic-like conduction (Rakitin *et al.*, 2001). Electrochemical and AC impedance studies have shown M-DNA to have a much faster rate of electron transfer and a much lower resistance than B-DNA, respectively (Li *et al.*, 2003; Long *et al.*, 2003). Because DNA structure is based on sequence, it is possible to engineer different types of structures including three-way junctions. A fluorescence study was done using a three-way junction with two different quenchers on two of the ends. By changing the redox state of one of the quenchers or making a sequence-specific protein to bind one of the branches, it is possible to change the degree of quenching, thereby creating a molecular switch (Aich *et al.*, 1999; Wettig *et al.*, 2003). The potential of M-DNA in nanotechnology is great and the need to study its structure is obvious. As we study more about M- and B-DNA three dimensional structures, eventually the only limit to what we can engineer will be our imaginations.

1.3 X-Ray Crystallography

1.3.1 Crystal Growth and Diffraction

A crystal is a solid in which the constituent atoms, molecules, or ions are packed in a regularly ordered, repeating pattern extending in all three spatial dimensions. Due to their large and irregular shapes, proteins and nucleic acids are not ideally suited for stacking into the periodic lattice required for crystal formation. Protein and nucleic acid crystals typically contain significant amounts of aqueous solvent and are thus very fragile and sensitive to a variety of environmental stresses (Giegé and Ducruix, 1999). The crystals are held together by weak forces such as hydrogen bonds, hydrophobic interactions and salt bridges. Growing suitable crystals is one of the major stumbling blocks for X-ray crystallographers. In order to crystallize, the macromolecule has to separate from bulk aqueous solution and form a distinct, and hopefully, well ordered crystalline solid phase (Rhodes, 1993; Giegé and Ducruix, 1999). In order to create this phase change, one must control the level of saturation and supersaturation in solution. At saturation, there is no net change in the amount of macromolecule in the solid and liquid phases, hence no crystal growth. In supersaturated conditions, the kinetics are in favour of solid formation; this creates the perfect conditions for either crystal growth or precipitation (Rhodes, 1993; Giege and Ducruix, 1999). Temperature and volume are two of the simplest manipulations used to achieve varied levels of supersaturation. The higher the supersaturation, the more nuclei will form and the smaller the crystals will be. The metastable region is a level of supersaturation perfect for crystal growth, and prohibiting nucleation (Giege and Ducruix, 1999). There are many parameters that affect the growth of crystals, including: protein purity, pH of solution, ionic strength, temperature, cleanliness, protein concentration, precipitants, vibration and sound, convection, source and age of macromolecule, as well as the presence of ligands and additives. Each macromolecule is unique and will crystallize under its own set of conditions. Many different conditions must be screened and optimized until a crystal produced is of a diffractable quality.

The quality of the crystal diffraction is critical to determining its structure and ultimately determines the worth of the final crystallographic model. W.L. Bragg demonstrated that the angles of diffracted X-ray intensities relative to the incident beam

produced by X-ray diffraction can be computed by treating them as if the diffracted beams were reflections from sets of equivalent, parallel planes of atoms in the crystal (Rhodes, 1993). One set of parallel planes acts as a single diffractor and produces one reflection. All sets of regularly spaced parallel planes in a crystal can be drawn through lattice indices labelled *hkl*. The indices *h*, *k* and *l* refer to the number of planes in the set per unit cell, in the *x*, *y*, and *z* directions, respectively. According to Bragg's law, only certain conditions will produce diffraction. Bragg showed that a set of parallel planes with index *hkl* and interplanar spacing d_{hkl} produce a diffraction beam when X-rays of wavelength λ impinge on the planes at an angle θ and are reflected at the same angle, only if θ meets the following conditions

$$2\, d_{hkl} \sin\theta = n\lambda \qquad 1.1$$

where n is an integer (Rhodes, 1993). When these conditions are met, the waves emerge from the crystal in phase with one another, interfering constructively and producing a strong diffraction. When these conditions are not met, the waves emerge from the crystal out of phase from one another, interfering destructively and producing a weak or absent reflection. In reality, this equation is a mathematical formalism, as most atoms in protein or nucleic acid molecules do not sit exactly on these planes, and the resultant diffraction amplitude for any particular *hkl* reflection depends on the relative distance of each of the atoms in the molecule normal to the *hkl* plane. This dependence of diffraction amplitude by positions of the N atoms in the molecule perpendicular to the *hkl* plane is given by the structure factor equation.

X-rays are diffracted by electrons. The relationship between the atom and its ability to scatter X-rays is called the scattering factor (*f*) of that atom.

$$f = \int_r \rho(r) e^{(2\pi i r \bullet S)} dr \qquad 1.2$$

$\rho(r)$ is the electron density of the atom at position r and *f* is dependent on |S|, where |S| = (2sinθ)/λ. The scattering of X-rays by a molecule is the summation of the scattering factors for each atom within that molecule and is known as the structure factor (*F*(S))

$$F(S) = \sum_{j=1}^{n} f_j e^{(2\pi i r_j \bullet S)} \qquad 1.3$$

The structure factor depends on both the type of atoms present as well as their location within the unit cell. The spacing of unit cells in the actual crystalline lattice (the real lattice) is inversely proportional to the spacing of unit cells in the lattice on film (the reciprocal lattice) (Rhodes, 1993). Reflections are related to the reciprocal lattice spacing. Therefore, the unit cell dimensions can be determined from the dimensions of the reciprocal lattice, i.e. the direction and number of reflections on the diffraction image. Fourier transform is a mathematical relationship between real space and reciprocal space. Thus, it is used to describe the relationship between an object and its diffraction pattern (Rhodes, 1993). In turn, the electron density at position x, y, z [$\rho(xyz)$] in a crystal is the Fourier transform of the structure factor $F(hkl)$.

$$\rho(xyz) = \frac{1}{V}\int_h\int_k\int_l F(hkl)\exp[-2\pi i(hx+ky+lz)]dhdkdl \qquad 1.4$$

where V is the volume of the unit cell and the structure factors $F(hkl)$ may be expressed as $F(hkl) = |F(hkl)|\exp[i\alpha(hkl)]$; $|F(hkl)|$ is the structure factor amplitude of the reflection (hkl) with coordinates h, k and l in reciprocal space and $\alpha(hkl)$ is the phase angle. Due to Bragg's Law, diffraction data are collected at discrete positions therefore, the above integral is reduced to a Fourier summation

$$\rho(xyz) = \frac{1}{V}\sum_h\sum_k\sum_l |F(hkl)|\exp[-2\pi i(hx+ky+lz)+i\alpha(hkl)] \qquad 1.5$$

Each term in the Fourier series is a simple three-dimensional wave whose frequency is h in the x direction, k in the y direction and l in the z direction. For each of the possible hkl values, the associated wave has an amplitude F_{hkl} and phase α_{hkl} (Rhodes, 1993). Computer programs are readily available to perform these mathematical calculations from the amplitudes of the structure factors collected *via* diffraction experiments. However, there is no way to obtain phase information through diffraction. This is known as the phase problem of X-ray crystallography.

1.3.2 Molecular Replacement and Model Refinement

The diffraction experiment provides the intensities and positions of the reflections, but it provides no information about the phase angle $\alpha(hkl)$. Several methods have been developed to solve the phase problem and provide a starting $\alpha(hkl)$ from which to

initially determine electron density. The most widely used method for structure solving in macromolecule crystallography is molecular replacement, in which estimates of the phases are computed from a previously-determined structure presumed to approximately represent the structure of the macromolecule being studied. The search model should be of a high quality and as close to the unknown structure as possible. In molecular replacement, the search model is related to the unknown structure by six parameters: three rotations α, β and γ and three translations x, y and z. First, the rotation parameters are determined, followed by the translation parameters. In the rotation step, the spatial orientation of the known and unknown molecule is determined with respect to one another. In the translation step, the translation needed to superimpose the now correctly orientated known model onto the true position of the unknown molecule is calculated (Drenth, 1994). In order to compare the search model with the unknown structure, the Patterson functions $P(uvw)$ of the molecules are used. The Patterson function is the summation of the square of the structure factor amplitudes with all phase angles equal to zero.

$$P(uvw) = \frac{1}{V}\sum_{hkl}\left|F(hkl)\right|^2 \cos\left[2\pi(hu + kv + lw)\right] \qquad 1.6$$

Each peak in the Patterson function can be shown to be the product of two atomic scattering factors separated by some vector u,v,w. In other words, the Patterson function is a pairwise sum of interatomic vectors, each term being the product of the electron density of the two contributing atoms. If two atoms are within the same molecule, the distance of the vector will be small; these vectors represent the self-Patterson vectors and are used to determine the rotation parameters. If the two atoms are from two different molecules, the vector will be relatively large; these vectors are referred to as cross-Patterson vectors and are used to determine the translation parameters.

In order to determine the correct rotation orientation for the search model, it is first placed in an arbitrary unit cell and moved to the origin followed by calculation of its diffraction pattern or Patterson function. The Patterson function is then calculated for a large number of different orientations. The solution that has a good overlap between the search models Patterson function and that of the unknown structure most likely represents the correct orientation of the model in the unknown structures unit cell. Next,

the correct position of the molecule within the unit cell is determined by translation. This is often done using the cross-Patterson vectors within the correct unit cell and space group of the unknown structure.

Once the search model is orientated and located within the unit cell of the unknown structure, it is necessary to refine the model because the initial phases are generally not satisfactorily accurate. The first step in refinement involves manually changing the model to correspond to the correct sequence with that of the unknown molecules. Once this is complete along with any other major change that has to be made, more automated refinement processes may be used. It is necessary to restrict the model building and refinement using known data concerning protein or nucleic acid structure. Some possible restraints include bond lengths, bond angles, van der Waals contact distances, keeping planar groups planar and maintaining the chirality of chiral centers (Rhodes, 1993; Drenth, 1994). Once the model is sufficiently refined, water molecules are often added to the electron density map. The ultimate goal of model refinement is to obtain a model that is as closely related to the original diffraction data as possible.

The R-factor and free R-factor are two mathematical representations of how close the model is to the diffraction data. The lower the R-factor, the better the model. The R-factor, otherwise known as the agreement factor or residual factor, determines the mathematical agreement between the observed and calculated structure factors

$$R = \frac{\sum |F_O| - |F_C|}{\sum_{hkl} |F_O|} \qquad 1.8$$

The problem of overfitting the model can be circumvented by using most, but not all, of the diffraction data to refine the atomic model (Drenth, 1994). The remaining data can then be used to verify how well the model fits by calculating the free R-factor.

1.3.3 DNA Crystallography

The first step in any crystallographic experiment is the production of large, well diffracting crystals. This step is often considered the "bottleneck" of crystallography. It is the most limiting step in obtaining a final crystallographic model. Although DNA oligonucleotides are no more difficult than proteins to crystallize, good diffracting crystals of oligonucleotides are much less common (Timsit and Moras, 1992). Crystals

of nucleic acids have several tendencies that make their average diffraction quality less than that of protein crystals. For one, the plasticity of DNA structures helps contribute to the lack of good diffracting crystals. Several DNA forms can coexist in the same crystal. One oligonucleotide sequence can crystallize in several different structures and crystal packing arrangements. The DNA molecule is highly hydrated and the crystals often have a very large solvent content resulting in a high degree of freedom for the phosphodiester backbone, ions and solvent (Subirana and Soler-Lopez, 2003). It should also be noted that the crystallization environment can stabilize important DNA structural alterations and induce structural transitions, such as the high salt induction of B-Z transitions (Timsit and Moras, 1992). Because the creation of good quality crystals is the most limiting step in X-ray crystallography, it is necessary to examine the tendencies specific to the crystallization of oligonucleotides and how these can affect diffraction quality.

1.3.3.1 Choosing Oligonucleotides for Crystallization

The first step in producing large, well diffracting crystals is choosing and preparing the oligonucleotide. Perhaps even more important than screening crystallization conditions is screening different oligonucleotide sequences (Scott *et al.*, 1995). One can choose a sequence based on its biological significance, ease of preparation or its crystallization potency. For example, about 85% of all B-form duplexes solved to date start with C and only two sequences have been crystallized consisting purely of AT base pairs (Egli, 2004). To try and crystallize a sequence consisting of only AT base pairs may be of more interest, but it has also proven to be more difficult. If you want to choose a sequence that is more likely to crystallize, it is wise to start the sequence with a C residue. The DNA sequence can also help to determine crystal packing. For example, B-form duplexes with terminal CGCxxxxxGCG boxes preferentially lead to crystallization in the orthorhombic $P2_12_12_1$ space group by favouring specific hydrogen bonding between the minor grooves of the duplexes, whereas having C residues at particular positions will promote major groove-backbone interactions and lead to crystallization in the trigonal R3 space group. The practical application of this is to design B-DNA molecules containing these crystal packing driving boxes in order to

guide the crystallization into well-defined crystal lattices (Timsit and Moras, 1992; Dock-Bregeon *et al.*, 1999).

Another important feature for the crystallization of an oligonucleotide is its size. There are definitely tendencies with regards to the size of the oligonucleotides that have been reported in the Nucleic Acid Database (Figure 1.10) (Berman *et al.*, 1992). It is apparent from Figure 1.10 that oligonucleotides crystallize more readily with either six, eight, ten or twelve base pairs. Yet, there are also more specific tendencies among the different forms of oligonucleotides. For example, almost all of the Z-DNA structures are six base pairs and most of the B-form (both DNA and RNA) structures are either twelve or ten base pairs. The A-form structures are mostly eight or ten base pairs, however they exhibit more variation than either of the B or Z-forms (Figure 1.10). As indicated in Figure 1.10, there are several structures of an unusual type that is neither a B-, A- or Z-form structure. This group contains interesting structures such as triplexes, quadruplexes, and junctions just to name a few. It is obvious that even numbers of base pairs result in more successfully solved structures. Perhaps this is because people most often choose palindromic self-complementary sequences which are favoured for crystallography. With self-complementary sequences, an odd number would mean either a mismatched base pair in the sequence or an overhanging end. It is important to consider all of these factors when designing an oligonucleotide for crystallization.

1.3.3.2 Crystallization Conditions

In order to obtain a crystal, the molecules must assemble into a periodic lattice. The first step is to make a highly concentrated solution of the oligonucleotide and then slowly bring it to a supersaturated state. If the environment is favourable for nucleation and the formation of the first ordered aggregates, a small crystal may start to grow. Finding the right conditions can be hit-and-miss and often quite a lot of conditions must be tried in order to succeed. However, with nucleic acids there are some general features worth considering. For example, the most often used precipitant in nucleic acid crystallization is 2-methyl-2,4-pentanediol (MPD) (Dock-Bregeon *et al.*, 1999). Nucleic acids are less sensitive to pH than proteins, but it is still an important factor in crystallization trials as the pH may have an effect on crystal packing or perhaps on the

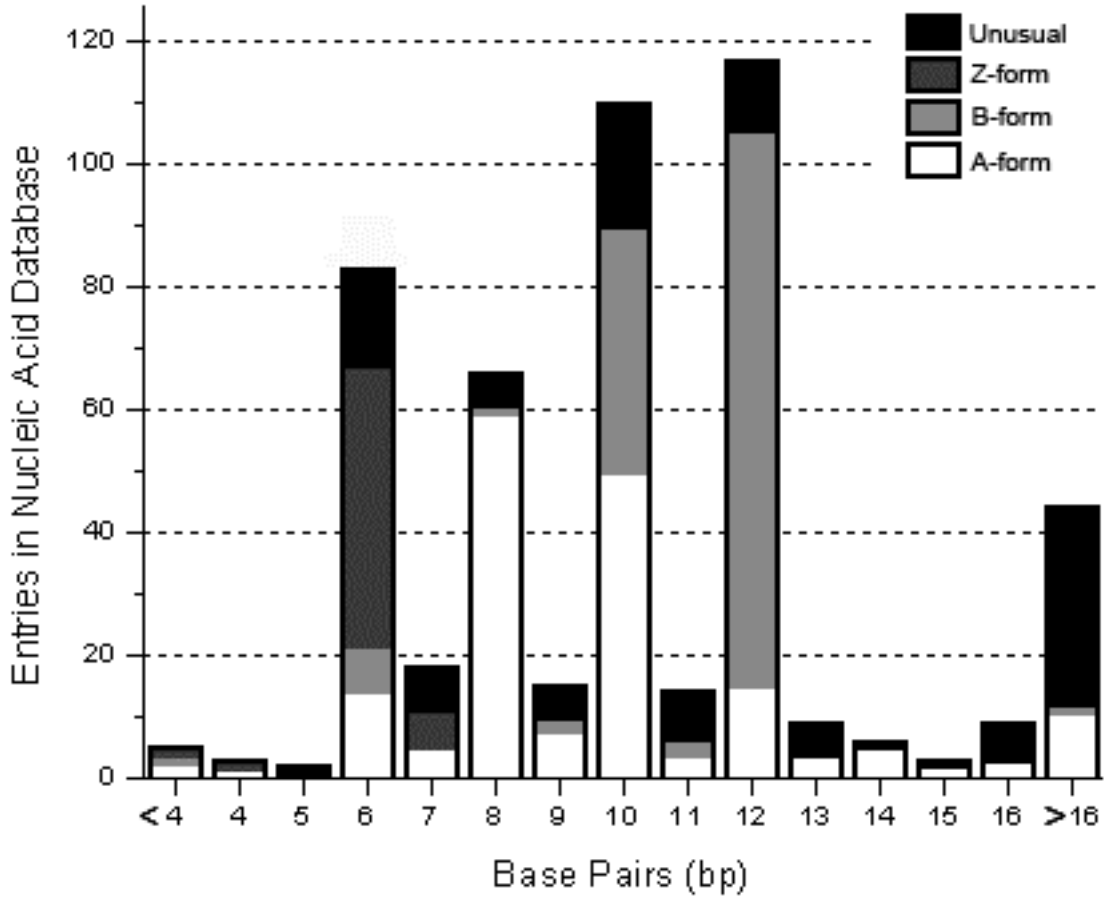

Figure 1.10: Oligonucleotide size in base pairs verses the number of entries in the Nucleic Acid Database. The Nucleic Acid Database compiles all of the nucleic acid structures solved and reported. There are definite tendencies regarding the sizes in which different forms of oligonucleotides crystallize. The number of entries only includes DNA or RNA structures. In other words, it does not include DNA or RNA complexed with protein or drug molecules.

structural form of the oligonucleotide. It is important to remember that nucleic acids are polyelectrolytes and therefore counterions are important additives for crystallization.

Two different types of additives are generally used, polyamines and metal cations (Dock-Bregeon *et al.*, 1999). Some examples of polyamines often used in crystallization trials include spermidine and spermine. The most often used polyamine is spermine, a linear molecule with four positive charges at neutral pH. Although spermine is used in many crystallization trials, the detection of spermine molecules within the crystals of B-DNA is a difficult task. According to molecular dynamic computer simulations, this is because spermine, as a flexible polyamine molecule, has several binding modes and interacts irregularly with different sites within the DNA. Spermine molecules do not form structurally stable complexes with DNA (Korolev *et al.*, 2001). Spermine, sodium and water molecules often compete with each other to bind the bases within the minor groove. Together, they influence the structure of the DNA hydration shell. Polyamines can bind within the minor groove or form bridges between different DNA helices in the crystal as well as across the major and minor groove.

Divalent and monovalent ions are good at neutralizing some of the negative charge on the oligonucleotides. They are often used to help stabilize different crystal packing arrangements; however, care must be given when using these metal ions because they can also induce structural changes in the helical structure. It is important to consider these factors when designing a screen for initial crystal growing conditions. It can be hit-and-miss and the key is to try as many different conditions as possible. Crystals of poor quality are very common with nucleic acids. This may be due to the geometry of the helices, which can pack easily despite rotational disorder. The key to improving crystal quality is to introduce structural change by varying the additives, temperature or pH. Another option is the addition of small molecules which could act as a lever promoting lattice building. If poor diffraction remains an issue, as stated above, one of the most important factors for getting crystals that diffract well is sequence. It may be necessary to go back to the drawing board and start by screening different oligonucleotide sequences.

1.4 X-ray Absorption Spectroscopy

X-ray absorption fine structure (XAFS) is a practical way to determine the chemical state and local atomic structure for a selected atomic species. EXAFS spectra measures X-ray absorption by an atom at energies near and above the core-level binding energies of that atom (Newville, 2004). EXAFS at the metal *K*-edge will not tell a lot about the complete three dimensional structure of duplex DNA, but may give some important information regarding the local structure surrounding the divalent metals. In combination with X-ray crystallography, EXAFS may play a vital role in determining the interactions of metal cations with duplex DNA. Unlike crystallography, crystallinity is not essential for EXAFS data collection. EXAFS spectra can be collected for non-crystalline, highly disordered materials. The X-ray absorption spectra are divided into two regimes, namely the fine structure in absorption close to the X-ray edge known as X-ray absorption near edge spectroscopy (XANES) and the fine structure in the absorption well above an X-ray edge or extended X-ray absorption fine-structure spectroscopy (EXAFS) (Rehr and Albers, 2000). XANES is strongly sensitive to the coordination chemistry and formal oxidation state of the absorbing atom, whereas, EXAFS is more often used to determine the distances, coordination number and the species of the neighbours of the absorbing atom (Newville, 2004). For purposes regarding this thesis, it is important to focus on how EXAFS yields geometric structure.

1.4.1 X-ray Absorption

X-rays are absorbed by all matter and emit electrons through a process known as the photo-electric effect described by Einstein in 1905. In the photo-electric effect, an X-ray photon is absorbed and a core level electron is promoted out of the atom, which leaves the atom with an empty electronic level (Figure 1.11) (Reinert and Hufner, 2005). The energy of the incident X-ray must be greater than the binding energy of the electronic core level in order for the core level to participate in the absorption. When this occurs, the X-ray is absorbed, the electron is removed and the excess energy is given to the photo-electron that is ejected from the atom (Newville, 2004). When the incident X-ray's energy is equal to the binding energy of a core level, there is a sharp increase

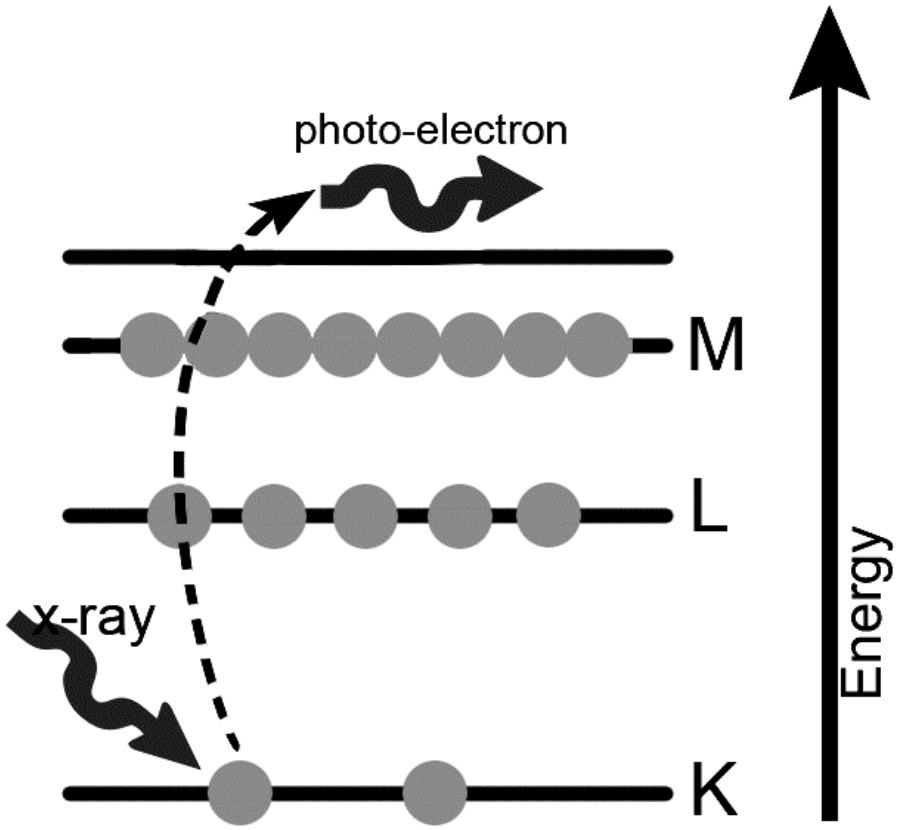

Figure 1.11: Diagram illustrating the photo-electric effect. During photo-electric effect an X-ray is absorbed into an atom and a core level electron is promoted out, resulting in an empty electronic level and a photo-electron. M, L and K represent electronic core levels.

in absorption, corresponding to the absorption edge. The absorption process is governed by the dipole selection rules and the normalized edge jump reflects the unoccupied final state per atom. There are two main methods in which the excited atom decays. The first is X-ray fluorescence, where a higher energy core-level electron fills the hole created by the absorption event and an X-ray of a well-defined energy are ejected (Newville, 2004). The second mechanism is the Auger effect, in which an electron drops from a higher electron level and a second electron is ejected out of the atom (Attwood, 2000). Every atom has core-level electrons with well defined binding energies. Therefore, the X-ray energy can readily be adjusted to probe a specific core level of a particular element.

The outgoing photo-electron wave scatters from the electrons of the neighbouring atoms and can also return to the absorbing atom. The absorption coefficient is modulated by the presence of a photo-electron backscattering from the absorbing atoms neighbours (Teo, 1980). This oscillation in the absorption coefficient is the EXAFS. The distance between the neighbouring atoms and the absorbing atom directly affects the frequency of each EXAFS wave because the photo-electron must travel from the absorbing atom to the neighbouring atom and back. The amplitude of each EXAFS wave depends directly on the number and types of neighbouring atoms as well as their distance from the absorber (Teo, 1980). Therefore, from analysis of the frequency and amplitude of the EXAFS waves, it is possible to determine the distance as well as the number of each type of atom surrounding the absorber.

1.4.2 The EXAFS Equation

For EXAFS, we are most interested in the absorption coefficient, μ, which gives the probability of an X-ray being absorbed according to Beer's law,

$$I = I_0 e^{-\mu t} \tag{1.9}$$

where I_0 is the X-ray intensity incident on a sample, I is the intensity transmitted through the sample and t is the sample thickness (Newville, 2004). The absorption coefficient is most often a smooth function of energy that depends on sample density, atomic number, atomic mass and X-ray energy, E. In EXAFS, The absorption coefficient can be measured either in transmission as

$$\mu(E) = \log(I_0 / I) \tag{1.10}$$

or in X-ray fluorescence (or Auger emissions) as

$$\mu(E) \propto \log(I_f / I) \tag{1.11}$$

where I_f is the intensity of fluorescence or electron emission and μ(E) is the measured absorption coefficient (Newville, 2004).

Typically the XANES spectrum is within 30 eV of the main absorption edge, whereas the oscillations well above the absorption edge are studied for the EXAFS spectrum (Rehr and Albers, 2000). The EXAFS spectrum, χ, is defined as the normalized, oscillatory part of the X-ray absorption above a given absorption edge, i.e.,

$$\chi(E) = \frac{\mu(E) - \mu_0(E)}{\Delta\mu_0(E)} \tag{1.12}$$

where μ_0(E) is the background function, and $\Delta\mu_0$ is a normalization factor that is often approximated as the measured jump in absorption at the threshold energy, E_0 (Teo, 1980; Rehr and Albers, 2000; Newville, 2004). However, EXAFS are best understood in terms of structural parameters. Therefore, it is useful to convert the E into the photo-electron wave vector, *k*, by the relationship

$$k = \sqrt{\frac{2m(E - E_0)}{\hbar^2}} \tag{1.13}$$

where *m* is the electron mass and ħ refers to the reduced Planck's constant or Dirac's constant (Teo, 1980; Newville, 2004). Transforming of $\chi(E)$ into $\chi(k)$ in *k* space allows for writing the EXAFS equation as

$$\chi(k) = \sum_j \frac{N_j f_j(k) e^{-2k^2\sigma_j^2} e^{-2R_j/\lambda(k)} f_j(k)}{kR_j^2} \sin\left[2kR_j + \delta_j(k)\right] \tag{1.14}$$

where $f(k)$ is the scattering amplitude of the neighbouring atoms, $\delta(k)$ is the phase-shift of the neighbouring atoms, *N* is the number of neighbouring atoms, *R* is the distance to the neighbouring atom, σ^2 is the mean-square-displacement in *R*, j is the individual coordination shell of identical atoms at approximately the same distance from the absorbing atom and finally, λ is the mean-free-path of the photo-electron (i.e. how far it usually travels before scattering inelastically and before the core-hole is filled) (Teo, 1980; Newville, 2004). From this equation, it is possible to make some conclusions about EXAFS. If we know $\lambda(k)$, $f(k)$ and $\delta(k)$, the EXAFS equation allows us to

determine N, R, and σ^2 (Newville, 2004). Because of limitations such as $\lambda(k)$, EXAFS is considered only a local probe and can not give information much further than about 5 Å from the absorbing atom. The scattering factors depend strongly on the atomic species, Z, of the neighboring atoms and therefore Z can in principle be determined *via* EXAFS analysis (Newville, 2004). The EXAFS oscillations will consist of varying frequencies corresponding to each coordination shell's distance; hence, it is necessary to use Fourier transforms in the analysis to separate the overlapping oscillations.

1.4.3 EXAFS Measurements and Data Analysis

In EXAFS analysis, it is essential to have very accurate and precise measurements. It is also necessary to have access to a tunable energy source. Most often the X-ray source used for EXAFS is synchrotron radiation, which can fulfill both of these requirements (Newville, 2004). In recent years third-generation synchrotron radiation sources have undergone significant improvements in beamline optics, controls and detectors. Good quality EXAFS data on millimolar concentration samples can be obtained within seconds or minutes (Bunker *et al.*, 2005). For transmission measurements, it is important that the sample is homogeneous and free from pinholes. Measuring fluorescence can be somewhat more difficult than transmission. However, it is often preferred for thick samples or samples with a lower concentration (element of interest < 10% of sample) (Newville, 2004). Regardless of whether the absorption coefficient is measured in transmission or fluorescence, the data analysis is essentially the same (Newville, 2004).

Before the raw data can be analyzed using the EXAFS equation, they must be first reduced. The first step in data reduction is converting the measured intensities to $\mu(E)$ and then subtracting a smooth pre-edge function in order to eliminate any instrumental background and absorption from other edges (Newville, 2004; Bunker *et al.*, 2005). Next, the threshold energy, E_0, must be identified and $\mu(E)$ is normalized to go from 0 to 1. Then a smooth post-edge background function is required to estimate $\mu_0(E)$ and isolate the EXAFS, $\chi(k)$ (Newville, 2004). The final step is to k-weight the EXAFS and Fourier transform into R-space (Newville, 2004). The EXAFS are k-weighted usually k^2 or k^3, in order to compensate for the diminishing amplitudes at high k values (Teo, 1980; Newville, 2004). The Fourier transform step is critical to isolate and identify different

coordination spheres around the absorbing atom. Now the sample structure is ready to be analyzed using the EXAFS equation.

First, the scattering amplitude and phase shifts are theoretically calculated using computer programs such as FEFF (Ankudinov *et al.*, 1998). These theoretical scattering factors are then used in the EXAFS equation to refine structural parameters from the data. The structural parameters, N, R, and σ^2, and also E_0 are allowed to change until the best-fit to $\chi(k)$ is achieved for the data (Teo, 1980; Newville, 2004). This refinement may be done with either the measured $\chi(k)$ or the Fourier transformed data. Working in R-space is often preferred, as it allows us to selectively ignore higher coordination shells (Newville, 2004). Care must be taken when choosing the theoretical standards to calculate the scattering factors. Similar to crystallography's Molecular Replacement, there may be bias placed on the structure based on the model chosen. Neither crystallography nor EXAFS are 100% accurate structural analysis tools. Both have their limitations and their strengths that must be carefully weighed and considered.

1.5 Justification and Objectives

There remains a lot still unknown regarding the structure of M-DNA and until an X-ray crystallographic model is available, M-DNA is unlikely to find wide-spread acceptance. M-DNA has a wide variety of exciting potential future applications. However, until we can be completely confident in its three dimensional structure, it will be difficult to fully understand how M-DNA functions and how best to utilize it. The ultimate goal of the work presented in this thesis is the structural analysis of M-DNA. If crystal packing arrangements can be found for DNA grown in conditions that stabilize M-DNA, crystallography will be a vital tool in the analysis of M-DNA structure. Using tools such as EXAFS and X-ray crystallography together will be important in finally deciding on an accurate model for the three-dimensional structure of M-DNA.

In order to study M-DNA structure, several DNA sequences have been screened in conditions favouring the formation of M-DNA, with a particular emphasis on sequences containing the modified base 5FU. Because 5FU has a lower pK_a for its imino proton, M-DNA is stabilized in a less alkaline environment, hence, providing a more flexible pH range for the crystallization conditions. X-ray crystallographic data were collected for

the deoxyoligonucleotides d(CGUGUGCACACG) in the presence of Zn^{2+} and d(GAUUAAUUC) in the presence of Co^{2+} (where U = 5FU). The data have been analyzed *via* Molecular Replacement using a B-DNA structure as a model. Due to the limitations of crystallography on nucleic acids, EXAFS analysis was also performed on M-DNA samples. The two methods combined were used cooperatively to gain as much knowledge as possible about the three-dimensional structure of M-DNA. For comparison purposes, both Ni^{2+}-M-DNA and B-DNA complexed with Ni^{2+} samples were studied using EXAFS. Ultimately, the work presented in this thesis hopes to either confirm or discard the M-DNA structure accepted to date.

2.0 MATERIALS AND METHODS

2.1 Reagents, Supplies and Equipment

Biological and chemical reagents, supplies and equipment used in the experiments throughout this thesis along with their commercial suppliers, are listed in Table 2.1. The addresses of the companies are listed in Table 2.2.

2.2 Crystal Structures of DNA-Metal Complexes

2.2.1 Oligonucleotide Screening

Crystallization trials began with the screening of seventeen different deoxyoligonucleotide sequences including two sequences with no modified bases, d(GGCTAGCC) and d(CGTGTGCACACG), two sequences containing 5-bromo-deoxyuridine (5BrU), d[GG(5BrU)ACC] and d[A(5BrU)(5BrU)AA(5BrU)], two containing inosine (I), the nucleoside of hypoxanthine, d(IICICC) and d[I(5BrU)A(5BrU)AC], one sequence containing 4-thiothymidine (s^4T), d[(s^4T)G(s^4T)G(s^4T)GCACACA] and the remaining ten sequences all contain 5-fluorodeoxyuridine (5FU), d[CG(5FU)G(5FU)GCACACG], d[GA(5FU)(5FU)AA-(5FU)C], d[GG(5FU)ACC], d[C-(5FU)AA(5FU)(5FU)AG], d[GA(5FU)A(5FU)C], d[GA(5FU)AUCGC], d[CG(5FU)A(5FU)GCA(5FU)ACG], d[CGAA(5FU)(5FU)AA-(5FU)(5FU)CG], d[(5FU)ACAA(5FU)(5FU)G] and d[CA(5FU)A(5FU)G]. The phosphoramidite method was used to synthesize all of these sequences (Caruthers *et al.*, 1992), followed by purification from trityl-on reversed phase high performance liquid chromatography (Brown and Brown, 1992). Finally, the oligonucleotides underwent desalting and removal of the dimethoxytrityl groups by ethanol precipitation. The procedures mentioned so far were carried out by the suppliers. Once the samples were received, they were dissolved in distilled autoclaved water and stored at -20 °C. Other chemicals used for the crystallization trials were stored at 4 °C in plastic Falcon tubes.

Table 2.1: Biological and chemical reagents, supplies and equipment.

Item	Supplier
Biological reagents	
Synthetic d[CG(5FU)G(5FU)GCACACG]	Alpha DNA
Synthetic d[GA(5FU)(5FU)AA(5FU)C]	Alpha DNA
Synthetic d[GG(5FU)ACC]	Alpha DNA
Synthetic d[GG(5BrU)ACC	Alpha DNA
Synthetic d[I(5BrU)A(5BrU)AC]	Alpha DNA
Synthetic d[A(5BrU)(5BrU)AA(5BrU)]	Alpha DNA
Synthetic d(IICICC)	Synthegen
Synthetic d(GGCTAGCC)	Synthegen
Synthetic d(CGTGTGCACACG)	Synthegen
Synthetic d[C(5FU)AA(5FU)(5FU)AG]	Alpha DNA
Synthetic d[GA(5FU)A(5FU)C]	Alpha DNA
Synthetic d[GA(5FU)AUCGC]	Alpha DNA
Synthetic d[CG(5FU)A(5FU)GCA(5FU)ACG]	Alpha DNA
Synthetic d[CGAA(5FU)(5FU)AA(5FU)(5FU)CG]	Alpha DNA
Synthetic d[(5FU)ACAA(5FU)(5FU)G]	Alpha DNA
Synthetic d[CA(5FU)A(5FU)G]	Alpha DNA
Synthetic d[(s^4T)G(s^4T)G(s^4T)GCACACA]	Alpha DNA
Chemical reagents	
Aqua sil – siliconizing agent	Hampton
Cobalt Chloride hexahydrate ($CoCl_2$•$6H_2O$)	Sigma
Cobalt Perchlorate hexahydrate ($Co(ClO_4)_2$·$6H_2O$)	Sigma
Glycerol	BDH
Hydrochloric acid (HCl)	BDH
(±)-2-Methyl-2,4-pentanediol (MPD)	Fluka
Nickel Chloride hexahydrate ($NiCl_2$•$6H_2O$)	Sigma
Nitrogen (liquid)	Praxair
Sodium Hydroxide (NaOH)	BDH
Spermine Tetrahydrochloride	ICN
N-Tris-(hydroxymethyl)methyl-3-aminopropanesulfonic acid, sodium salt (TAPS)	ICN
Zinc Chloride	Sigma
Supplies and equipment	
Accumet Basic pH Meter	Gilford
Cryoloops	Hampton
Dow Corning High Vacuum Grease	VWR
Falcon Tubes	VWR
Microcentrifuge Tubes	VWR

Micro Cover Glass (No. 1, 22 mm square)	VWR
Silicon Graphics Indigo^2 Computer	SGI
Syringes	Becton Dickinson
24-Well VDX Plates	Hampton
X8 Proteum	Bruker AXS

Table 2.2: Names and addresses of suppliers.

Company	Address
Alpha DNA	Alpha DNA, Montreal, PQ, Canada.
BDH	British Drug House, Saskatoon, SK, Canada.
Becton Dickinson	Becton Dickinson and Co., Franklin Lakes, NJ, USA.
Bruker AXS	Bruker AXS, Inc., Madison, WI, USA.
Fluka	Sigma-Aldrich Canada Ltd., Oakville, ON, Canada.
Gilford	Gilford Instrument Laboratories Inc., Oberlin, OH, USA.
Hampton	Hampton Research, Lugana Niguel, CA, USA.
ICN	ICN Biomedical Canada Ltd., Saint Laurent, PQ, Canada.
Praxair	Praxair, Saskatoon, SK, Canada.
SGI	Silicon Graphics Inc., Mountain View, CA, USA.
Sigma	Sigma Chemical Co., Saint Louis, MO, USA.
Synthegen	Synthegen, LLC, Houston, TX, USA.
VWR	VWR, Mississauga, ON, Canada.

The crystals were grown by vapour diffusion using the hanging drop method (Rhodes, 1993). In hanging drop vapour diffusion, 1-8 μL drops are placed on a square glass cover slip that is then placed over the reservoirs of a 24-well plate and sealed by vacuum grease. In most trials, to keep the drops compact and uniform, the cover slips were treated with Aqua sil, a siliconizing agent that provides a hydrophobic water repellent surface. The cover slip treatment involved rinsing them in the Aqua sil solution, followed by distilled water and then allowing them to air dry overnight. Inside the reservoir of each trial is a precipitant solution. Among other things, this solution contains a higher concentration of precipitant, most often 2-methyl-2,4-pentanediol (MPD), than does the drop on the cover slip. This allows a slow equilibrium to develop over time between the drop and the reservoir solution. As the water vapour diffused from the drop to the reservoir solution, the concentration of MPD and DNA-metal complexes in the drop slowly increased, hence encouraging crystal formation and growth. A few other vapour diffusion methods were experimented with, but the hanging drop method proved to be the most consistently successful. All of the crystallization trials were set up at room temperature and then allowed to continue either at room temperature or 4 °C. Crystal growth was observed using an optical microscope and pictures were recorded through a camera mounted on the microscope.

2.2.2 Crystallization Conditions

2.2.2.1 d[GA(5FU)(5FU)AA(5FU)C]-Metal Complexes

The d[GA(5FU)(5FU)AA(5FU)C] oligomer was crystallized and diffracted in the presence of both cobalt and zinc. The crystallization conditions were discovered and optimized through trail and error. The d[GA(5FU)(5FU)AA(5FU)C]-Co^{2+} complex was crystallized at 4 °C in either a four or a one μL drop containing 0.75 mM of oligomer, 40 mM N-tris-[hydroxymethyl]methyl-3-aminopropanesulfonic acid (TAPS) (pH 8.5), 6.0 mM of $Co(ClO_4)_2$, 2.0 mM of spermine, 2% (v/v) of glycerol and 5% (v/v) of MPD. The drop was equilibrated against a either a 0.8 mL or a 0.4 mL reservoir containing 40 mM TAPS (pH 8.5), 6.0 mM $Co(ClO_4)_2$, 2% glycerol, and 30% MPD. The d[GA(5FU)(5FU)AA(5FU)C]-Zn^{2+} complex has been crystallized using the same method and solutions as the Co^{2+} complex, with the following exceptions, the $Co(ClO_4)_2$

was replaced with 0.50 mM of $Zn(ClO_4)_2$, the pH was 7.7, the reservoir consisted purely of 40% MPD and the concentration of DNA in the drop was 1.25 mM.

2.2.2.2 d[CG(5FU)G(5FU)GCACACG]-Zn^{2+} Complex

Crystallization conditions for growing crystals of the d[CG(5FU)G(5FU)GCACA-CG]-Zn^{2+} complex were determined by trail and error and optimized by Shaunivan Labiuk. The crystals grew at 4 °C in 2 μL drops containing 0.70 mM of d[CG(5FU)G(5FU)GCACACG], 1.8 mM of $ZnCl_2$, 40 mM of TAPS at pH 7.75, 2.0 mM of spermine, 7% (v/v) of glycerol and 5% (v/v) of MPD. The drops were equilibrated against a 0.4 mL reservoir containing 40 mM TAPS at pH 7.75, 1.8 mM $ZnCl_2$, 7% glycerol, and 45% MPD.

2.2.3 Cryoprotection of Crystal Samples

For diffraction data collected at the Advanced Photon Source (APS) at Argonne National Laboratory (Argonne, IL, USA), the crystals were mounted in a cryoprotectant solution prior to immediate cooling in liquid nitrogen. Cryoprotectant solutions consisted of the same components that make up the drops in the crystallization trials with the following exceptions; no DNA oligomer or metal and an increased concentration of MPD and glycerol. MPD concentrations of 50%, 55%, 60% and 65% (v/v) were combined with either 5% or 10% (v/v) glycerol. Before mounting the crystals, the cryoprotectant solutions were brought to the temperature of the crystals to be mounted. Under a microscope, about 10 μL of cryoprotectant was added to the drop containing a crystal to be mounted. The crystal was then transferred into a drop containing the pure cryoprotectant solution, from which it was mounted on a loop of an appropriate size and dipped immediately into liquid nitrogen. Crystals were stored mounted on the loop under liquid nitrogen until being transferred in a dry shipper to the synchrotron radiation source.

For the diffraction data screened on the DX8 Proteum diffractometer at the Saskatchewan Structural Sciences Centre at the University of Saskatchewan (Saskatoon, SK, Canada), a slightly different method was used for the cryoprotection of the sample. The entire crystallization trial setup was transported to the diffractometer in a styrofoam

box containing ice. Once at the diffractometer, the crystal was quickly mounted in a loop and placed under the nitrogen stream to cool followed by immediate data collection. Because the MPD concentration in the drop was relatively high, no extra cryoprotectant solution was used for these crystals.

2.2.4 Data Collection and Processing

X-ray data were collected at BioCARS beamline 14-ID-B of the APS at Argonne National Laboratory (Argonne, IL, USA) for the d[GA(5FU)(5FU)AA(5FU)C]-Co^{2+} crystals. Five data sets were collected in a nitrogen stream at 110 K by a Mar CCD 165 detector with 1˚ per image oscillation around the omega axis. For each data set, 180 images were collected at λ = 1.1271 Å. Data from the best diffraction set were indexed and merged using Mosflm and scaled using Scala from the CCP4 suite (CCP4, 1994). Data collection and processing parameters are given for both Laue groups 4/m and 4/mmm in Table 2.3. Additional screening of d[GA(5FU)(5FU)AA(5FU)C] crystals with both Co^{2+} and Zn^{2+} were done on the DX8 Proteum diffractometer at the Saskatchewan Structural Sciences Centre at the University of Saskatchewan (Saskatoon, SK, Canada). A diffraction image from the Co^{2+} sample was collected from a two minute exposure in a nitrogen stream at 110 K by a Proteum 4K CCD detector with a 1˚ oscillation around the omega axis. Diffraction images of the Zn^{2+} sample were collected from thirty second exposures with a 0.3 ˚ per image oscillation around the omega axis, also with the use of the Proteum 4K CCD detector.

X-ray data for the d[CG(5FU)G(5FU)GCACACG]-Zn^{2+} crystals were collected at BioCARS beamline 14-BM-C of the APS at Argonne National Laboratory (Argonne, IL, USA). The data set was collected in a nitrogen stream at 110 K by an ADSC Quantum-315 detector with 0.5˚ oscillation per image around the omega axis. A total of 720 images were collected at λ = 0.90 Å. Three hundred sixty images were indexed and merged using Mosflm and scaled using Scala from the CCP4 suite (CCP4, 1994). The

Table 2.3: Data collection and processing parameters.

d[GA(5FU)(5FU)AA(5FU)C]-Co^{2+}		
Laue group	4/m	4/mmm
Unit cell dimensions	a = 129.6 b = 129.6 c = 44.1 α = 90.0 β = 90.0 γ = 90.0	a = 129.6 b = 129.6 c = 44.1 α = 90.0 β = 90.0 γ = 90.0
Detector distance (mm)	160	160
X-ray wavelength (Å)	1.1271	1.1271
Total reflections collected	18195	16955
Unique reflections	4005	2149
Resolution range (outer shell) (Å)	58.3-4.7 (5.0-4.7)	45.8-4.7 (5.0-4.7)
Redundancy (outer shell)	4.5 (3.9)	7.9 (6.9)
R_{merge} (outer shell)	0.164 (0.460)	0.182 (0.446)
Completeness (outer shell) (%)	97.5 (97.5)	98.2 (98.2)
Mosaicity (˚)	1.6	1.6
I/σI (outer shell)	2.9 (1.6)	2.7 (1.7)

d[CG(5FU)G(5FU)GCACACG]-Zn^{2+}					
Laue group	$\bar{3}$	$\bar{3}$ m1	$\bar{3}$ 1m	6/m	6/mmm
Unit cell dimensions	a = 26.0 b = 26.0 c = 99.1 α = 90.0 β = 90.0 γ = 120.0	a = 26.0 b = 26.0 c = 99.1 α = 90.0 β = 90.0 γ = 120.0	a = 26.0 b = 26.0 c = 99.1 α = 90.0 β = 90.0 γ = 120.0	a = 26.0 b = 26.0 c = 99.1 α = 90.0 β = 90.0 γ = 120.0	a = 26.0 b = 26.0 c = 99.1 α = 90.0 β = 90.0 γ = 120.0
Detector distance (mm)	300	300	300	300	300
X-ray wavelength (Å)	0.9000	0.9000	0.9000	0.9000	0.9000
Total reflections collected	8700	8560	8560	8588	8625
Unique reflections	1769	990	990	910	611
Resolution range (Å)	50-2.8	25-2.8	22-2.8	21-2.8	50-2.8
Outer shell resolution (Å)	2.95-2.80	2.95-2.80	2.95-2.80	2.95-2.80	2.95-2.80
Redundancy (outer shell)	4.9 (5.1)	7.9 (8.7)	8.6 (9.3)	9.4 (9.8)	14.1 (16.0)
R_{merge} (outer shell)	0.04 (0.17)	0.04 (0.18)	0.04 (0.18)	0.04 (0.18)	0.04 (0.19)
Completeness (outer shell) (%)	95.9 (95.9)	96.3 (96.3)	96.2 (96.2)	96.5 (96.5)	96.8 (96.8)
Mosaicity (˚)	0.7	0.7	0.7	0.7	0.7
I/σI (outer shell)	9.7 (4.3)	9.3 (4.0)	11.4 (3.9)	10.0 (4.1)	11.4 (3.8)

data collection and processing parameters for several possible Laue groups are given in Table 2.3.

2.2.5 Solution and Refinement of Structure

Molecular replacement was the chosen method of analyzing the collected diffraction data because of the low resolution of the collected data, as well as the large availability of B-DNA models.

2.2.5.1 Starting Models

Due to the strong stacking reflections observed at approximately 3.2 Å in both the d[GA(5FU)(5FU)AA(5FU)C]-Co^{2+} and the d[CG(5FU)G(5FU)GCACACG]-Zn^{2+} diffraction patterns, it is apparent that these complexes are in a B-like conformation. Therefore, several B-DNA models with similar size and sequence were chosen from the Nucleic Acid Database (NDB). In particular, models were chosen with a resolution better than 1.5 Å. For each set of data at least five B-DNA models were used in combinations with each other, as well as one A-DNA and one Z-DNA model. The A and Z-DNA models were chosen as controls and were not expected to result in good solutions. The models chosen were adapted from the NDB entries BD0001, BD0018, BD0067, BDJB50, BDL084, AD0021 and ZDH030 (Kumar *et al.*, 1992; Hahn and Heinemann, 1993; Rozenberg *et al.*, 1998; Shui *et al.*, 1998; Minasov *et al.*, 1999; Egli *et al.*, 2001; Woods *et al.*, 2004). To prepare the starting models, the water molecules were removed as well as enough flanking base pairs in order to create a model with the same number of base pairs as the unknown structure. The models were prepared as both single strands and duplexes and were all rotated and translated to an equivalent position using Superpose from the CCP4 suite (CCP4, 1994). For both the d[GA(5FU)(5FU)AA(5FU)C]-Co^{2+} and the d[CG(5FU)G(5FU)GCACACG]-Zn^{2+} starting models, Phaser, from the CCP4 suite was used for both the rotational orientation and the translational orientation searches (McCoy *et al.*, 2005; Storoni *et al.*, 2005). Phaser uses the molecular replacement methods known as likelihood-enhanced fast rotation and likelihood-enhanced fast translation.

2.2.5.2 Refinement Procedures

2.2.5.2.1 d[GA(5FU)(5FU)AA(5FU)C]

Refinement procedures for orientated models were carried out using the program CNS version 1.1 (Brunger *et al.*, 1998) using nucleic acid specific parameter files (Parkinson *et al.*, 1996). The diffraction data were separated into a working set containing 90% of the reflections and a reference set containing 10% of the reflections. A model from one of the most plausible rotation/translation solutions was carefully rebuilt with the sequence d(GAUUAAUC) using Coot (Emsley and Cowtan, 2004) and regularized using the program Refmac5 from the CCP4 suite (CCP4, 1994). This same model was then rotated and translated over the coordinates from the other possible solutions *via* Superpose from the CCP4 suite (CCP4, 1994). The different solutions were then merged together creating asymmetric units with two duplexes. Each of the solutions were then refined separately as well as combined by rigid body refinement using CNS (Brunger *et al.*, 1998). Each rigid body refinement procedure consisted of two cycles of twenty steps of refinement. For all of the following refinement steps B-DNA restraints for base planarity, sugar pucker and Watson-Crick hydrogen bonding were imposed, as well as non-crystallographic restraints among the models with more than one duplex. All of the phosphate and C1' atoms were considered fixed atoms in order to prevent over refining the model. Because of the low resolution of the available data, without these restraints the models had a tendency to over-refine significantly. After rigid body refinement, 25 trials of simulated annealing using torsion angle dynamics were used to improve the model. The simulated annealing stages consisted of a slowcooling stage with a starting temperature of 2500 K and a drop in temperature of 50 K per cycle followed by 100 steps of minimization (Brunger *et al.*, 1998).

2.3 X-ray Absorption Spectroscopy

EXAFS experiments were carried out in the transmission mode for powder samples of Nickel (II) phthalocyanine (NiPC), Ni^{2+} M-DNA and Ni^{2+} B-DNA on the PNC-CAT at the APS at Argonne National Laboratory (Argonne, IL, USA) using beamline BM-20 by Dr. R. Sammynaiken and Dr. D. T. Jiang. Samples were ground and mounted on 3M Scotch Tape. The number of layers of tape and sample was adjusted until the absorption

at the metal edge was close to 1. Solid DNA samples were prepared by Mr. R. Skinner using the following procedure: Calf thymus DNA was dissolved in either 40 mM Tris-HCl pH 7.5 or 8.6, to a concentration of 1.125 mM in base pairs. It was sheared 5 times by passing through a 30-gauge needle on ice. For M-DNA, $NiCl_2$ was added to a final concentration of 33.75 mM at pH 8.6 and incubated for 2 hrs. During this time the DNA precipitated, it was collected by spooling onto a glass rod and dried *in vacuuo*. For B-DNA at pH 7.5 the DNA did not precipitate during addition of the metal ions and it was precipitated with 2 volumes of ethanol before drying.

2.3.1 Data Analysis

Three sets of raw data for the NiPC sample, eight sets for the Ni^{2+} M-DNA sample and five sets for Ni^{2+} B-DNA were collected. For each of the samples, NiPC, M-DNA and B-DNA, the data were averaged followed by background subtraction and normalization correction. After background subtraction, data were transformed to a function of photoelectron wave number, *k*. Since experimental artefacts and non-EXAFS processes often preclude the use of data at high-*k* and low-*k*, it was appropriate to use a subset of data for the Fourier transform. The *k* ranges selected are as large as possible yet still maintain fits of good statistical quality. The *k*-range chosen for both the M-DNA and the B-DNA data was 2.000 – 14.850 $Å^{-1}$, for the NiPC data a *k*-range of 2.000 – 14.350 $Å^{-1}$ was choosen. All three sets were *k*-weighted at k^1. These steps were performed using the program Athena version 0.8.041 (Ravel and Newville, 2005).

The next step in the data analysis was the creation of theoretical standards with which to analyze the experimental data. The coordinates used for the theoretical structure of M-DNA were determined by Les W. Tari (Aich *et al.*, 1999). The program Atoms version 3.0 (Ravel, 2001) was used to create an input file out of the coordinates from the theoretical structure (Table A1). The space group used to create the input file was P1, chosen due to its lack of symmetry. A space group with higher symmetry would cause the program to consider all symmetry related interactions which is unnecessary with a model of a double helix. In EXAFS analysis, the space group would be very important for a small molecule structure, but less so when it is only interhelical interactions that are of interest. Because the samples were most likely not 100% pure,

more input files were created to represent the nickel-N7 bound and the nickel-OP bound (Table A2). For data analysis of the B-DNA data, only the nickel-N7 and the nickel-OP input files were used. It is assumed that no M-DNA bounds will be present in the B-DNA sample. This is consistent with the experimental data regards to M-DNA to B-DNA conversion. The input file used in the analysis of the NiPC data was created using previously published information known about the structure of NiPC (Table A3) (Robertson and Woodward, 1937).

The program Artemis version 0.8.000 (Ravel and Newville, 2005) was used to fit the theoretical standards to the data *via* the EXAFS equation. Table A4 in Appendix A lists the constraints and mathematical expressions applied to the data in order to limit some of the variables involved in the fit. For example, the passive electron reduction factor, S_O^2, and the change in half path length, δR, were allowed to differ between each of the single scattering paths while σ^2 was limited to positive numbers. Finally, the structural parameters are refined and the data are ready for structural analysis.

3.0 RESULTS
3.1 d[GA(5FU)(5FU)AA(5FU)C]
3.1.1 Crystallization and Cryoprotection

Crystals of d[GA(5FU)(5FU)AA(5FU)C] grew reproducibly at 4 °C in the presence of Co^{2+} as described in Section 2.2.2.1. Initially, the Co^{2+} complex crystals grew as rectangular prisms after a period of one week having approximate maximum dimensions of 0.10 x 0.08 x 0.03 mm^3. These crystals diffracted only poorly, with the best resolution being around 7 Å. It was obvious that optimization was necessary to obtain more useful diffraction data. In order to improve the quality of these crystals many different additives and methods were tried. In particular, additives such as cobalt hexamine, spermidine, sodium chloride, magnesium chloride, potassium sulphate, sodium carbonate, polyethylene glycol and potassium chloride were included. Different buffers and concentrations of buffers were also manipulated along with the concentrations of all of the ingredients in the crystallization trials. Also, different temperatures, different materials for the slides and different crystallization methods such as microbatch and sitting drop vapour diffusion methods were examined.

The greatest improvement in the size and quality of the crystals came when the metal source was altered. Originally, cobalt chloride was used as a source of divalent cobalt. When cobalt perchlorate was investigated, the improvement in crystal size was obvious immediately. The crystal dimensions were then at an approximate maximum of 0.30 X 0.15 X 0.5 mm^3 with diffraction to about 3 Å, a significant improvement from the original results (Figure 3.1). Similar crystallization conditions were found to work also for the Zn^{2+} and Ni^{2+} complexes. In particular, the Zn^{2+} complex crystals grew even larger than the Co^{2+} crystals; however, the crystals appeared lower in quality, often growing more then one crystal from a single nucleation site (Figure 3.2).

The d[GA(5FU)(5FU)AA(5FU)C]-Co^{2+} crystals were mounted in a cryoprotectant solution containing 50% MPD and 10% glycerol using the first method described in Section 2.2.3. The entire procedure took less than one minute. The d[GA(5FU)(5FU)-

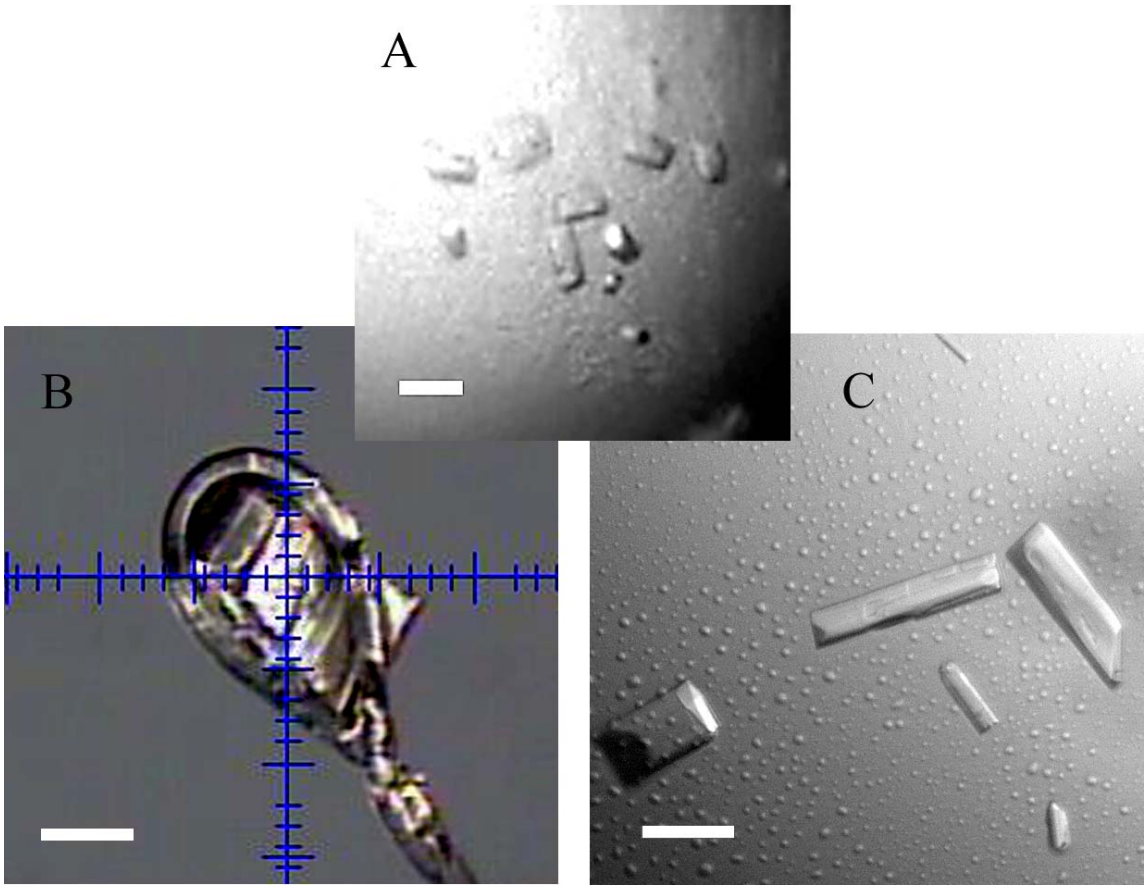

Figure 3.1: Crystals of d[GA(5FU)(5FU)AA(5FU)C] grown with Co^{2+} at pH 8.5. The photographs include A) the initial crystals before optimization and crystals after optimization both B) mounted after flash cooling in liquid nitrogen and C) in the drop of mother liquor. The scale bar represents a distance of approximately 0.1 mm.

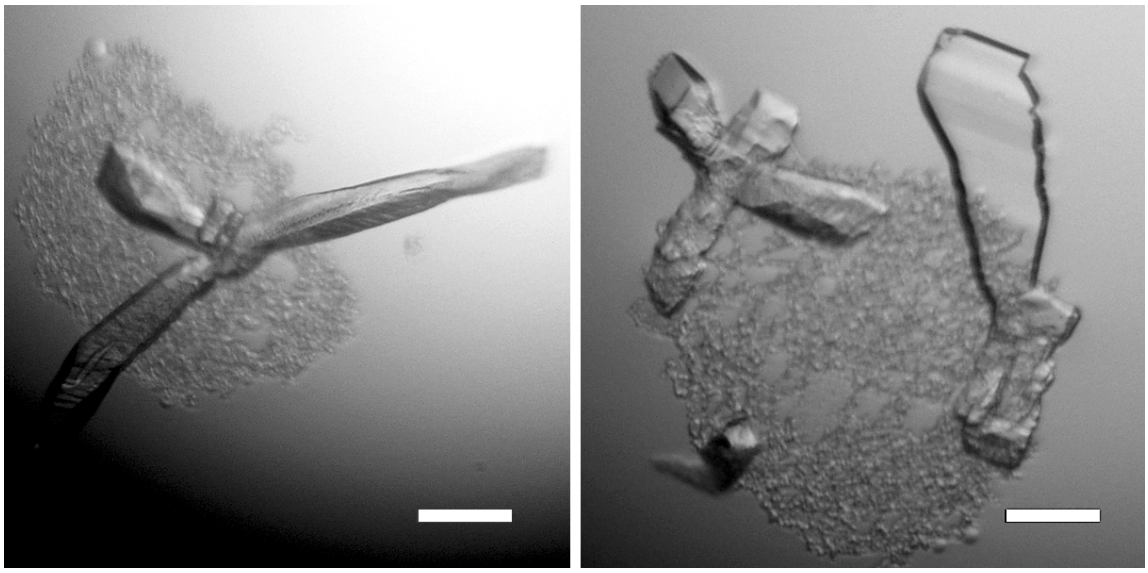

Figure 3.2: Crystals of d[GA(5FU)(5FU)AA(5FU)C] grown with Zn^{2+} at pH 8.3. The scale bar represents a distance of approximately 0.1 mm.

AA(5FU)C]-Zn^{2+} crystal was diffracted on the DX8 Proteum diffractometer at the Saskatchewan Structural Sciences Centre at the University of Saskatchewan (Saskatoon, SK, Canada) often undergoing the second cryoprotection method as described in Section 2.2.3.

3.1.2 Diffraction

The crystal of d[GA(5FU)(5FU)AA(5FU)C] grown with Co^{2+} diffracted to approximately 3.2 Å at BioCARS beamline 14-ID-B of the APS at Argonne National Laboratory (Argonne, IL, USA) (Figure 3.3). Using the CCP4 program HKLVIEW (CCP4, 1994), simulated precession images were calculated from the diffraction data in 4/m. The *hk1* and *hk2* planes were used in order to help distinguish between 4/m and 4/mmm (Figure 3.4). By looking at the precession images, one can see the presence of mirror planes on each axes, as well as on the diagonal between the axes. This additional symmetry indicates 4/mmm diffraction symmetry. Therefore, the possible space groups are P422, $P4_222$, $P4_122$, $P4_322$, $P42_12$, $P4_22_12$, $P4_12_12$and $P4_32_12$. These space groups can be distinguished by looking at the systematic absences of reflections along *00l* and/or *h00*. Table 3.1 lists the equivalent indices for 4/m and 4/mmm. The [GA(5FU)(5FU)AA(5FU)C] diffraction data appears to have no systematic absences, suggesting the space group P422. However, because the quality of the data is so poor and the data are scarce due to the low resolution of diffraction, the systematic absences are very unreliable and translation function searches were performed in all of the possible space groups. The merging R-value (a measure of agreement among multiple measurements of the same reflections) in P422 was 0.182 and 468 reflections were rejected due to cutting the resolution off at 4.7 Å. The data were processed with the following cell dimensions; a = b = 130.26 Å, c = 44.28 Å, $\alpha = \beta = \gamma = 90.00$ º.

The crystal of d[GA(5FU)(5FU)AA(5FU)C] grown in the presence of Zn^{2+} diffracted to approximately 2.7 Å on the DX8 Proteum Diffractometer in the Saskatchewan Structural Sciences Centre (Figure 3.5). The best initial diffraction yet seen was collected from this crystal. Yet, moments after the initial diffraction was collected the diffraction disappeared. Some unexplained event had occurred to this crystal once it was mounted and exposed to the X-rays.

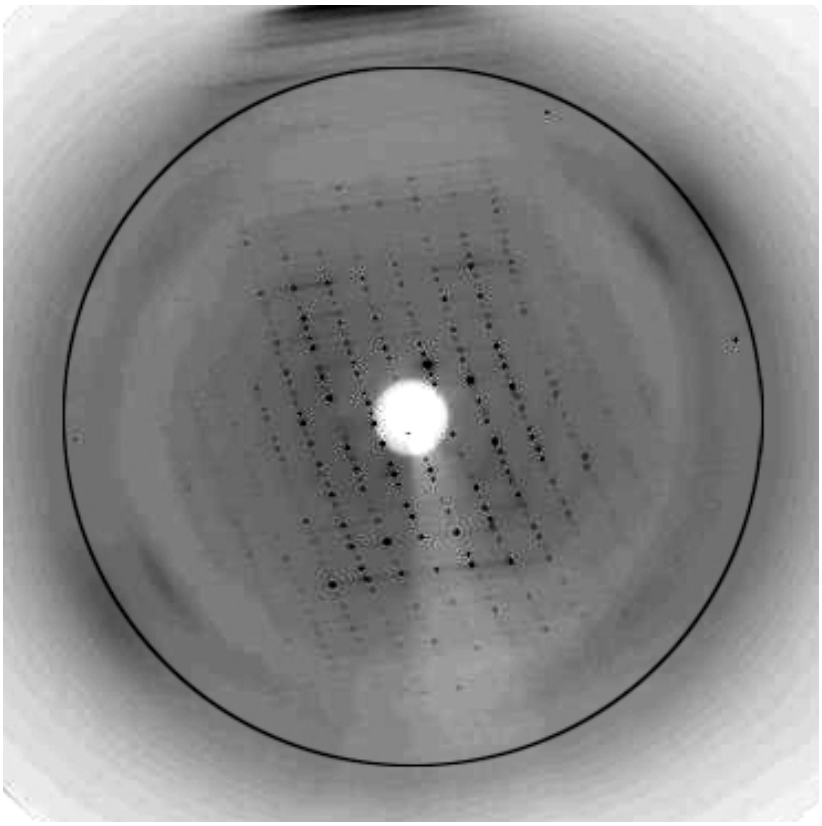

Figure 3.3: Diffraction image for a d[GA(5FU)(5FU)AA(5FU)C] crystal grown in the presence of Co^{2+} at pH 8.5. The diffraction image was collected at BioCARS beamline 14-ID-B of the APS at Argonne National Laboratory (Argonne, IL, USA). The circle shows an approximate resolution boundary of 3.0 Å.

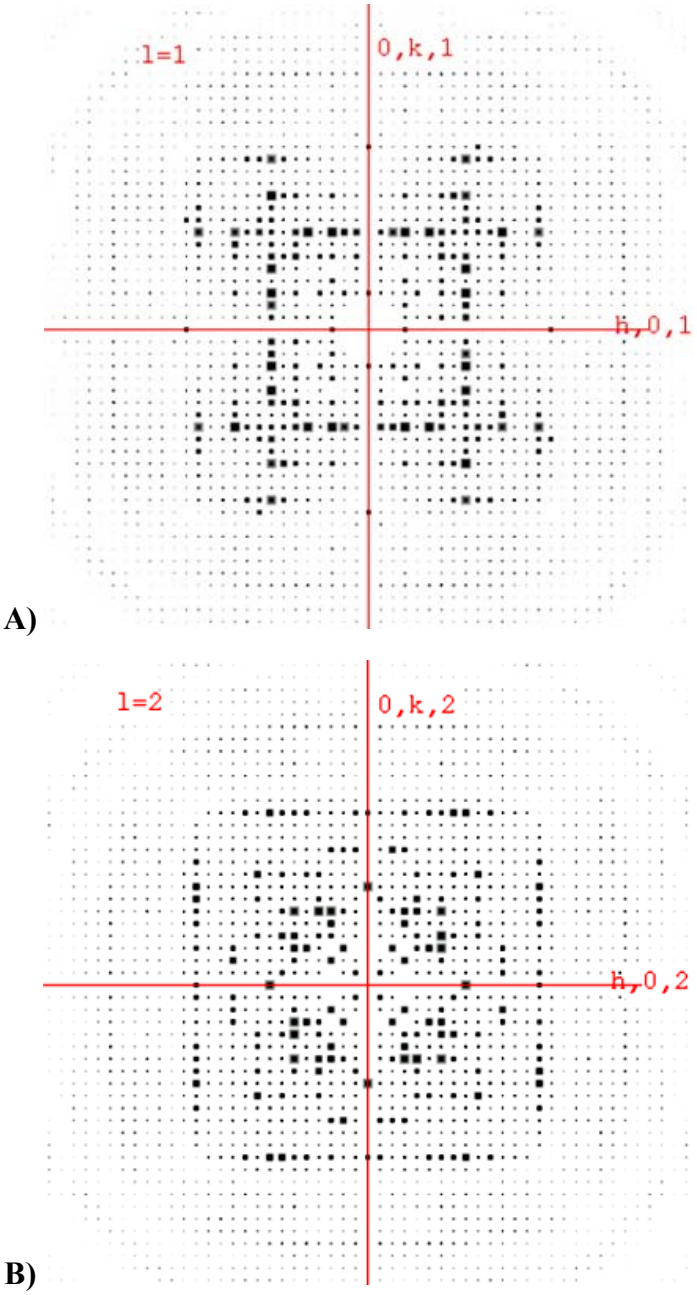

Figure 3.4: Simulated precession images of the A) *hk1* and B) *hk2* planes of diffraction data for d[GA(5FU)(5FU)AA(5FU)C] grown in the presence of Co^{2+} at pH 8.5. These data were processed assuming 4/m symmetry.

Table 3.1: Equivalent indices for the Laue groups 4/m and 4/mmm from the International Tables for X-ray Crystallography (The International Tables of Crystallography Vol I, 1968).

Laue Group	Equivalent Indices
4/m	$(hkl) = (-h-k-l) = (hk-l) \neq (-hkl);\ (-hkl) = (h-kl) = (khl)$
4/mmm	$(hkl) = (-h-k-l) = (-hkl) = (h-kl) = (hk-l) = (khl)$

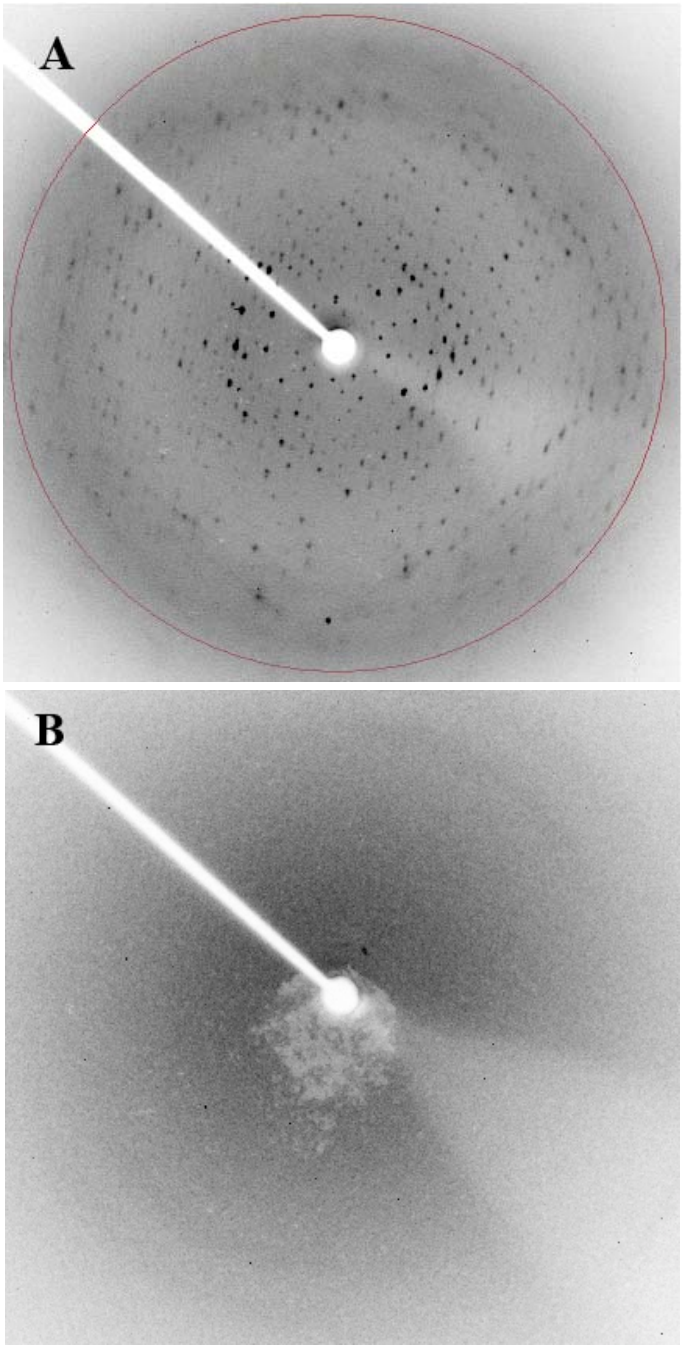

Figure 3.5: Diffraction images collected from a d[GA(5FU)(5FU)AA(5FU)C] crystal grown with a source of divalent zinc at pH 8.0. The diffraction images were collected on the DX8 Proteum Diffractometer in the Saskatchewan Structural Sciences Centre. The crystal diffracted to approximately A) 3.0 Å, yet, B) moments later, diffraction disappeared after an unexplained event. The red circle represents an approximate resolution boundary of 3.0 Å.

3.1.3 Rotation and Translation

In order to determine the correct rotational orientation of the search model in the cell of the unknown structure, it was necessary to begin with a rotation search using Phaser from the CCP4 suite (Storoni *et al.*, 2005). Brute force likelihood rotation was performed in the space group 4/mmm (the symmetry of the Patterson function for any of the possible space groups) with the seven different models described in section 2.2.5.1 using both single stranded and double stranded versions as well as several different resolution cut off limits. The best solutions came from using the double stranded models with three different resolution cut off limits. Table 3.2 provides a list of the rotation search results with regards to Euler angles. The use of Euler angles is common in molecular replacement programs as a way to describe the orientation of the search model. In this convention, the coordinate system is rotated by an angle α around the original *z* axis, then by an angle *β* around a new *y* axis, and then by an angle *γ* around the new *z* axis. For a rotation function, the correct solution may be in the list with a Z-score (number of standard deviations above the mean value) under 4, and will not be found until a translation function is performed and picks out the correct solution. Of the results reported in Table 3.2 the Z-scores are too low to indicate an obvious solution, however, some solutions were consistent between the different search models and the separation between the top z-scores with that of the next solution appear hopeful. It is also important to note that no likely solutions were found using the A and Z-DNA models. This provided good controls as no real solutions were expected if the unknown structure is significantly close to B-DNA. The highest Z-scores with the clearest separation between solutions resulted with the resolution cut off of 5.0 Å – 15.0 Å. Therefore, brute force likelihood translation was performed in Phaser from the CCP4 suite using the bottom solution set from Table 3.2 (McCoy *et al.*, 2005).

Translation was performed in the space group P422, as well as $P4_22_12$, $P4_32_12$, $P4_122$, $P4_322$, $P4_222$, $P4_12_12$ and $P42_12$. The frequency of occurrences of these space groups in the NDB can be used as a crude guide in identifying likely translation function solutions. In general, space groups with no translational (screw axis) symmetry are very rare for crystalline nucleic acids or proteins (e.g. P422), whereas space groups containing one or

Table 3.2: Rotation search results for the d[GA(5FU)(5FU)AA(5FU)C] search models performed in the space group P422 using Phaser from the CCP4 suite (Storoni *et al.*, 2005).

Ensemble	α	β	γ	Z-score
Resolution cut off: 5.0 Å – 8.0 Å				
bd0001	71.956	60.056	254.850	2.89
	72.484	80.019	247.094	2.85
bd0018	52.776	34.003	274.177	3.68
	41.434	23.822	295.704	3.28
	70.015	64.547	249.031	3.09
bd0067	79.559	61.184	248.857	3.52
	79.199	80.395	243.261	3.36
bdjb50	85.746	50.188	251.733	3.43
bdl084	79.472	80.128	246.917	3.79
	79.448	60.383	253.109	3.72
	72.096	30.765	261.161	3.34
Resolution cut off: 5.0 Å – 10.0 Å				
bd0001	71.956	60.056	254.850	3.49
	72.484	80.019	247.094	3.22
bd0018	52.776	34.003	274.177	3.40
	41.434	23.822	295.704	3.14
	70.015	64.547	249.031	3.03
bd0067	79.199	80.395	243.261	3.50
	79.559	61.184	248.857	3.49
	68.326	32.369	270.054	2.95
bdl084	79.472	80.128	246.917	3.70
	72.096	30.765	261.161	3.10
Resolution cut off: 5.0 Å – 15.0 Å				
bd0001	71.956	60.056	254.850	3.53
	72.484	80.019	247.094	3.32
bd0018	52.776	34.003	274.177	3.51
	41.434	23.822	295.704	3.27
	70.015	64.547	249.031	3.10
bd0067	79.199	80.395	243.261	3.53
	79.559	61.184	248.857	3.46
	68.659	51.579	259.662	2.90
bdl084	79.472	80.128	246.917	3.75

more screw axes occur with high frequency (e.g. $P4_32_12$, $P4_12_12$, etc.). This is essentially, because the introduction of a screw axis allows many more potential ways for molecules to pack in a crystal. The translation search results for the space groups consistent with 4/mmm diffraction symmetry are listed in Table 3.3. Generally, correct solutions to the translation function should have a Z-score over 6 and be well separated from the rest of the solutions in terms of signal to noise. There were several possible solutions based on this criterion. However, after analyzing the solutions graphically for physically reasonable molecular packing within the unit cell using either TURBO-FRODO or Coot (Roussel and Cambillau, 1992; Emsley and Cowtan, 2004), many potential solutions were ruled out. The solution at this point is one of the two bold face solutions in Table 3.3 or a combination of both solutions (Figure 3.6). The correct space group is likely $P4_22_12$ and the asymmetric unit most likely contains two duplexes. In order to check for the presence of a third duplex in the asymmetric unit, Phaser was run again with the two known solutions input into the program. Some possible solutions resulted, however, due to bad packing, it was concluded that there is not a third duplex in the asymmetric unit. However, based on packing, a fourth solution was found. In the fourth solution, solution B from figure 3.6 was extended by one duplex (Figure 3.6D). This solution was found using Phaser with a relatively low Z-score. For each solution, the ensemble was rotated and translated using the program Superpose from the CCP4 suite (CCP4, 1994) followed by careful rebuilding with the correct sequence, d(GAUUAAUC) and further refinement using the CNS software (Brunger *et al.*, 1998).

3.1.4 Refinement

The refinement process took place as described in sections 2.2.5.2.1 for each of the four final solutions illustrated in Figure 3.6. Refinement progress statistics are given in Table 3.4. Unfortunately, none of the models refined very successfully. In fact, none of the R-factors came below 50% indicating none of the potential solutions can be verified by refinement of the atomic coordinates. With such weak diffraction data, this result is not surprising. If one looks at the root mean square deviation (rmsd) values of the bond lengths and angles, all of the solutions A through D seem plausible (Table 3.4). Rmsd represent how well the bond distances and angles compare to that of what is

Table 3.3: Translation search results for the d[GA(5FU)(5FU)AA(5FU)C] search models performed in Phaser from the CCP4 suite (McCoy *et al.*, 2005). The two bold face solutions are the chosen final solutions. *Log-likelihood gain (LLG).

Ensemble	α	β	γ	transX	transY	transZ	LLG*	Z-score
Space group: **$P4_12_12$**								
bd0001	72.0	60.1	254.8	0.033	0.188	0.222	39.9011	4.13
	72.5	80.0	247.1	0.029	0.703	0.074	39.1088	4.56
				0.060	0.783	0.445	38.0378	4.34
bd0018	52.8	34.0	274.2	0.083	0.930	0.074	46.2181	4.79
				0.365	0.202	0.074	44.0906	4.39
	41.4	23.8	295.7	0.233	0.857	0.297	44.8553	5.44
	70.0	64.5	249.0	0.411	0.389	0.074	38.7394	4.70
bd0067	79.2	80.4	243.3	0.461	0.774	0.389	41.3045	4.70
	79.6	61.2	248.9	0.110	0.046	0.389	45.0498	5.45
				0.141	0.006	0.167	41.7976	4.81
	68.7	51.6	259.7	0.152	0.948	0.056	37.3651	5.23
	68.3	32.4	270.1	0.222	0.841	0.130	38.9539	5.35
bdl084	79.5	80.1	246.9	0.164	0.175	0.074	46.7657	5.16
Space group: **P422**								
bd0001	72.0	60.1	254.8	0.307	0.306	0.093	68.7650	6.48
				0.488	0.313	0.093	63.7254	5.81
	72.5	80.0	247.1	0.469	0.462	0.222	57.8766	5.15
bd0018	52.8	34.0	274.2	0.311	0.175	0.222	65.5793	5.48
	41.4	23.8	295.7	0.369	0.623	0.445	57.4196	5.03
	70.0	64.5	249.0	0.276	0.355	0.148	59.5807	5.00
bd0067	79.2	80.4	243.3	0.264	0.460	0.204	64.7970	5.74
				0.272	0.460	0.056	58.1313	4.91
	79.6	61.2	248.9	0.477	0.320	0.093	68.8514	6.63
				0.477	0.320	0.426	62.1753	5.75
	68.7	51.6	259.7	0.369	0.213	0.426	56.4673	6.27
				0.396	0.260	0.389	55.5947	6.14
				0.446	0.213	0.463	54.7746	6.02
	68.3	32.4	270.1	0.264	0.148	0.185	51.9042	5.17
bdl084	79.5	80.1	246.9	0.473	0.442	0.037	61.6542	5.29
Space group: **$P42_12$**								
bd0001	72.0	60.1	254.8	0.492	0.315	0.074	51.5344	5.58
				0.492	0.315	0.222	51.0699	5.50
	72.5	80.0	247.1	0.295	0.469	0.111	49.5448	5.46
bd0018	52.8	34.0	274.2	0.311	0.175	0.445	52.5422	5.36
	41.4	23.8	295.7	0.369	0.119	0.019	50.6552	5.59
				0.369	0.119	0.463	47.9622	5.11
	70.0	64.5	249.0	0.349	0.349	0.371	48.5327	5.39
				0.345	0.346	0.019	46.4349	5.00
bd0067	79.2	80.4	243.3	0.268	0.449	0.408	49.4993	5.28
				0.415	0.467	0.426	48.7563	5.15
	79.6	61.2	248.9	0.477	0.823	0.037	55.2170	6.53
				0.477	0.823	0.297	54.3304	6.38
				0.477	0.823	0.111	52.2889	6.01
				0.477	0.823	0.222	51.4092	5.85
	68.7	51.6	259.7	0.500	0.262	0.111	40.5725	5.13

	68.3	32.4	270.1	0.264	0.647	0.204	39.3168	4.72
bdl084	79.5	80.1	246.9	0.407	0.943	0.334	51.1617	5.10

Space group: **$P4_122$**

bd0001	72.0	60.1	254.8	0.137	0.502	0.037	50.8631	4.79
	72.5	80.0	247.1	0.241	0.460	0.130	52.4084	5.84
				0.226	0.620	0.315	49.1056	5.29
bd0018	52.8	34.0	274.2	0.291	0.306	0.130	57.4732	5.23
	41.4	23.8	295.7	0.106	0.462	0.371	58.5161	5.65
				0.064	0.447	0.352	58.2145	5.61
	70.0	64.5	249.0	0.044	0.984	0.074	55.4856	5.12
bd0067	79.2	80.4	243.3	0.465	0.268	0.111	51.3796	4.54
	79.6	61.2	248.9	0.245	0.493	0.019	54.4396	5.21
	68.7	51.6	259.7	0.334	0.268	0.259	48.1014	5.07
	68.3	32.4	270.1	0.399	0.173	0.426	43.7975	4.88
bdl084	79.5	80.1	246.9	0.233	0.603	0.334	54.7511	4.88

Space group: **$P4_322$**

bd0001	72.0	60.1	254.8	0.434	0.335	0.148	47.3497	4.40
	72.5	80.0	247.1	0.033	0.750	0.111	44.3928	4.48
bd0018	52.8	34.0	274.2	0.326	0.179	0.130	54.0633	4.37
	41.4	23.8	295.7	0.102	0.790	0.371	55.0962	5.36
	70.0	64.5	249.0	0.473	0.442	0.074	56.8025	5.18
bd0067	79.2	80.4	243.3	0.006	0.079	0.130	52.6810	4.90
	79.6	61.2	248.9	0.245	0.186	0.019	57.5319	5.74
				0.496	0.433	0.019	54.8414	5.35
	68.7	51.6	259.7	0.268	0.141	0.445	47.3853	5.09
	68.3	32.4	270.1	0.423	0.173	0.093	44.4968	4.80
bdl084	79.5	80.1	246.9	0.029	0.075	0.148	52.4985	4.61

Space group: **$P4_222$**

bd0001	72.0	60.1	254.8	0.492	0.315	0.482	59.7412	5.39
	72.5	80.0	247.1	0.241	0.442	0.111	56.7857	5.90
				0.033	0.950	0.111	53.0271	5.33
bd0018	52.8	34.0	274.2	0.314	0.676	0.074	61.0004	5.39
	41.4	23.8	295.7	0.369	0.623	0.445	56.6989	5.42
	70.0	64.5	249.0	0.276	0.369	0.148	57.0810	5.08
bd0067	79.2	80.4	243.3	0.199	0.961	0.204	57.6555	5.77
				0.268	0.462	0.037	56.8167	5.65
				0.264	0.309	0.185	54.6271	5.33
	79.6	61.2	248.9	0.477	0.828	0.093	56.4715	5.43
	68.7	51.6	259.7	0.434	0.268	0.037	56.2500	5.90
				0.357	0.242	0.000	50.4468	5.11
	68.3	32.4	270.1	0.411	0.135	0.185	53.5901	5.90
				0.446	0.141	0.185	53.5273	5.89
				0.480	0.162	0.185	50.0550	5.38
bdl084	79.5	80.1	246.9	0.473	0.442	0.037	60.6024	5.83
				0.295	0.447	0.056	60.0694	5.75

Space group: **$P4_22_12$**

bd0001	72.0	60.1	254.8	0.423	0.320	0.056	51.8093	5.79
				0.307	0.306	0.130	51.5223	5.74
				0.492	0.315	0.074	50.1488	5.49
	72.5	80.0	247.1	0.492	0.315	0.074	50.1488	5.49

bd0018	52.8	34.0	274.2	0.311	0.175	0.259	51.2480	5.37
	41.4	**23.8**	**295.7**	0.369	0.119	0.019	51.5230	6.66
				0.369	**0.119**	**0.315**	**49.9057**	**6.33**
				0.369	0.119	0.463	49.3480	6.21
	70.0	64.5	249.0	0.345	0.346	0.019	47.4068	6.12
				0.276	0.360	0.167	45.9250	5.79
bd0067	79.2	80.4	243.3	0.268	0.453	0.056	48.7186	5.77
				0.415	0.467	0.426	47.7635	5.59
	79.6	61.2	248.9	0.226	0.322	0.037	50.4083	5.93
				0.295	0.322	0.445	49.1871	5.71
	68.7	51.6	259.7	0.434	0.268	0.297	40.7940	5.94
				0.249	0.255	0.408	39.7885	5.72
	68.3	32.4	270.1	0.372	0.153	0.056	40.8601	5.49
bdl084	**79.5**	**80.1**	**246.9**	**0.442**	**0.456**	**0.445**	**51.6668**	**5.94**
				0.291	0.440	0.093	51.2228	5.86
Space group: **$P4_32_12$**								
bd0001	72.0	60.1	254.8	0.326	0.790	0.259	41.8604	4.43
	72.5	80.0	247.1	0.179	0.202	0.297	42.3577	5.34
bd0018	52.8	34.0	274.2	0.365	0.202	0.074	46.2802	5.16
	70.0	64.5	249.0	0.276	0.373	0.167	40.3909	4.93
bd0067	79.2	80.4	243.3	0.064	0.175	0.259	42.0966	4.65
	79.6	61.2	248.9	0.469	0.039	0.130	40.8583	4.64
	68.7	51.6	259.7	0.496	0.981	0.056	37.5025	5.21
	68.3	32.4	270.1	0.118	0.268	0.000	32.6274	3.99
bdl084	79.5	80.1	246.9	0.461	0.173	0.315	48.6008	5.34

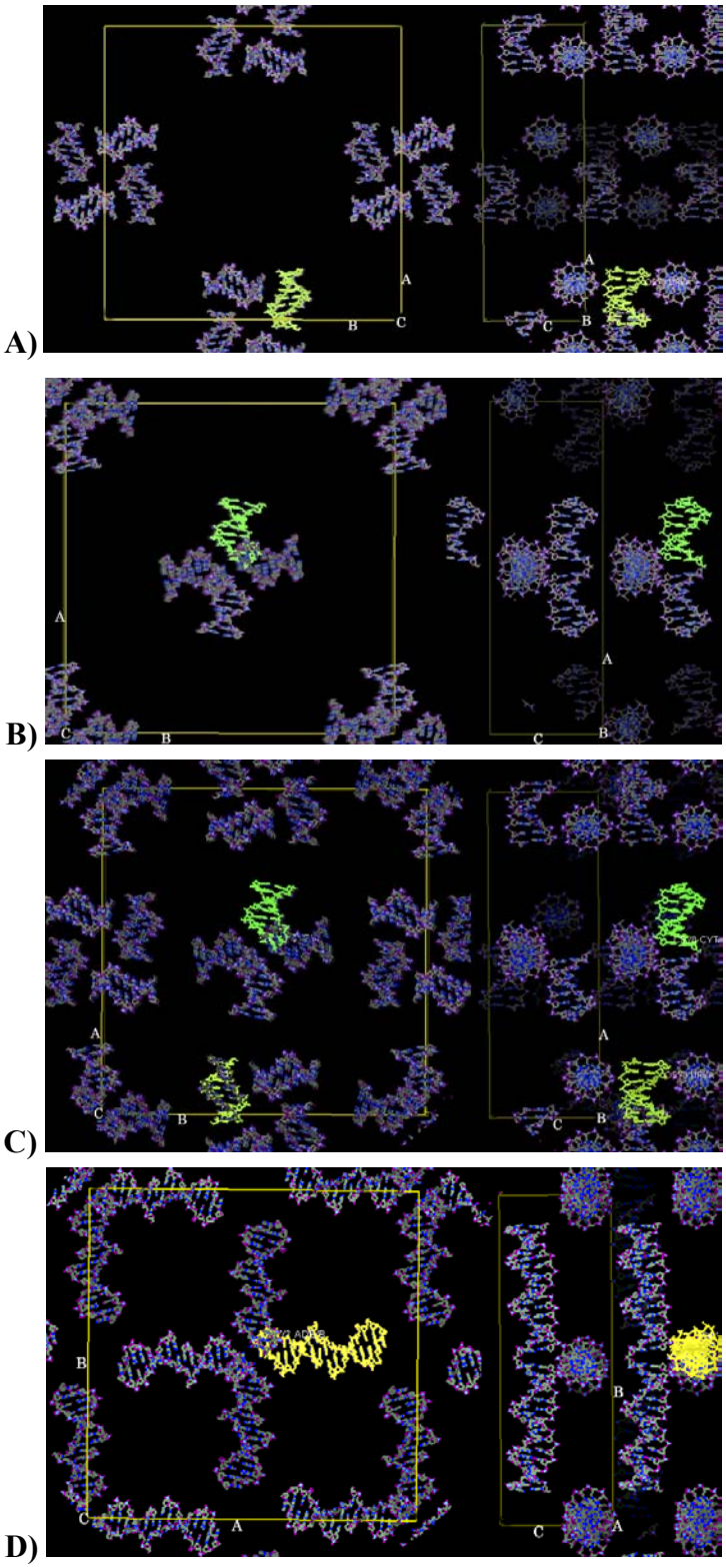

Figure 3.6: Rotation and translation solutions A) found with ensemble bd0018 and B) ensemble bdl084, a C) combination of the two solutions and D) solution B extended by one duplex. The yellow or green models represent the rotation/translation solutions. The blue models represent the symmetry mates of the solutions in $P4_22_12$.

Table 3.4: Crystallographic and refinement parameters

Parameter	Solution A	B	C	D
Crystallographic data				
Space group	$P4_22_12$	$P4_22_12$	$P4_22_12$	$P4_22_12$
Cell constants				
a (Å)	129.59	129.59	129.59	129.59
b (Å)	129.59	129.59	129.59	129.59
c (Å)	44.06	44.06	44.06	44.06
α (º)	90.00	90.00	90.00	90.00
β (º)	90.00	90.00	90.00	90.00
γ (º)	90.00	90.00	90.00	90.00
Refinement progress (R_{work}(R_{free}))				
Rigid body	0.650(0.660)	0.638(0.653)	0.648(0.672)	0.621(0.645)
5 x Anneal	0.568(0.617)	0.569(0.622)	0.562(0.600)	0.540(0.612)
10 x Anneal	0.570(0.601)	0.556(0.578)	0.565(0.599)	0.545(0.612)
15 x Anneal	0.573(0.611)	0.533(0.564)	0.562(0.598)	0.543(0.614)
20 x Anneal	0.572(0.614)	0.556(0.608)	0.559(0.605)	0.544(0.608)
25 x Anneal	0.569(0.613)	0.543(0.569)	0.558(0.605)	0.543(0.587)
Refinement results				
Resolution range (Å)	50.0-4.7	50.0-4.7	50.0-4.7	50.0-4.7
Reflection in workset	1943	1943	1486	1943
Reflections in test set	192	192	649	192
Rmsd in bond lengths (Å)	0.0096	0.0032	0.0080	0.0069
Rmsd in bond angles (°)	2.201	1.552	1.946	1.965

currently known about refined molecular structures. The lower the value, the more realistic the solution, although, due to the very large R-factors, it is difficult to assume any of these solutions are a realistic representation of the data presented so far. Solutions B and D do appear to be a little better than A and C with regards to the R-factor, although the difference is not significant. In fact, an R-factor greater than 0.55 is generally accepted to represent a random arrangement of atoms in the unit cell. Lack of even medium resolution data exacerbates the problem significantly. It will be necessary to collect higher resolution data before it will be possible to solve this problem more completely. For now, we can be confident that the solutions represented here are a reasonable representation of the likely DNA packing within a $P4_22_12$ unit cell.

3.2 d[CG(5FU)G(5FU)GCACACG]

3.2.1 Crystallization and Cryoprotection

Crystals of the d[CG(5FU)G(5FU)GCACACG]-Zn^{2+} complex grew reproducibly at 4 °C using the method described in Section 2.2.2.2. After the crystallization conditions were optimized, the crystals grew as hexagonal prisms with approximate dimensions of 0.15 x 0.1 x 0.1 mm^3 (Figure 3.7). Using the method described in Section 2.2.3, the crystals were them mounted into a cryoprotectant solution containing 60% MPD.

3.2.2 Diffraction

The crystal of d[CG(5FU)G(5FU)GCACACG] grown with Zn^{2+} diffracted to approximately 2.8 Å at BioCARS beamline 14-IM-C of the APS at Argonne National Laboratory (Argonne, IL, USA). The diffraction data were processed in several different Laue groups. The processing statistics for $\bar{3}$m1, $\bar{3}$1m and 6/mmm were very similar (Table 2.3), suggesting that the higher symmetry is correct or the possibility of merohedral twinning. Simulated precession images were calculated from the diffraction data processed in Laue group $\bar{3}$ using the CCP4 program HKLVIEW (CCP4, 1994). The *hk4*, *hk7* and *hhl* planes were used to aid in the identification of a possible Laue group (Figure 3.8) as well as the intensities of the reflections at the equivalent indices (Table 3.5). The pseudo-precession image *hk4* and *hk7* shows six-fold symmetry, implicating the correct diffraction symmetry to be 6/m or 6/mmm and the image *hhl*

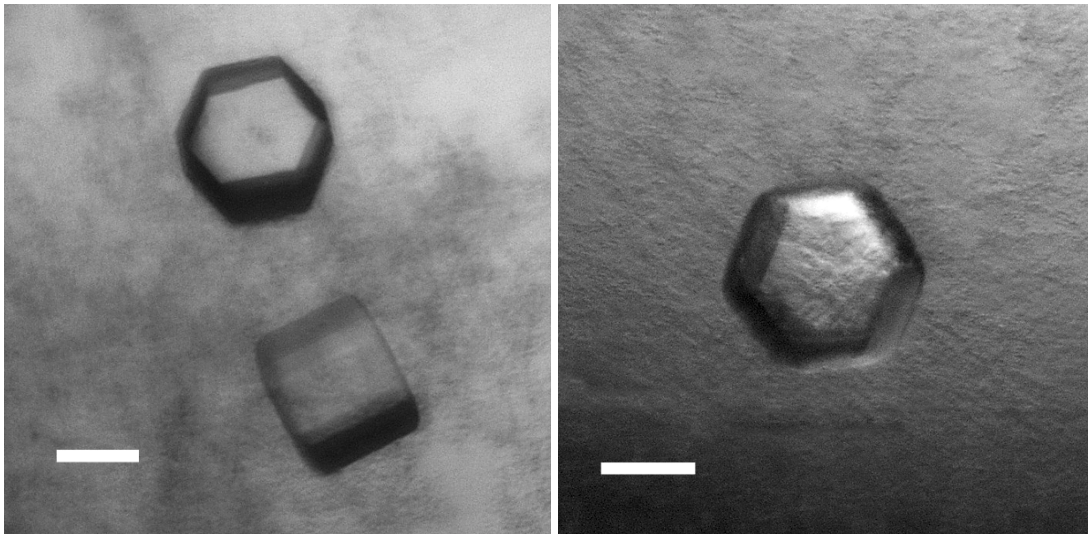

Figure 3.7: Crystals of d[CG(5FU)G(5FU)GCACACG] grown with Zn^{2+} at pH 7.75. The scale bar represents a distance of approximately 0.1 mm.

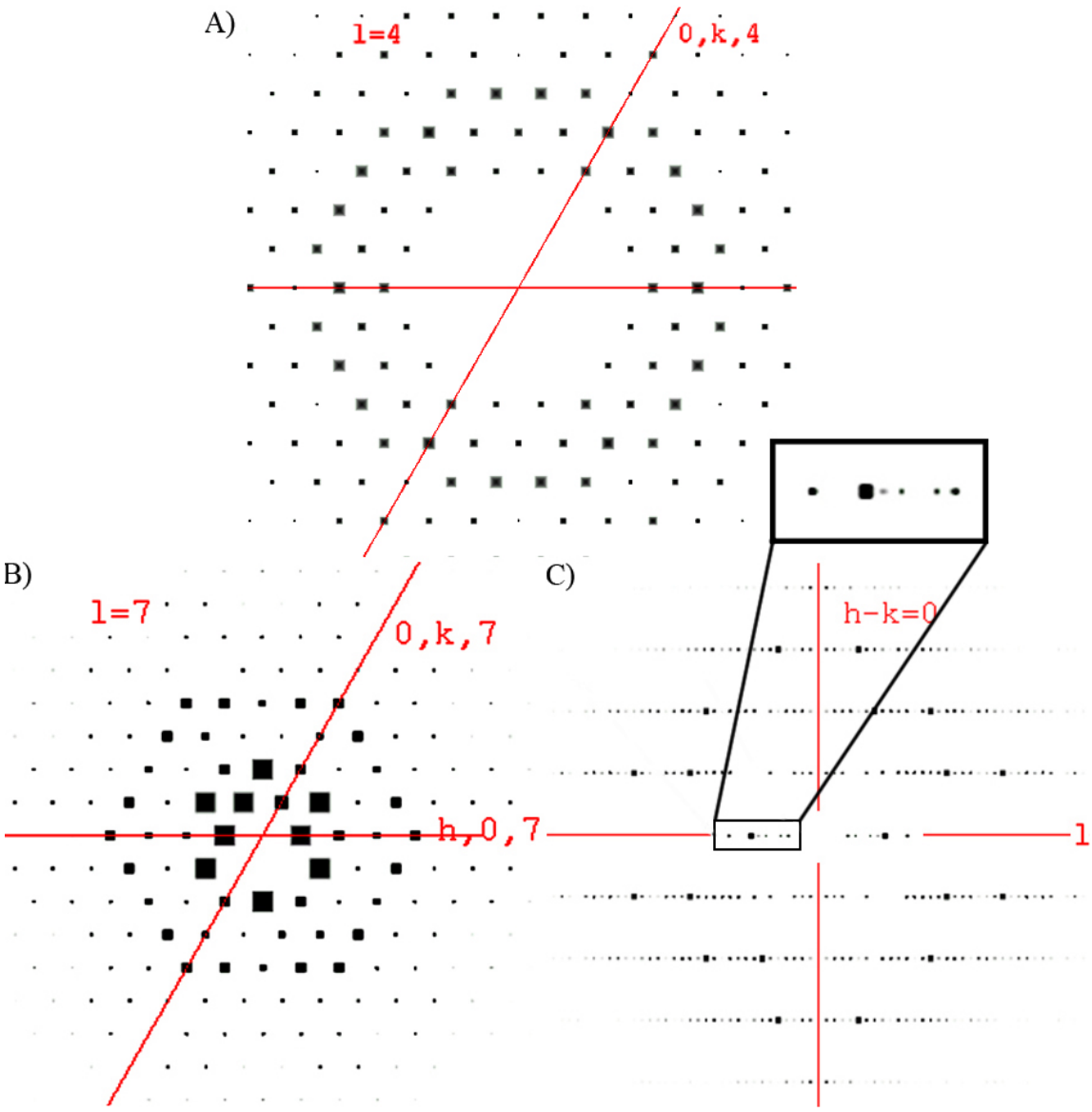

Figure 3.8: Simulated precession images of d[CG(5FU)G(5FU)GCACACG]-Zn^{2+} diffraction data processed in $\bar{3}$. These images were calculated using the CCP4 program, HKLVIEW (CCP4, 1994). A) *hk4* B) *hk7* and C) *hhl*. Note, in the precession image C) *hhl*, the reflections at 0,0,l have been enlarged respectively, in order to emphasize that there are no obvious systematic absences in the intensities of *00l* reflections.

Table 3.5: Equivalent indices for the Laue groups 1, 6/m, 6/mmm, $\bar{3}$, $\bar{3}$1m and $\bar{3}$m1 from the International Tables for X-ray Crystallography (The International Union of Crystallography, 1968).

Laue Group	Equivalent Indices
1	(*hkl*) = (-*h*-*k*-*l*) ≠ (-*hkl*) ≠ (*h*-*kl*) ≠ (*hk*-*l*)
6/m	(*hkl*) = (-*h*-*k*-*l*) = (*hk*-*l*) = (*kil*) ≠ (-*hkl*) ≠ (*khl*); (-*hkl*) = (*h*-*kl*)
6/mmm	(*hkl*) = (-*h*-*k*-*l*) = (*hk*-*l*) = (*kil*) = (*khl*) ≠ (-*hkl*); (*hkl*) = (*h*-*kl*)
-3	(*hkl*) = (-*h*-*k*-*l*) = (*hk*-*l*) = (*kil*) ≠ (-*hkl*) ≠ (*h*-*kl*) ≠ (*hk*-*l*) ≠ (*khl*)
-31m	(*hkl*) = (-*h*-*k*-*l*) = (*kil*) = (*khl*) ≠ (-*hkl*) ≠ (*h*-*kl*) ≠ (*hk*-*l*)
-3m1	(*hkl*) = (-*h*-*k*-*l*) = (*kil*) ≠ (-*hkl*) ≠ (*h*-*kl*) ≠ (*hk*-*l*) = (*khl*)

is useful to look for the presence of a mirror plane along *hhl* indicating the presence of 6/mmm symmetry. Therefore, several possible space group may be possible, namely, P622, $P6_322$, $P6_222$, $P6_422$, $P6_122$ and $P6_522$. The data was processed in P622 with the following unit cell dimensions a = 25.97 Å, b = 25.97 Å, c = 99.38 Å, $\alpha = 90$ º, $\beta = 90$ º, $\gamma = 120$ º. The R_{merge} was 0.041 for the resolution range 49.39 Å – 2.80 Å and the R_{merge} for the outer shell of 2.95 Å – 2.80 Å was 0.187.

3.2.3 Rotation and Translation

Phaser from the CCP4 suite was used in order to determine the correct rotational orientation of the search model in the cell of the unknown structure (Storoni *et al.*, 2005). Brute force likelihood rotation was performed in the Patterson space group 6/mmm with the seven different models described in section 2.2.5.1. The models were each separated into single strands labelled either A or B and different resolution cut off limits were used in several different runs of the program. The best solution separation came from the resolution limit of 3.0 Å – 15.0 Å. Table 3.6 provides a list of the rotation search results with regards to Euler angles for this resolution limit. The Euler coordinate system is described in sections 3.1.3. For the rotation function, the correct solution may be in the list with a Z-score (number of standard deviations above the mean value) under 4, and will not be found until a translation function is performed and a correct solution is selected. It is no surprise that no obvious solution presented itself in Table 3.6, therefore, brute force likelihood translation was performed in Phaser from the CCP4 suite using the entire solution set in Table 3.6 (McCoy *et al.*, 2005). A partial list of the solutions for the translation function is given in Table 3.7. Translation was preformed in all of the following space groups, P622, $P6_322$, $P6_222$, $P6_422$, $P6_122$ and $P6_522$.

Based on the log-likelihood gain (LLG) and Z-score, there are several good solutions that present themselves from the translation search. The top three solutions are bold faced in Table 3.7 and two of them are illustrated graphically in Figure 3.9 and Figure 3.10. It is obvious from looking at the solutions with their symmetry mates within the unit cell, that no real solution was found. Due to impossible packing, all of the solutions found through rotation and translation searches have to be ruled out. Based on

Table 3.6: Partial list of the best rotation search results for the d[CG(5FU)G(5FU)GC-ACACG] search models performed in the space group P622 using Phaser from the CCP4 suite (Storoni *et al.*, 2005).

Ensemble	α	β	γ	Z-score
bd0001B	25.510	70.224	135.864	3.40
bd0001B	1.274	75.874	138.971	3.24
bd0001A	54.996	66.636	49.787	3.11
bd0067A	6.242	13.796	142.084	3.71
bd0067A	36.842	13.796	142.797	3.53
bd0067A	59.462	72.337	314.075	3.30
bd0067B	24.649	14.149	109.043	3.78
bd0067B	54.745	14.149	110.127	3.45
bd0067B	56.712	19.984	97.524	3.33
bd0067B	0.789	14.149	114.194	2.94
bdl084A	11.534	13.386	139.307	3.53
bdl084A	49.974	19.280	140.581	3.44
bdl084B	8.246	11.766	108.580	3.41
bdl084B	41.005	11.766	106.716	3.34

Table 3.7: Translation search results for the d[CG(5FU)G(5FU)GCACACG] search models performed in Phaser from the CCP4 suite (McCoy *et al.*, 2005). Based on LLG and Z-score, the top three solutions are bold faced. *Log-likelihood gain (LLG).

Ensemble	α	β	γ	transX	transY	transZ	LLG*	Z-score
Space group: **P622**								
bd0001B	25.5	70.2	135.9	0.170	0.073	0.454	12.3781	4.53
	1.3	75.9	139.0	0.424	0.166	0.099	12.2144	4.30
bd0001A	55.0	66.6	49.8	0.154	0.065	0.439	9.63728	3.85
bd0067A	6.2	13.8	142.1	0.824	0.782	0.192	12.3700	4.36
	36.8	13.8	142.8	0.701	0.928	0.493	10.8425	3.86
bd0067B	24.6	14.1	109.0	0.547	0.528	0.360	11.1473	3.90
	54.7	14.1	110.1	0.724	0.767	0.395	14.0591	5.50
				0.447	0.073	0.493	13.5345	5.28
bdl084A	11.5	13.4	139.3	0.986	0.505	0.025	10.9698	4.09
	50.0	19.3	140.6	0.801	0.967	0.104	11.4498	4.07
bdl084B	8.2	11.8	108.6	0.863	0.813	0.168	12.7601	4.70
	41.0	11.8	106.7	0.562	0.882	0.069	10.6616	3.70
Space group: **$P6_122$**								
bd0001B	25.5	70.2	135.9	0.724	0.143	0.266	15.1175	5.33
	1.3	75.9	139.0	0.424	0.119	0.148	13.0136	4.76
bd0001A	55.0	66.6	49.8	0.686	0.643	0.321	10.5343	5.02
bd0067A	**6.2**	**13.8**	**142.1**	**0.424**	**0.374**	**0.444**	**18.9593**	**7.29**
				0.424	0.374	0.227	17.9959	6.88
				0.455	0.389	0.163	17.8555	6.82
				0.424	0.374	0.010	16.4707	6.24
				0.431	0.366	0.380	16.2838	6.17
	36.8	13.8	142.8	0.270	0.482	0.025	15.4683	5.29
bd0067B	24.6	14.1	109.0	0.817	0.813	0.276	11.3221	3.90
	54.7	14.1	110.1	0.331	0.073	0.375	14.4236	6.00
				0.316	0.019	0.074	13.8761	5.75
				0.285	0.004	0.207	13.4334	5.55
bdl084A	11.5	13.4	139.3	0.871	0.389	0.173	11.7446	4.39
	50.0	19.3	140.6	0.747	0.928	0.335	17.4761	6.82
				0.770	0.928	0.089	15.1006	5.79
bdl084B	8.2	11.8	108.6	0.755	0.782	0.498	15.4108	5.60
	41.0	11.8	106.7	0.986	0.274	0.281	13.6465	5.01
Space group: **$P6_522$**								
bd0001B	25.5	70.2	135.9	0.940	0.759	0.252	11.6096	4.36
	1.3	75.9	139.0	0.686	0.828	0.252	11.2244	4.85
bd0001A	55.0	66.6	49.8	0.886	0.258	0.158	8.00007	3.33
bd0067A	6.2	13.8	142.1	0.686	0.019	0.340	17.2029	6.61
				0.678	0.004	0.059	16.6145	6.36
				0.686	0.019	0.409	16.2099	6.19
	36.8	13.8	142.8	0.848	0.736	0.123	16.4702	6.44
				0.840	0.720	0.404	16.3616	6.39
bd0067B	24.6	14.1	109.0	0.894	0.158	0.074	17.5729	5.59
	54.7	14.1	110.1	0.709	0.713	0.409	16.9197	6.04
				0.709	0.713	0.340	16.4566	5.87
bdl084A	11.5	13.4	139.3	0.331	0.119	0.010	14.4834	5.48
	50.0	**19.3**	**140.6**	**0.378**	**0.189**	**0.335**	**18.9507**	**7.00**
				0.378	0.189	0.049	18.6754	6.88
				0.378	0.189	0.118	17.0619	6.23

bdl084B	8.2	11.8	108.6	0.478	0.158	0.015	14.8585	5.48
	41.0	11.8	106.7	0.362	0.042	0.084	16.7483	6.55
				0.709	0.713	0.084	16.3432	6.38
Space group: **$P6_222$**								
bd0001B	25.5	70.2	135.9	0.616	0.065	0.153	12.4337	4.57
	1.3	75.9	139.0	0.778	0.227	0.133	11.9651	4.30
bd0001A	55.0	66.6	49.8	0.778	0.065	0.281	9.16228	4.00
bd0067A	6.2	13.8	142.1	0.339	0.320	0.025	16.4867	6.51
				0.339	0.343	0.094	16.1069	6.34
	36.8	13.8	142.8	0.200	0.389	0.192	16.0075	5.82
bd0067B	24.6	14.1	109.0	0.909	0.166	0.325	17.3272	5.97
	54.7	**14.1**	**110.1**	**0.362**	**0.042**	**0.291**	**18.8825**	**7.10**
				0.378	0.050	0.158	18.8274	7.08
				0.378	0.050	0.227	18.7655	7.05
				0.378	0.050	0.444	18.1955	6.82
bdl084A	11.5	13.4	139.3	0.801	0.112	0.192	15.4981	5.94
	50.0	19.3	140.6	0.978	0.443	0.256	15.9905	6.59
				0.986	0.435	0.192	14.2857	5.79
bdl084B	8.2	11.8	108.6	0.385	0.204	0.084	16.6626	6.43
				0.385	0.181	0.301	15.8478	6.08
	41.0	11.8	106.7	0.817	0.836	0.168	16.1187	5.83
Space group: **$P6_422$**								
bd0001B	25.5	70.2	135.9	0.562	0.050	0.237	12.2408	4.56
	1.3	75.9	139.0	0.339	0.112	0.084	11.7144	4.22
bd0001A	55.0	66.6	49.8	0.131	0.227	0.291	9.15685	4.01
bd0067A	6.2	13.8	142.1	0.770	0.050	0.345	13.8612	5.10
	36.8	13.8	142.8	0.932	0.813	0.108	12.1377	4.52
bd0067B	24.6	14.1	109.0	0.824	0.158	0.192	11.4889	4.09
	54.7	14.1	110.1	0.378	0.073	0.227	14.2562	5.78
bdl084A	11.5	13.4	139.3	0.732	0.828	0.360	12.4210	5.12
	50.0	19.3	140.6	0.632	0.790	0.454	11.4615	4.23
bdl084B	8.2	11.8	108.6	0.408	0.158	0.330	14.4301	5.14
	41.0	11.8	106.7	0.293	0.019	0.429	13.2974	4.87
Space group: **$P6_322$**								
bd0001B	25.5	70.2	135.9	0.639	0.019	0.271	14.2183	5.65
	1.3	75.9	139.0	0.239	0.027	0.039	11.3881	4.16
bd0001A	55.0	66.6	49.8	0.678	0.443	0.039	9.27738	4.21
bd0067A	6.2	13.8	142.1	0.385	0.412	0.429	15.0214	5.06
	36.8	13.8	142.8	0.239	0.374	0.355	15.4294	5.79
bd0067B	24.6	14.1	109.0	0.447	0.882	0.049	9.11687	3.95
	54.7	14.1	110.1	0.586	0.559	0.108	12.5409	4.07
bdl084A	11.5	13.4	139.3	0.986	0.505	0.025	13.0834	4.40
	50.0	19.3	140.6	0.794	0.859	0.385	14.9965	5.29
bdl084B	8.2	11.8	108.6	0.863	0.836	0.168	14.8526	5.28
	41.0	11.8	106.7	0.986	0.320	0.064	11.7243	4.09

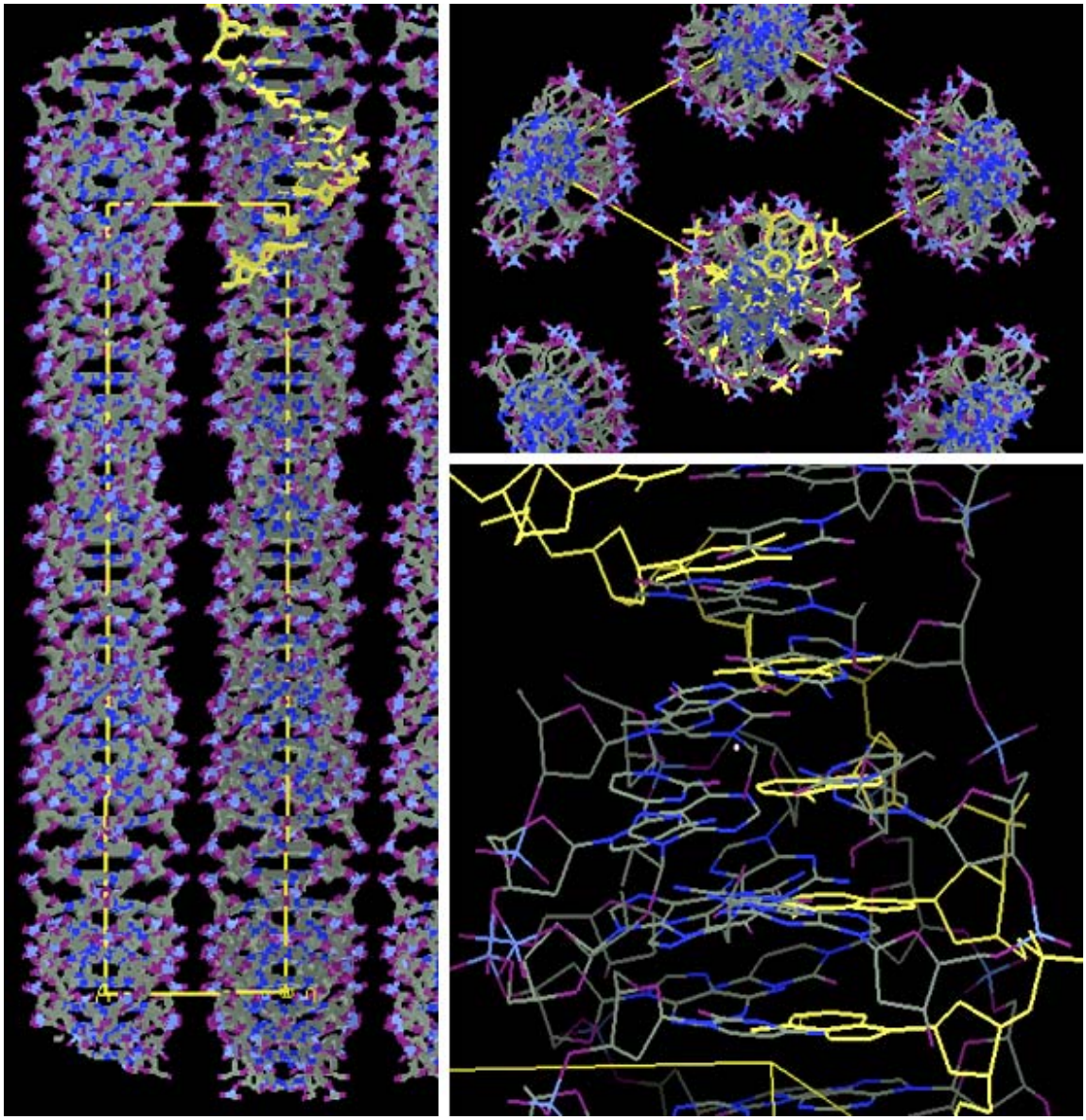

Figure 3.9: Graphic illustration of the top translation solution in the $P6_222$ space group. The image was viewed using the program Coot (Emsley and Cowtan, 2004). The yellow single strand represents ensemble BD0067B and the rotation/translation solution, the blue and purple represents its symmetry mates.

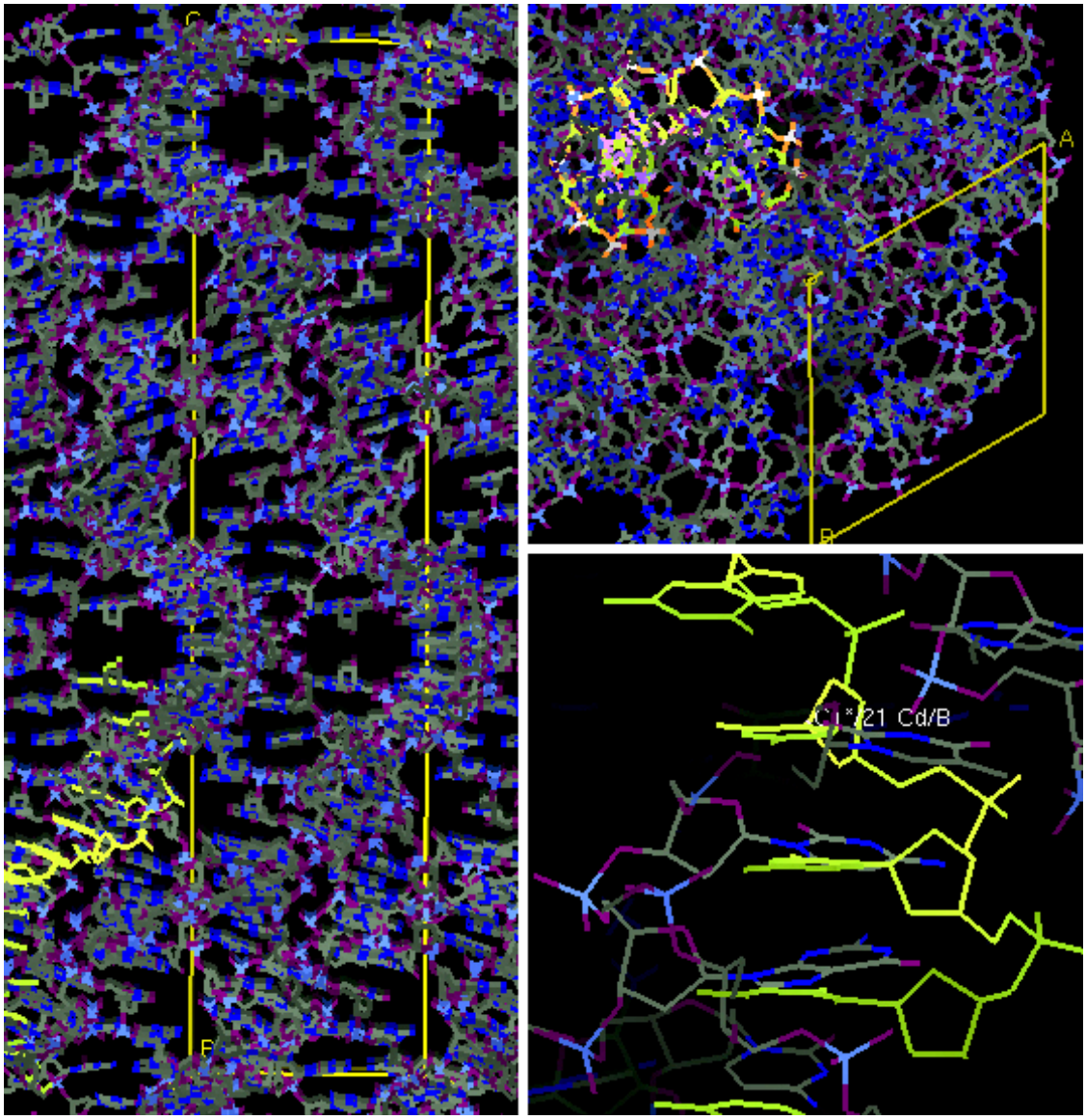

Figure 3.10: Graphic illustration of the top translation solution in the $P6_522$ space group. The image was viewed using the program Coot (Emsley and Cowtan, 2004). The yellow single strand represents ensemble BDl084A and the rotation/translation solution, the blue and purple represents its symmetry mates.

these observations, it is likely that the true symmetry of the lattice is lower (e.g. $P3_221$) but the apparent diffraction symmetry appears higher due to merohedral twinning of the crystals. Hence, the crystal would be composed of two subdomains related by a rotation of 60 ° about the c axis yielding an almost equal contribution of diffraction intensity from each of the two twinned domains. There is not enough room within the 6/mmm Laue group for the DNA to pack appropriately. It is most likely that the lower symmetry solution is the correct one. Perhaps the data is actually two $\bar{3}$m1 or two $\bar{3}$1m unit cells growing together, hence appearing like 6/mmm. Solving this data set would require more time and is beyond the scope of this thesis.

3.3 X-ray Absorption Spectroscopy
3.3.1 NiPC

Phthalocyanines are symmetrical 18 π-electron aromatic macrocycles, closely related to naturally occurring porphyrins. They are most commonly used as dyes or pigments. Figure 3.11 is an illustration of the nickel form of phthalocyanine. NiPC was chosen as a model compound for EXAFS analysis because the four nitrogen atoms are attached at similar distances from the center nickel. Because the crystal structure is known, this model provides an understanding of phase changes and parameter adjustments that would be required for fitting M-DNA.

X-ray absorption spectra for NiPC samples were collected and analyzed as described in section 2.3. The averaged normalized absorption spectra for the NiPC data is shown in Figure 3.12A followed by the isolated oscillatory structures, $\chi(k)$, after background removal in Figure 3.12B. Using the theoretical input file from Table A3 and the constraints listed in Table A4, the best fit analysis was calculated for the Fourier transformed NiPC absorption data (Figure 3.13). A list of the resulting EXAFS statistical data is given in Table 3.8. The fit results for the final refinement, after correcting for the scattering phase-shift (recall that the EXAFS goes as sin $[2kR + \delta]$) which is typically 0.5 Å, gives a Ni -N13 distance of R = 1.88 ± 0.05 Å and a mean square disorder of σ^2 = 0.001 ± 0.003 Å^2, a Ni - N9 distance of R = 1.94 ± 0.05 Å with a mean square disorder of σ^2 = 0.001 ± 0.003 Å^2, a Ni - C12 distance of R = 2.96 ± 0.05 Å and a mean square disorder of σ^2 = 0.001 ± 0.003 Å^2, a Ni - C14 distance of R = 2.98 ±

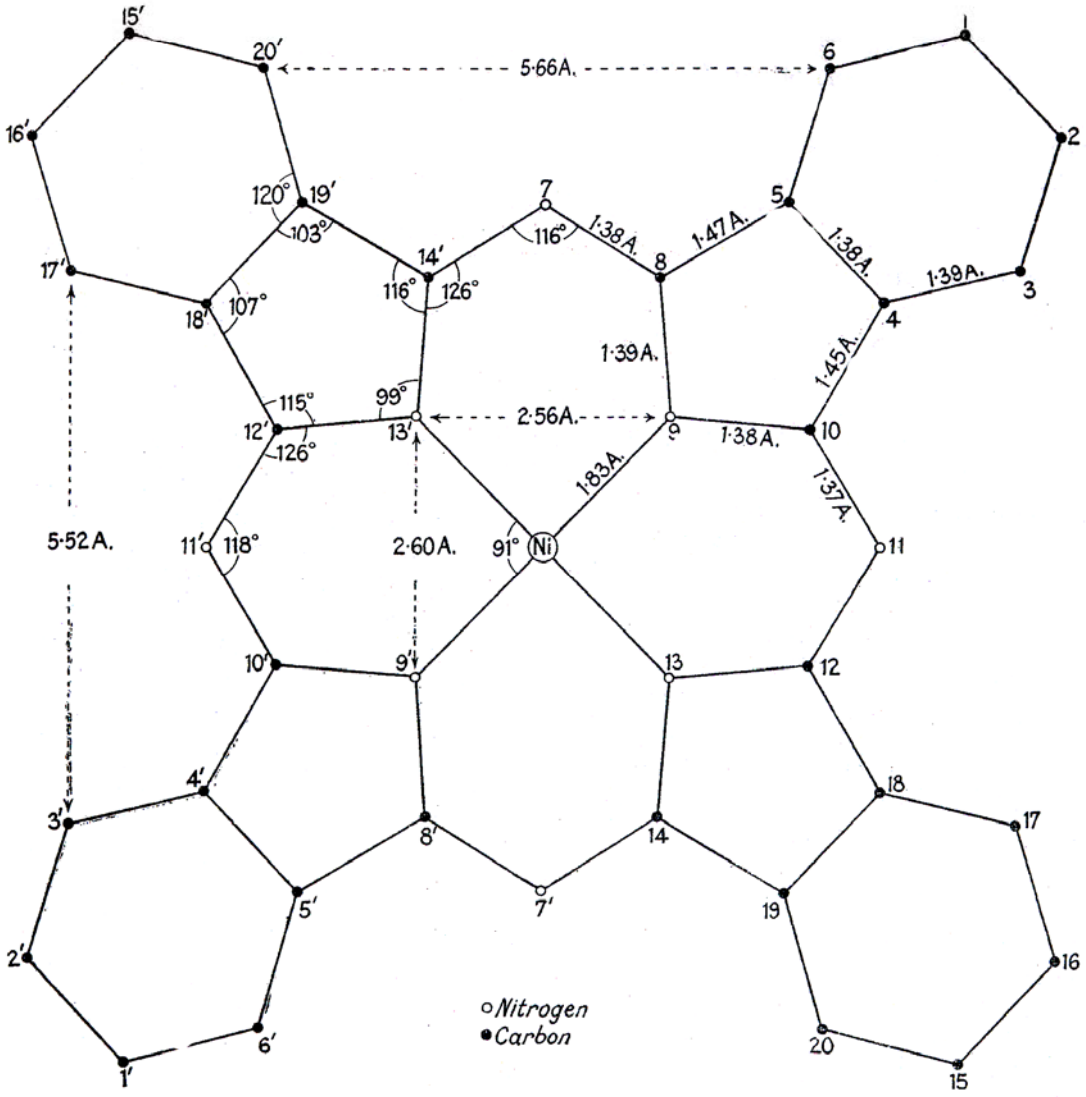

Figure 3.11: Dimensions of NiPC molecule. Note the four nitrogens surrounding the center nickel atom. Image adapted from Robertson and Woodward, 1937.

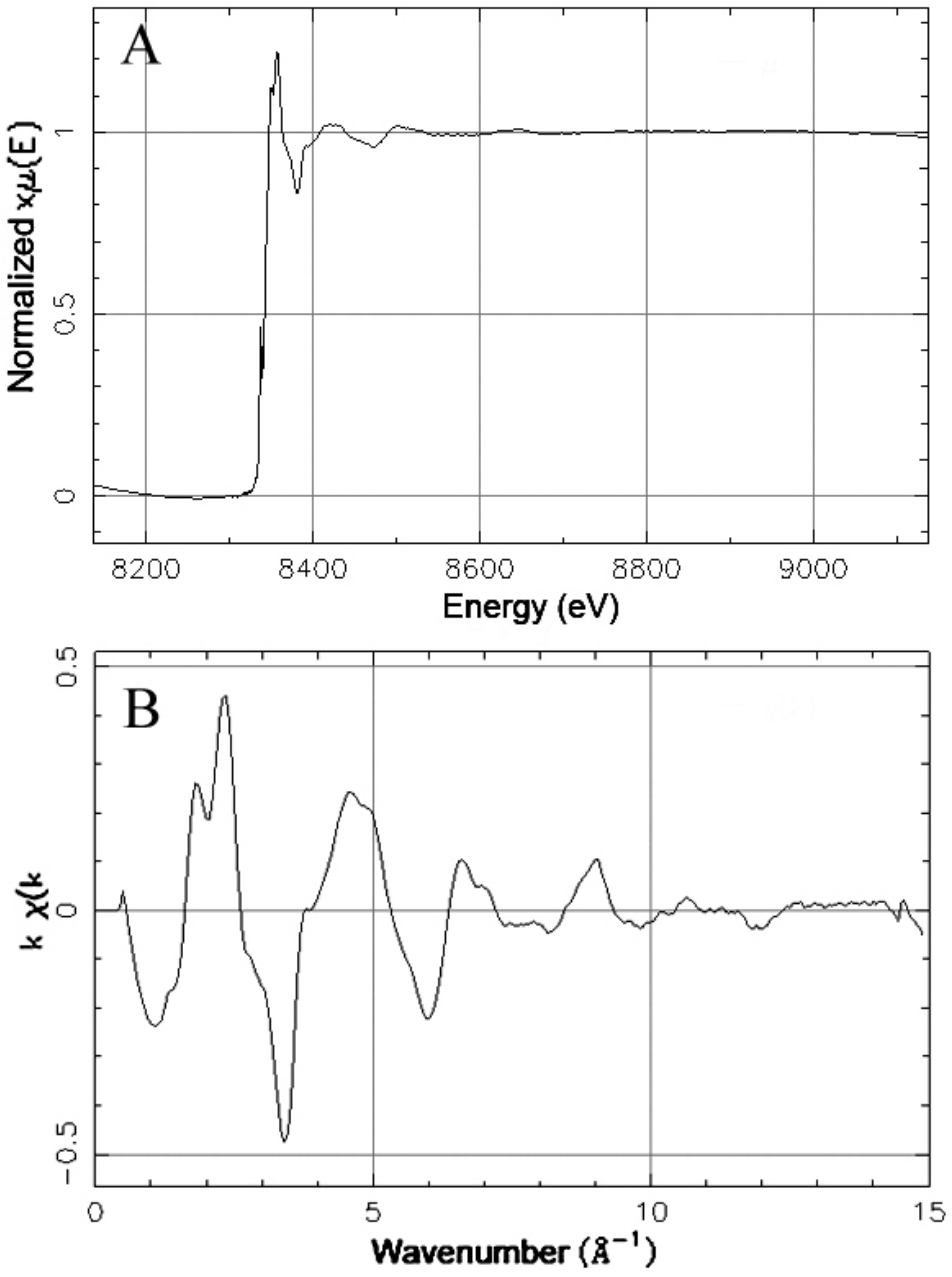

Figure 3.12: NiPC EXAFS results. A) The average normalized absorption data $x\mu(E)$ for the nickel *K* edge and B) the $\chi(k)$ spectra for the Ni *K* edge.

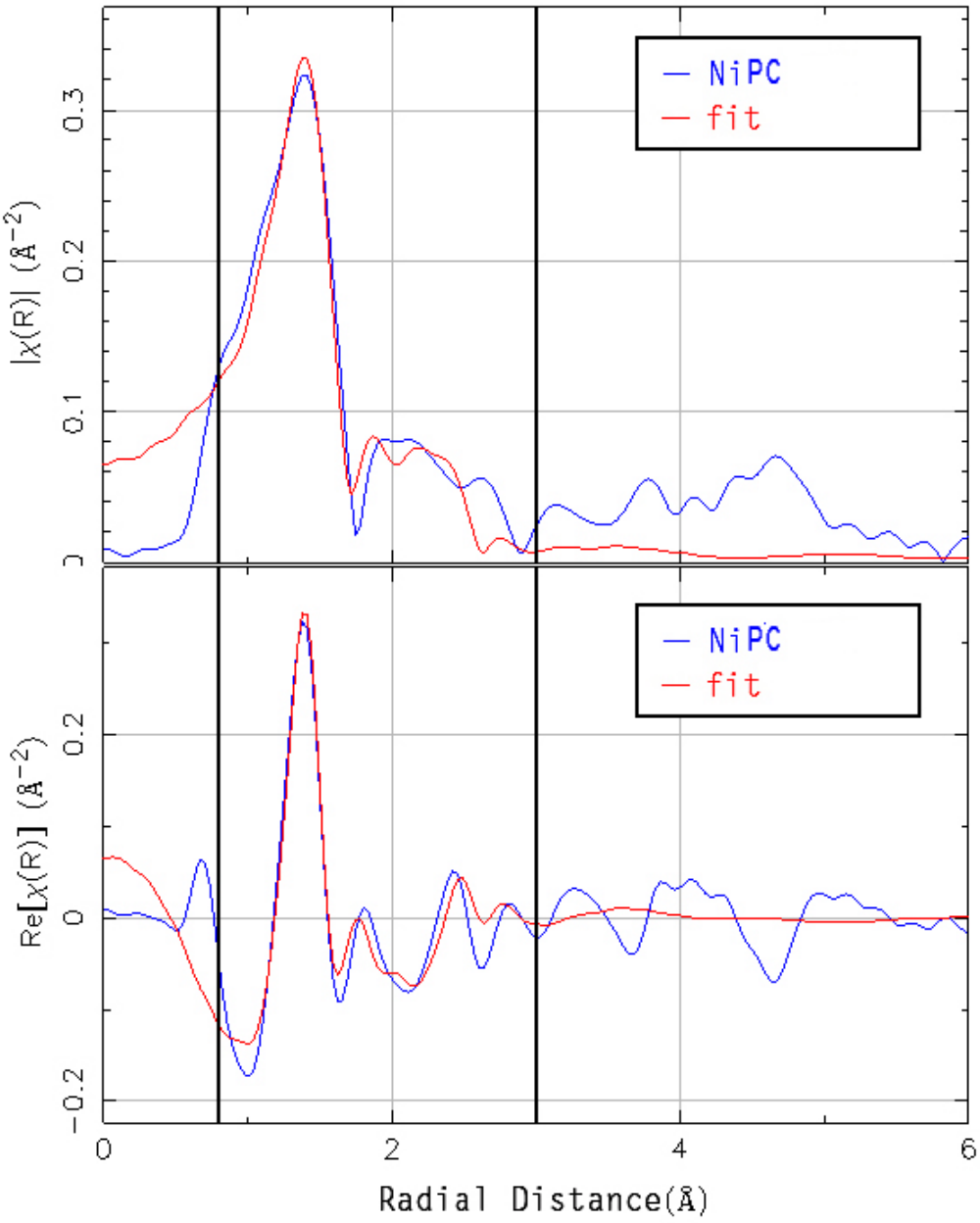

Figure 3.13: NiPC fit results in *R* space for the Ni *K* edge. The top panel shows the magnitude of $\chi(R)$, the complex Fourier transform of $k\,\chi(k)$, and the bottom panel shows the real part. The thick vertical lines represent the range over which the data were Fourier transformed for analysis.

Table 3.8: NiPC EXAFS statistical data..

Parameter		Result/Best Fit
Fitting Statistics		
	k-range ($Å^{-1}$)	2.0 – 14.35
	R-range (Å)	0.8 – 3.0
	χ^2	1255
	Reduced χ^2	138
	R-factor	0.044
	k-weight	1.0
	E (eV)	8337
	Fitting space	*R*
Guess parameters		
Path Ni – N13	*R*	1.88
	S_0^2	0.99
	E_0	3.33
	ΔR	0.05
	σ^2	0.001
Path Ni – N9	*R*	1.94
	S_0^2	0.99
	E_0	3.33
	ΔR	0.05
	σ^2	0.001
Path Ni – C12	*R*	2.97
	S_0^2	0.99
	E_0	22.11
	ΔR	0.05
	σ^2	0.044
Path Ni – C14	*R*	2.99
	S_0^2	0.99
	E_0	22.11
	ΔR	0.05
	σ^2	0.044
Path Ni –C10	*R*	2.51
	S_0^2	0.99
	E_0	22.11
	ΔR	-0.47
	σ^2	0.003

0.05 Å and a mean square disorder of $\sigma^2 = 0.044 \pm 0.032$ Å^2 and, finally, a Ni – C10 distance of R = 2.51 ± 0.47 Å and a mean square disorder of $\sigma^2 = 0.003 \pm 0.006$ Å^2. All of these distances for the first five paths in the fit agree within the limits of error with the crystallographic data for NiPC. Additional scattering paths and impurities in the 85% pure NiPC were not determined.

3.3.2 Nickel B and M-DNA

X-ray absorption spectra for both nickel B and M-DNA samples were collected and analyzed as described in section 2.3. The averaged normalized absorption spectra for both data sets are shown in Figure 3.14 followed by the isolated oscillatory structures, $\chi(k)$, after background removal in Figure 3.15. Using the theoretical standards described in Table A1 and Table A2 and the constraints from Table A4, fits were calculated for the Fourier transformed nickel B and M-DNA absorption data (Figure 3.16). A list of the resulting EXAFS statistical data is given in Table 3.9.

The results for the Ni^{2+} M-DNA data vary only slightly from the results for the Ni^{2+} B-DNA data (Figure 3.17), however, it is these variations that are of interest in this thesis and a few important points shall be made. The first things to notice in Figure 3.17 are the circles labelled A, B and C which correspond to the Ni^{2+}-N7, the Ni^{2+} within the base pair and the Ni^{2+}-OP distances, respectively. The differences observed after circle C are beyond the resolution limits examined within this experiment. The next thing to notice is that these three distances correspond to 1.3 Å, 1.6 Å and 1.8 Å in Figure 3.17. In order to determine potential real distances, we have to correct for the scattering phase-shift (recall that the EXAFS goes as sin $[2kR + \delta]$) which is typically 0.5 Å. Hence, the new distances are approximately 1.8 Å, 2.1 Å and 2.3 Å. A closer look at the fit results for the final refinement gives a Ni^{2+}-G N1 distance of $R = 2.04 \pm 0.03$ Å and a mean square disorder of $\sigma^2 = 0.002 \pm 0.378$ Å^2, a Ni^{2+}-C N3 distance of $R = 1.99 \pm 0.23$ Å and a mean square disorder of $\sigma^2 = 0.005 \pm 1.224$ Å^2 and a Ni^{2+}- O distance of $R = 2.27 \pm 0.14$ Å and a mean square disorder of $\sigma^2 = 0.000 \pm 0.133$ Å^2. Based on the EXAFS results presented in this thesis, a model for the structure of M-DNA is presented with these bond distances (Figure 3.18).

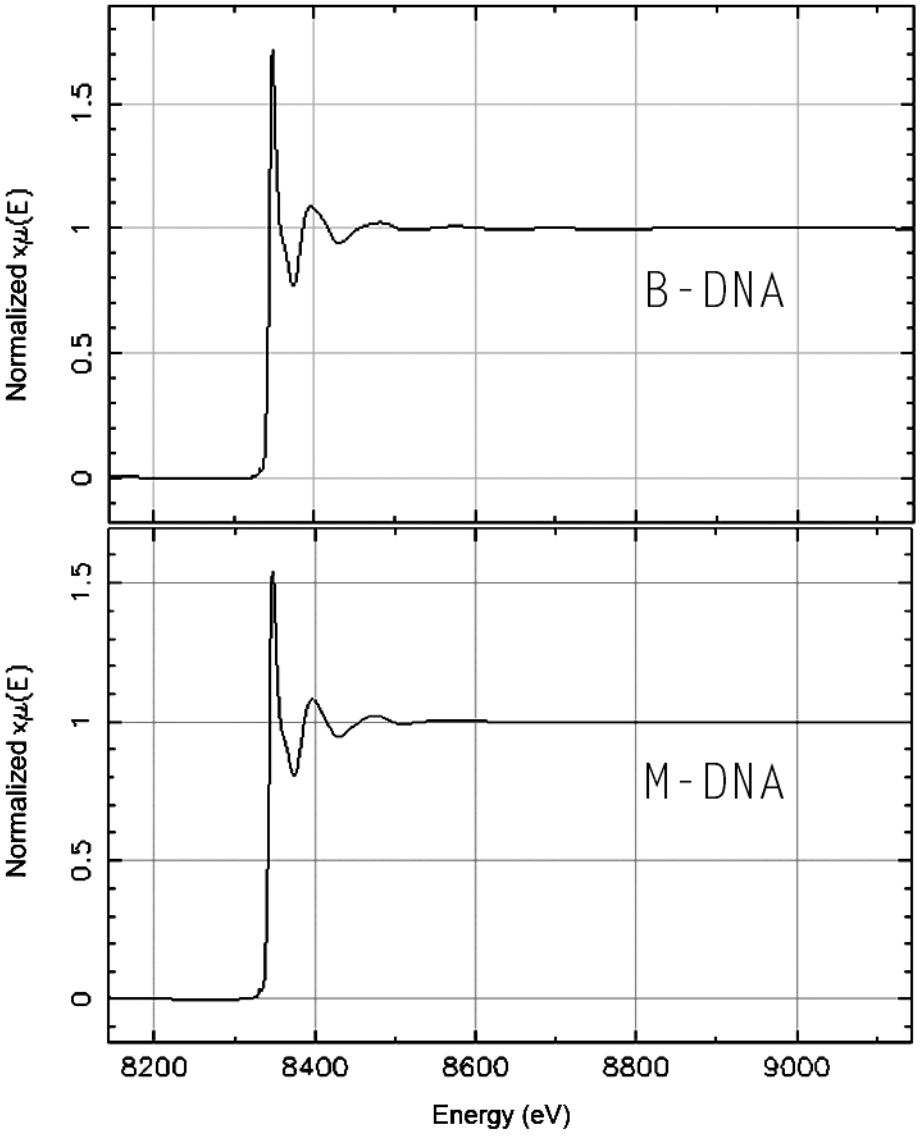

Figure 3.14: The average normalized absorption data $x\mu(E)$ for the nickel *K* edge of both Ni^{2+} (top) B-DNA and (bottom) M-DNA. The B-DNA data is an average obtained from five separate sets of data collection, whereas the M-DNA data is an average of eight.

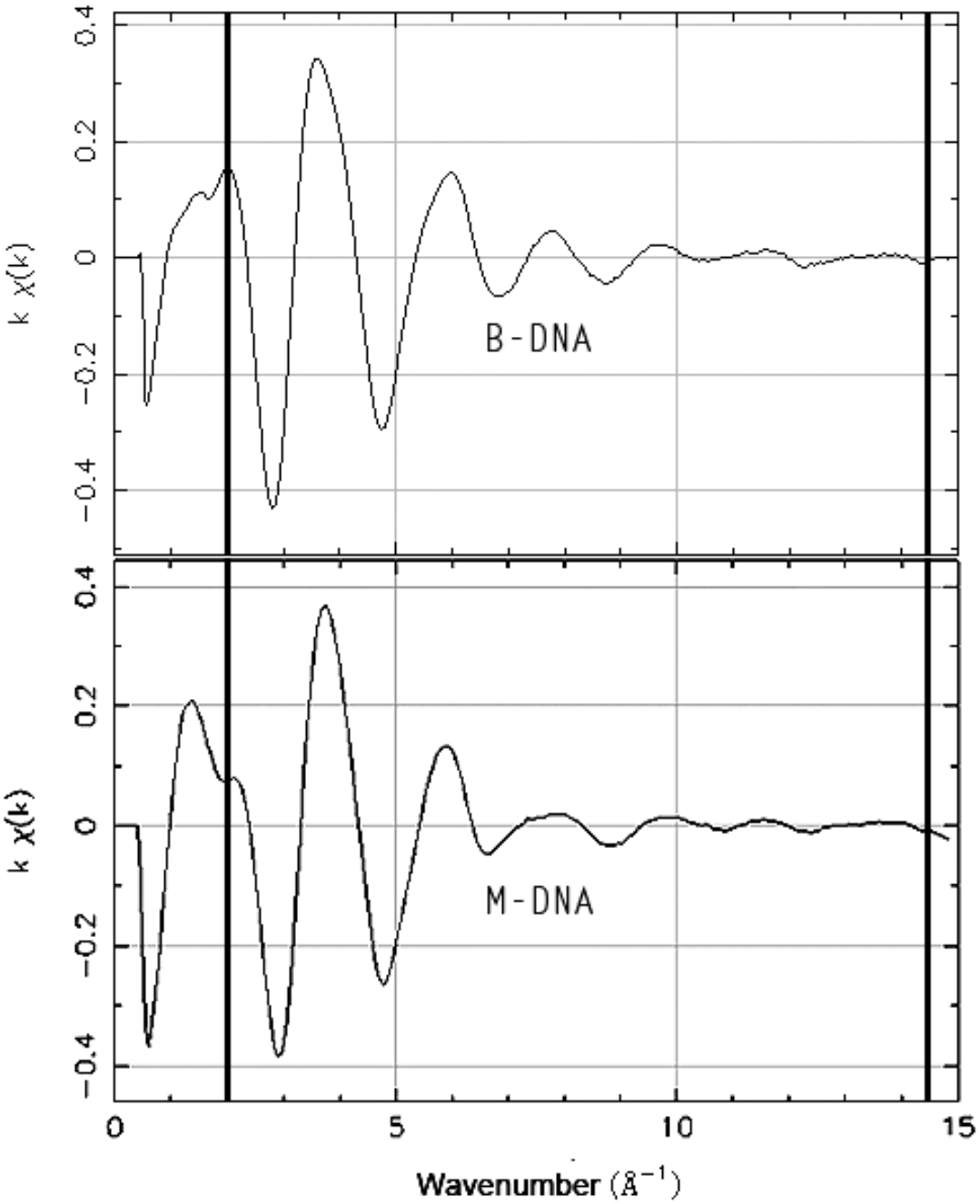

Figure 3.15: $\chi(k)$ spectra for the Ni *K* edge of both Ni^{2+} (top) B-DNA and (bottom) M-DNA. The thick vertical lines represent the range over which the data were Fourier transformed for analysis.

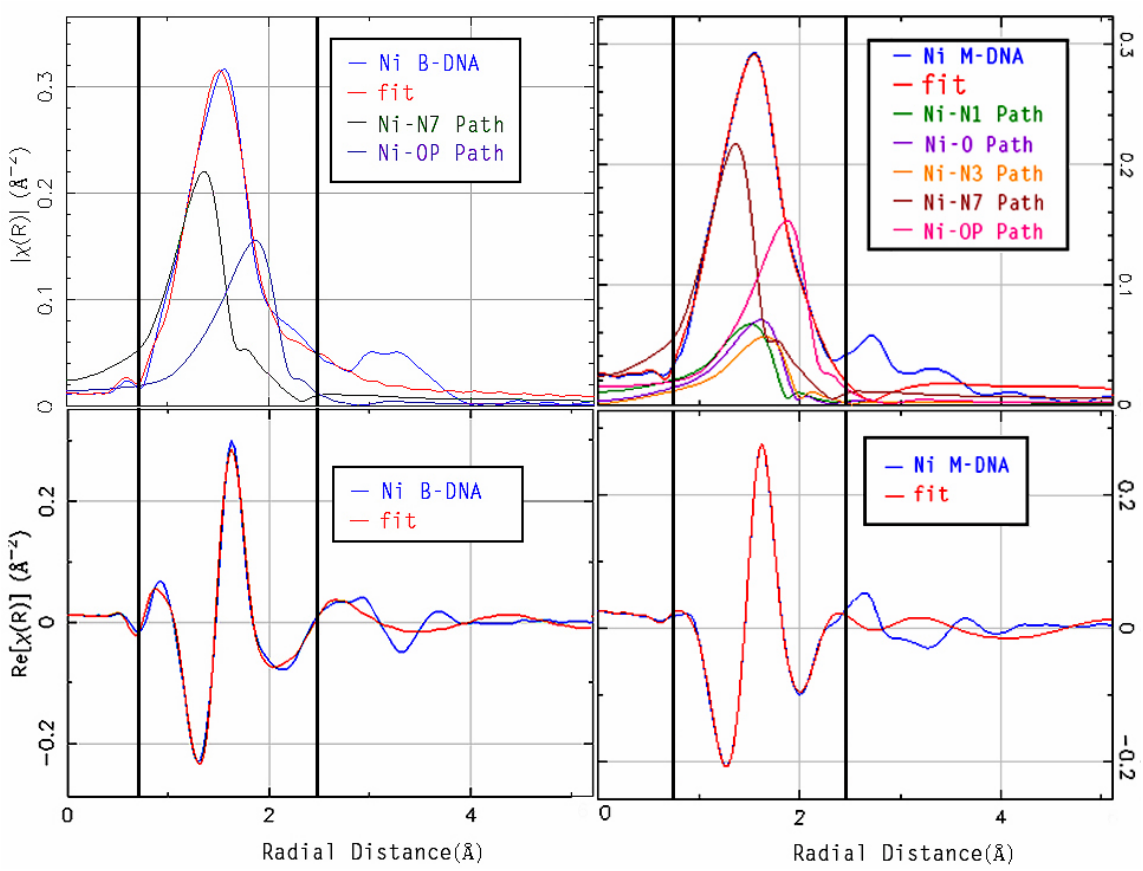

Figure 3.16: Fit results in *R* space for the Ni *K* edge of both Ni^{2+} (left) B-DNA and (right) M-DNA. The top panels show the magnitude of $\chi(R)$, the complex Fourier transform of $k\,\chi(k)$, and the bottom panels shows the real part. The thick vertical lines represent the range over which the data were Fourier transformed for analysis.

Table 3.9: Nickel B and M-DNA EXAFS statistical data.

Parameter		Result/Best Fit B-DNA	M-DNA
Fitting Statistics			
	k-range (Å^{-1})	2.0 – 14.7	2.0 – 15.6
	R-range (Å)	0.8 – 2.4	0.8 – 2.4
	χ^2	6082	12149
	Reduced χ^2	1423	121493
	R-factor	0.006	0.006
	k-weight	1.0	1.0
	E (eV)	8345	8343
	Fitting space	*R*	*R*
Guess parameters			
Path Ni – N7	*R*	1.93	2.01
	S_0^2	-5.39	0.88
	E_0	0.30	-10.04
	ΔR	0.10	0.18
	σ^2	0.014	0.006
Path Ni – OP	*R*	1.99	2.15
	S_0^2	4.70	0.88
	E_0	-0.60	5.95
	ΔR	-0.37	-0.21
	σ^2	0.024	0.003
Path Ni – N1	*R*	----	2.04
	S_0^2	----	0.88
	E_0	----	-10.04
	ΔR	----	-0.03
	σ^2	----	0.001
Path Ni – O	*R*	----	2.27
	S_0^2	----	0.88
	E_0	----	5.95
	ΔR	----	0.14
	σ^2	----	0.001
Path Ni – N3	*R*	----	1.99
	S_0^2	----	0.88
	E_0	----	-10.04
	ΔR	----	-0.23
	σ^2	----	0.005

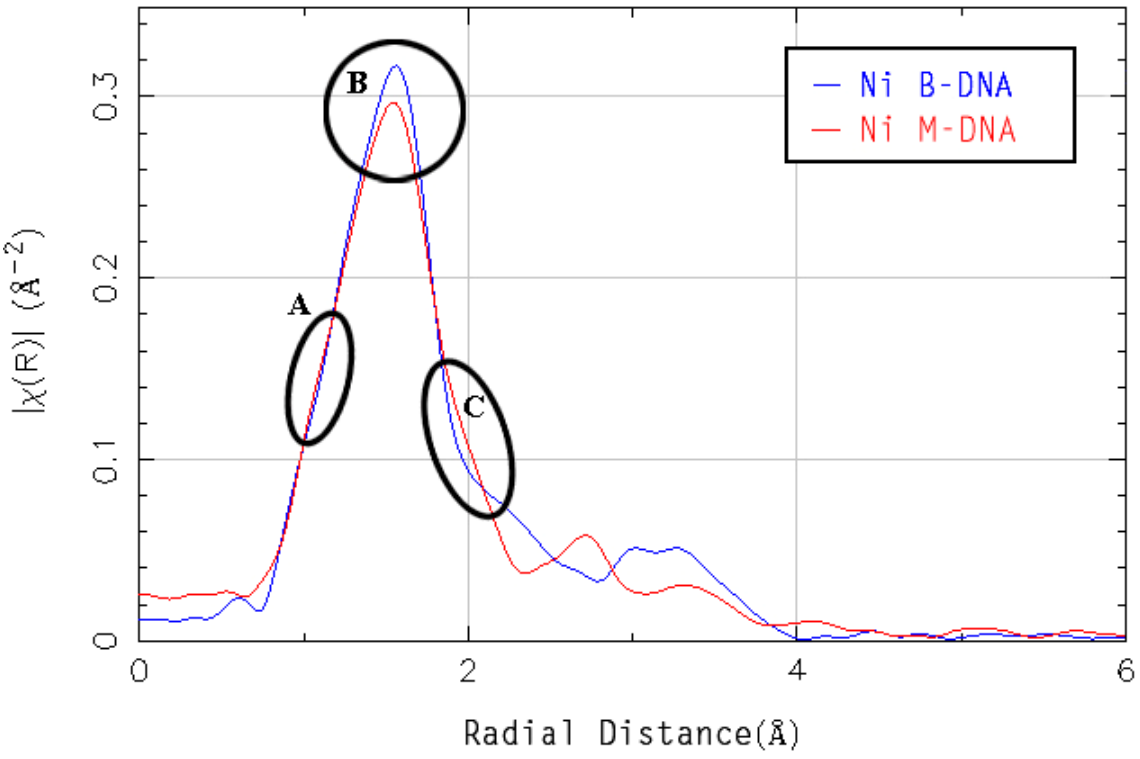

Figure 3.17: Differences in *R* space for the M-DNA and the B-DNA EXAFS spectra represented by the magnitude of $\chi(R)$, the complex Fourier transform of $k\,\chi(k)$, for the Ni *K* edge of both Ni^{2+} (blue) B-DNA and (red) M-DNA. The black circles represent three specific areas where the M-DNA spectrum is different from the B-DNA spectrum. The differences in the oscillations after circle C are not fitted because they are due to scattering from the aromatic rings and beyond the resolution of the data.

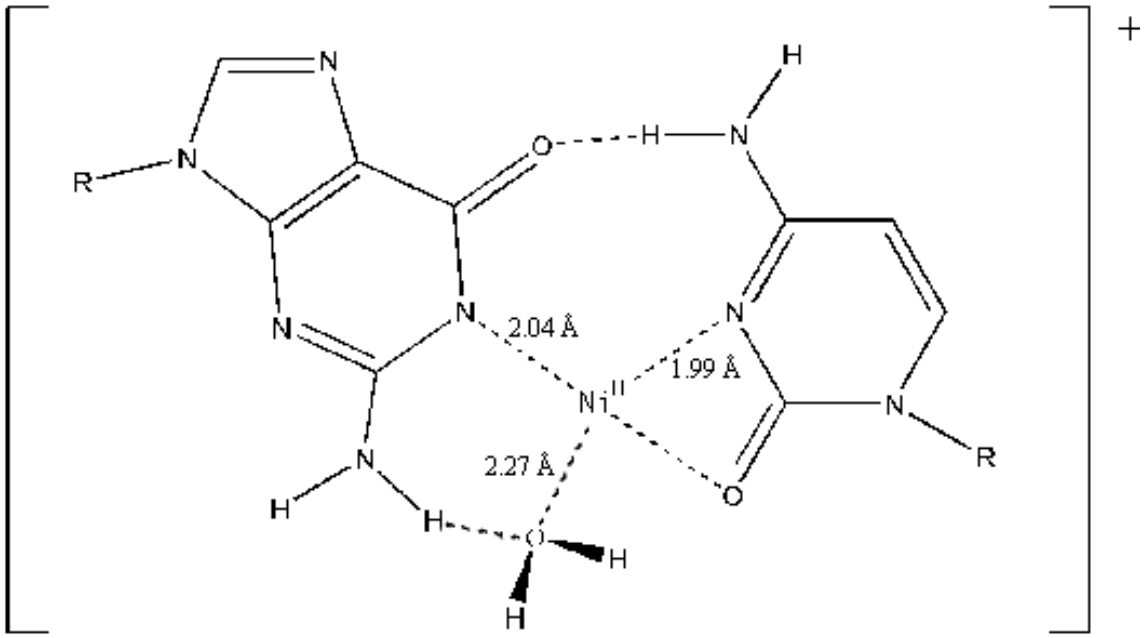

Figure 3.18: Proposed distances in an M-DNA model for the base pair G-C determined from EXAFS analysis.

4.0 DISCUSSION

4.1 Crystallography

The oligonucleotide, d[GA(5FU)(5FU)AA(5FU)C], was crystallized successfully in the presence of either Co^{2+} or Zn^{2+}. The best diffraction came from the crystals grown with Zn^{2+} diffracted on the DX8 Proteum Diffractometer in the Saskatchewan Structural Sciences Centre at the University of Saskatchewan. The crystal initially diffracted to approximately 2.7 Å followed by a complete disappearance of all the diffraction. This phenomenon is one that was quite unheard of previously. It was first thought that it was simply a case of crystal slippage, a one time experimental error. However, after this diffraction disappearance occurred a second time under very similar circumstances, it was apparent that another explanation was needed. It is possible that the crystals are simply not stable enough in their cryoprotectant and are disintegrating due to the X-ray exposure. This phenomenon was not observed for lower resolution crystals with the same cryoprotectant, therefore it is unlikely that it is a fault of the cryoprotectant. This phenomenon has only been observed on the DX8 Proteum Diffractometer. Perhaps it is a reaction specific to the X-ray source or wavelength of 1.54 Å. To date there has only been speculation as to why the high resolution oligonucleotide crystal diffraction disappears after initial X-ray exposure. More work is necessary in order to determine the true cause. Some suggestions include working with different wavelengths, temperatures and cryoprotectants in order to try and make this disappearance more reproducible and consistent. It will be necessary to determine if this is a phenomenon specific to this sequence, to M-DNA, or to DNA crystals in general. Perhaps it can be explained by experimental error, or may be it is something explained through the experimental conditions.

The results from the diffraction data collected from the oligonucleotide, d[GA(5FU)-(5FU)AA(5FU)C], in the presence of Co^{2+} were very limited due to the poor resolution of the data. There are definite consistencies to be observed about the crystal packing,

however, no useful conclusions can be made with regards to the location and binding of any metal within or around that duplex. Ultimately, better resolution diffraction data are needed and in order to get better resolution diffraction data, one must return to the crystal growing process.

Much more work needs to be done with the data collected on the d[CG(5FU)G-(5FU)GCACACG] crystals. The resolution of this data is quite good, at least as high as 2.8 Å and possibly as high as 2.5 Å could be usable. Before this data can be solved it must be appropriately detwinned. Currently there is software available to detwin data sets. However, due to the particular diffraction symmetry apparent with these crystals, detwinning will not be straight forward. There are several different solution possibilities that will have to be explored and considered carefully.

Some final conclusions may be made regarding the structure of crystals grown in M-DNA favouring conditions and the way the duplexes pack within the unit cell. The results from both the d[CG(5FU)G(5FU)GCACACG] and the d[GA(5FU)(5FU)AA-(5FU)C] crystals indicate duplexes stacking consistently end to end creating continuous columns throughout the crystals. In the d[CG(5FU)G(5FU)GCACACG] data, the columns were commonly arranged parallel to the *c*-axis, whereas in the d[GA(5FU)(5FU)AA-(5FU)C] data, the columns were both parallel and at 90˚ to the axes in the unit cell. This end to end type of packing is not uncommon throughout the DNA structures solved to date. However, it is often accompanied by poor quality diffraction. The end to end stacking of the DNA duplexes encourages few stabilizing interactions to occur along the phosphate backbones of the duplexes. It is common to have problems with diffraction quality, especially when considering DNA duplexes that are stacked end to end. In order to improve the diffraction quality of the DNA crystal, it may be helpful to include one or more additives in the crystal growing solutions that may function to stabilize the packing. This is easier said then actually done. Certain additives may encourage more stable packing, or they may hinder it. There is little known about what works and what doesn't in crystallography. Until something has been attempted, there is no way of knowing whether it will work or not. Several additives have been experimented with so far, however, there is always something new to try. This is the joy and the frustration that is crystallography. It may also be helpful to continue to

look at different sequences, perhaps find one that can crystallize in a structure that is not end to end, yet still can crystallize in M-DNA favouring conditions.

Before substantial conclusions can be drawn about the detailed structure of M-DNA from X-ray crystallography, the quality of data has to see a significant improvement. This ultimately requires improving the crystals from which the diffraction is obtained. Once the crystal quality has been improved, solving the structure using a method other than molecular replacement may be necessary. Molecular replacement is based on having a model similar to the unknown structure. Because M-DNA is a novel structure, limiting it to a B-like form may not be appropriate. One method that would be ideal for solving an X-ray crystallographic model for M-DNA is known as multiple wavelength anomalous diffraction (MAD). In MAD, changes are induced in the atomic scattering factor of a heavy-atom bound to the DNA by measuring diffraction data at three of four different X-ray energies where the anomalous scattering factors of the heavy atom are significantly different from one another (Stura and Gleichmann, 1999). In order to collect MAD data, it is necessary to have high quality crystals. Lack of high resolution data is the reason that MAD was not the method used in this thesis. What it all comes down to in X-ray crystallography, are the crystals.

4.2 X-Ray Absorption Spectroscopy

4.2.1 NiPC

EXAFS analysis on NiPC was performed more as an example or practice with a molecule containing multiple nitrogen bonds of varying length to the absorber, or the nickel in this case. It was a way of verifying the programs and methods used in analyzing the EXAFS data. The NiPC EXAFS results did coincide with that previously known of the structure of NiPC, although, it must be noted that an accurate model was used to create the theoretical scattering amplitude and phase shifts. It is no surprise that the EXAFS results output the same distances as the crystallographic results when the crystallographic results are part of the programs input. This is an example of what should happen when a correct model is used to calculate the theoretical scattering factors. Things would be more difficult if the theoretical scattering factors were calculated from a structure that is dissimilar to the real structure.

4.2.2 Nickel B and M-DNA

The program, Artemis (Ravel and Newville, 2005), used in the EXAFS analysis is based on a best fit calculation varying several mathematical parameters. The more paths you add to the calculation, the more parameters and the better the fit will be regardless of whether the paths are realistic or not. In order to limit the addition of more parameters, there is a limit to the number of free variables allowed in the fit. This maximum number of free variables that can effectively fit in a refinement depends on the *k* and *R*-ranges considered (Newville, 2004). For the M-DNA spectrum, the number of variable parameters was approaching the upper limit allowed. The R-factors for both the M and the B-DNA data were 0.006, representing very good statistical fits.

This type of EXAFS analysis method is very biased towards the theoretical standards used in the fit, creating similar problems as with molecular replacement in X-ray crystallography. Because M-DNA is a novel structure, it would be ideal to find a method of structural analysis that does not require theoretical input. The EXAFS analysis presented in this thesis has established a final model for an M-DNA G-C base pair that is very similar to the one created by Les W. Tari (Aich *et al.*, 1999), the only differences being in the bond distances surrounding the absorber, or in this case, the Ni^{2+}. The old bond distances, 2.217 Å, 2.065 Å and 2.137 Å have been replaced with 1.99 Å, 2.04 Å and 2.27 Å for the C N3-Ni, the G N1-Ni and the O-Ni bonds, respectively. However, the old bond distances are near the extreme ends of the error limit of the new bond distances, indicating no real significant change in the new model for the G-C M-DNA base pair. The end result is an experimental confirmation that M-DNA is formed by replacement of protons within the base pairs. Fitting the parameters for the A-T base pair would improve the fitting of the experimental data. Once a model has been calculated for the A-T base pair of M-DNA, it may be beneficial to return to these EXAFS spectra for further analysis. After all, the EXAFS spectrum is an average of all of the different metal species in the sample. DNA isolated from calf thymus does contain many A-T base pairs as well as G-C base pairs. Comparing the M-DNA and the B-DNA clearly illustrates that the M-DNA scattering is different from the B-DNA. The oscillations in the bond distances at large distances are a clear indication of scattering from a large

aromatic electron density. The M-DNA and B-DNA have different electron densities in the aromatic rings.

Many of the experiments done to date using M-DNA have indicated some interesting differences between the Ni^{2+}, the Co^{2+} and the Zn^{2+} versions of the complex. For example, the different metal forms of M-DNA all have unique and consistent rates of conversion between M-DNA to B-DNA and vice versa, as well as different pH or metal to DNA ratio requirements for formation (Wood *et al.*, 2002). Does this difference suggest variations in the structures of the three different metal forms of M-DNA? The next step in the EXAFS analysis of M-DNA will be to analyze the raw data from both the Co^{2+} and Zn^{2+} M-DNA complexes. It is necessary in order to determine if there are any consistent differences in the structures of the three different metal forms of M-DNA.

Although EXAFS analysis can present useful structural information about the local environment around an absorber, it remains necessary to continue with techniques such as X-ray crystallography in order to learn about the entire environment of the DNA duplex. It may also be necessary to continue with other techniques such as NMR or X-ray crystallography because of the differences in sample preparation techniques within the different methods. Namely, in EXAFS analysis, the sample was prepared as a solid and its structure may be altered in the drying of the molecules. X-ray crystallography and NMR may be performed on liquid samples, hence encouraging a more realistic and practical environment. EXAFS analysis is not without its benefits and advantages. However, in order to obtain a more complete picture, further structural analysis of M-DNA must continue using a variety of methods and techniques.

4.3 Final Conclusion and Future Prospective

The work presented in this thesis is consistent with the theoretical model currently accepted for M-DNA. However, I have several recommendations for the continuation of this project. In order to get a more realistic image of the EXAFS, it is a good idea to collect more EXAFS data on liquid samples rather then dried solids. It will also make the fit more accurate if a model for the A – T base pair can be added to the scattering paths calculated. Also, the resolution range for the fit should be extended and some multiple scattering paths could be added. However, the more fits that are added, the

better the fit will be due to increasing the number of variables. There is a fine line between adding enough scattering paths to represent the true structure verses adding too many.

It will be possible to process the d[CG(5FU)G(5FU)GCACACG] data set further. It is one of the more complicated examples of a twined data set, however, solving it is definitely possible. The data is quite redundant, therefore, I recommend performing a heavy atom search using programs such as SOLVE or SHELX once the data is appropriately detwinned. Better resolution data is required for the d[GA(5FU)(5FU)-AA(5FU)C] crystals before this can go any further. Due to the end-to-end packing within these crystals it may not be likely to get the resolution any better. For this I would recommend spending more time exploring other sequences, rather then trying to improve this one. None of the alternate sequences screened for crystallization conditions throughout the work presented in this thesis have been previously crystallized. I recommend choosing new sequences that have been solved and are currently in the NDB. This would result in a starting point for the crystallization conditions for the B-DNA form of the duplex and may save time in searching for the initial crystallization conditions of novel sequences. A few other suggestions include, forming the M-DNA prior to making the crystallization set up and trying to find crystallization conditions at room temperature rather then 4 º due to the nature of M-DNA. I believe it will be possible to obtain and X-ray crystallographic model M-DNA.

5.0 REFERENCES

Abrescia, N. A., Huynh-Dinh, T. and Subirana, J. A. (2002). Nickel-guanine interactions in DNA: crystal structure of nickel-d[CGTGTACACG]2. J Biol Inorg Chem *7*, 195-199.

Ahrland, S., Chatt, J. and Davies, N. R. (1958). The Relative Affinities of Ligand Atoms for Acceptor Molecules and Ions. Quart. Rev. Chem. Soc. *12*, 265-276.

Aich, P., Labiuk, S. L., Tari, L. W., Delbaere, L. J., Roesler, W. J., Falk, K. J., Steer, R. P. and Lee, J. S. (1999). M-DNA: A complex between divalent metal ions and DNA which behaves as a molecular wire. J Mol Biol *294*, 477-485.

Aich, P., Skinner, R. J., Wettig, S. D., Steer, R. P. and Lee, J. S. (2002). Long range molecular wire behaviour in a metal complex of DNA. J Biomol Struct Dyn *20*, 93-98.

Ankudinov, A. L., Ravel, B., J.J., R. and Conradson, S. D. (1998). Real Space Multiple Scattering Calculation of XANES. Phys. Rev. B *58*, 7565.

Aoki, K. (1975). Crystallographic Studies of Interaction between Nucleotides and Metal Ions. Crystal Structures of the 1:1 Complexes of Cobalt and Nickel with Inosine 5'-Phosphate. Bull. Chem. Soc. Japan *48*, 1260-1271.

Aoki, K. (1976). The crystal and molecular structure of the polymeric complex of zinc(II) withcytosine 5'-phosphate: metal bonding to both N(3) and phosphate. Biochim Biophys Acta *447*, 379-381.

Attwood, D. (2000). Soft X-rays and Extreme Ultraviolet Radiation: Principles and Applications. (New York: Cambridge University Press).

Auffinger, P. and Westhof, E. (2001). Water and ion binding around r(UpA)12 and d(TpA)12 oligomers--comparison with RNA and DNA (CpG)12 duplexes. J Mol Biol *305*, 1057-1072.

Avery, O. T., MacLeod, C.M. & McCarty, M. (1944). Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types. Induction of Transformation by Desoxyribonucleic Acid Fraction Isolated from Pneumococcal Type III. J. Exp. Med. *79*, 137-159.

Berman, H. M., Olson, W. K., Beveridge, D. L., Westbrook, J., Gelbin, A., Demeny, T., Hsieh, S. H., Srinivasan, A. R. and Schneider, B. (1992). The nucleic acid database. A comprehensive relational database of three-dimensional structures of nucleic acids. Biophys J *63*, 751-759.

Brennan, R. G., Westhof, E. and Sundaralingam, M. (1986). Structure of a Z-DNA with two different backbone chain conformations. Stabilization of the decadeoxyoligonucleotide d(CGTACGTACG) by [Co(NH3)6]3+ binding to the guanine. J Biomol Struct Dyn *3*, 649-665.

Brown, T. and Brown, D. J. (1992). Purification of synthetic DNA. Methods Enzymol *211*, 20-35.

Brunger, A. T., Adams, P. D., Clore, G. M., Delano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, N., Pannu, N. S.*, et al.* (1998). Crystallography and NMR System (CNS): A new software system for macromolecular structure determination. Acta Cryst D *54*, 905-921.

Bunker, G., Dimakis, N. and Khelashvili, G. (2005). New methods for EXAFS analysis in structural genomics. J Synchrotron Radiat *12*, 53-56.

Cartwright, B. A., Goodgame, D. M., Jeeves, I. and Skapski, A. C. (1977). X-ray structure of a compound of cobalt with uridine 5'-monophosphate. Evidence for metal-phosphate bonding only. Biochim Biophys Acta *477*, 195-198.

Caruthers, M. H., Beaton, G., Wu, J. V. and Wiesler, W. (1992). Chemical synthesis of deoxyoligonucleotides and deoxyoligonucleotide analogs. Methods Enzymol *211*, 3-20.

CCP4 (1994). The CCP4 suite: programs for protein crystallography. Acta Cryst D Biol Crystallogr *50*, 760-763.

Chandrasekaran, R. and Radha, A. (1992). Structure of poly d(A).poly d(T). J Biomol Struct Dyn *10*, 153-168.

Chargaff, E., Lipschitz, R., Green, C. and Hodes, M. E. (1951). The Composition of the Deoxyribonucleic Acid of Salmon Sperm. J. Biol. Chem. *192*, 223-230.

Clark, G. R. and Orbell, J. D. (1975). Transmition-Metal-Nucleotide Complexes. X-Ray Crystal and Molecular Structures of Cobalt(II) and Cadmium(II) Complexes of Cytosine 5'-Monophosphate, [Co(CMP)(H20)] and [Cd(CMP)(H20)]•H20. J. Chem. Soc. Chem. Commun. *697-698*.

Cotton, F. A. and Wilkinson, G. (1980). Advanced Inorganic Chemistry: A Comprehensive Text, 4th edn (New York: John Wiley & Sons).

De Meester, P., Goodgame, D. M., Jones, T. J. and Skapski, A. C. (1974a). The Polymeric Structure of Zinc Inosine 5'-Monophosphate: X-Ray Evidence for Metal Binding to Both N(7) and Phosphate. Biochim. Biophys. Acta *353*, 392-394.

De Meester, P., Goodgame, D. M. L., Jones, T. J. and Skapski, A. C. (1974b). Mise end Evidence par Diffraction X de la Liaison Metal-N (7) dans Deux Complexes Hydrates du Cobalt avec la Guanosine 5'-monophosphate et L'inosine 5'-monophosphate. C. R. Acad. Sc. Paris. serie C *279*, 667-669.

Dickerson, R. E. (1992). DNA structure from A to Z. Methods Enzymol *211*, 67-111.

Dickerson, R. E., Bansal, M., Calladine, C. R., Diekmann, S., Hunter, W. N., Kennard, O., von Kitzing, E., Lavery, R., Nelson, H. C. M., Olson, W. K.*, et al.* (1989). Definitions and nomenclature of nucleic acid structure components. Nucleic Acids Res *17*, 1797-1803.

Dickerson, R. E., Drew, H. R., Conner, B. N., Wing, R. M., Fratini, A. V. and Kopka, M. L. (1982). The anatomy of A-, B-, and Z-DNA. Science *216*, 475-485.

Dock-Bregeon, A. C., Moras, D. and Giege, R. (1999). Nucleic Acids and their Complexes, In Crystallization of Nucleic Acids and Proteins, A. Ducruix, and R. Giege, eds. (New York: Oxford University Press), pp. 209-243.

Dohm, J. A., Hsu, M. H., Hwu, J. R., Huang, R. C., Moudrianakis, E. N., Lattman, E. E. and Gittis, A. G. (2005). Influence of ions, hydration, and the transcriptional inhibitor P4N on the conformations of the Sp1 binding site. J Mol Biol *349*, 731-744.

Donohue, J. and Trueblood, K. N. (1960). Base Pairing in DNA. J. Mol. Biol *2*, 363-371.

Drenth, J. (1994). Principles of Protein X-Ray Crystallography (New York: Springer-Verlag Inc.).

Drew, H. R., Wing, R. M., Takano, T., Broka, C., Tanaka, S., Itakura, K. and Dickerson, R. E. (1981). Structure of a B-DNA Dodecamer. Conformation and Dynamics. Proc Natl Acad Sci U S A *78*, 2179-2183.

Egli, M. (2004). Nucleic acid crystallography: current progress. Curr Opin Chem Biol *8*, 580-591.

Egli, M., Minasov, G., Teplova, M., Kumar, R. and Wengel, J. (2001). X-ray crystal structure of a locked nucleic acid (LNA) duplex composed of a palindromic 10-mer DNA strand containing one LNA thymine monomer. J. Chem. Soc., Chem. Commun., 651-652.

Eichhorn, G. L. and Shin, Y. A. (1968). Interaction of metal ions with polynucleotides and related compounds. XII. The relative effect of various metal ions on DNA helicity. J Am Chem Soc *90*, 7323-7328.

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

Franklin, R. E. and Gosling, R. G. (1953). Molecular Configuration in Sodium Thymonucleate. Nature *171*, 738-741.

Gao, Y. G., Sriram, M. and Wang, A. H. (1993). Crystallographic studies of metal ion-DNA interactions: different binding modes of cobalt(II), copper(II) and barium(II) to N7 of guanines in Z-DNA and a drug-DNA complex. Nucleic Acids Res *21*, 4093-4101.

Gellert, R. W., Shiba, J. K. and Bau, R. (1979). X-ray crystal and molecular structures of the Ni(II) and Co(II) complexes of 2'-deoxyguanosine-5'-monophosphate. Biochem Biophys Res Commun *88*, 1449-1453.

Ghosh, A. and Bansal, M. (2003). A glossary of DNA structures from A to Z. Acta Crystallogr D Biol Crystallogr *59*, 620-626.

Giegé, R. and Ducruix, A. (1999). An Introduction to the Crystallogenesis of Biological Macromolecules, In Crystallization of Nucleic Acids and Proteins, A. Ducruix, and R. Giege, eds. (New York: Oxford University Press), pp. 1-16.

Hahn, M. and Heinemann, U. (1993). DNA helix structure and refinement algorithm: comparison of models for d(CCAGGCm5CTGG) derived from NUCLSQ, TNT and X-PLOR. Acta Crystallogr D Biol Crystallogr *49*, 468-477.

Hall, L. D. (1963). Conformations of Some Ribofuranosides. Chem. Ind., 950-951.

Hartzell, B. and McCord, B. (2005). Effect of divalent metal ions on DNA studied by capillary electrophoresis. Electrophoresis *26*, 1046-1056.

Haschemeyer, A. E. V. and Rich, A. (1967). Nucleoside Conformations: An Analysis of Steric Barriers to Rotation about the Glycosidic Bond. J. Mol. Biol *27*, 369-384.

Huheey, J. E., Keiter, E. A. and Keiter, R. L. (1993). Inorganic Chemistry: Principles of Structure and Reactivity, 4th edn (New York: HarperCollins College Publishers).

Hunter, C. A. (1993). Sequence-Dependent DNA Structure. The Role of Base Stacking Interactions. J. Mol. Biol. *230*, 1025-1054.

Hunter, C. A., Lawson, K. R., Perkins, J. and Urch, C. J. (2001). Aromatic Interactions. J. Chem. Soc. Perkins Trans. *2*, 651-669.

Kleveckis, C. and Grisham, C. M. (1996). Phosphate-Metal Ion Interactions of Nucleotides and Polynucleotides, In Metal Ions in Biological Systems, A. Sigel, and H. Sigel, eds. (New York: Marcel Dekker, Inc.), pp. 1-28.

Korolev, N., Lyubartsev, A. P., Nordenskiold, L. and Laaksonen, A. (2001). Spermine: an "invisible" component in the crystals of B-DNA. A grand canonical Monte Carlo and molecular dynamics simulation study. J Mol Biol *308*, 907-917.

Kumar, V. D., Harrison, R. W., Andrews, L. C. and Weber, I. T. (1992). Crystal structure at 1.5-A resolution of d(CGCICICG), an octanucleotide containing inosine, and its comparison with d(CGCG) and d(CGCGCG) structures. Biochemistry *31*, 1541-1550.

Labiuk, S. L., Delbaere, L. T. and Lee, J. S. (2003). Cobalt(II), nickel(II) and zinc(II) do not bind to intra-helical N(7) guanine positions in the B-form crystal structure of d(GGCGCC). J Biol Inorg Chem *8*, 715-720.

Lee, J. S., Latimer, L. J. and Reid, R. S. (1993). A cooperative conformational change in duplex DNA induced by Zn2+ and other divalent metal ions. Biochem Cell Biol *71*, 162-168.

Levene, P. A. and Tipson, R. S. (1935). The Ring Structure of Thymidine. J. Mol. Biol. *27*, 87-106.

Li, C. Z., Long, Y. T., Kraatz, H. B. and Lee, J. S. (2003). Electrochemical Investigation of M-DNA Self-Assembled Monolayers of Gold Electrodes. J. Phys. Chem. B *107*, 2291-2296.

Long, Y. T., Li, C. Z., Kraatz, H. B. and Lee, J. S. (2003). AC impedance spectroscopy of native DNA and M-DNA. Biophys J *84*, 3218-3225.

Lu, X. J. (2001). A Standard Reference Frame for the Description of Nucleic Acid Base-pair Geometry. http://rutchem.rutgers.edu/~olson/Tsukuba/.

McCarty, M., Avery, O.T. (1946). Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types. II. Effect of Desoxyribonuclease on the Biological Activity of the Transforming Substance. J. Exp. Med. *83*, 89-96.

McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C. and Read, R. J. (2005). Likelihood-enhanced fast translation functions. Acta Cryst *D61*, 458-464.

Mignon, P., Loverix, S., Steyaert, J. and Geerlings, P. (2005). Influence of the pi-pi interaction on the hydrogen bonding capacity of stacked DNA/RNA bases. Nucleic Acids Res *33*, 1779-1789.

Minasov, G., Tereshko, V. and Egli, M. (1999). Atomic-resolution crystal structures of B-DNA reveal specific influences of divalent metal ions on conformation and packing. J Mol Biol *291*, 83-99.

Moreno-Herrero, F., Herrero, P., Colchero, J., Gomez-Navarro, C., Gomez-Herrero, J. and Baro, A. M. (2003). Topographic Characterization and Electrostatic Response of M-DNA Studied by Atomic Force Microscopy. Nanotechnology *14*, 128-133.

Newville, M. (2004). Fundamentals of EXAFS, In Consortium of Advanced Radiation Sources (Chicago, IL: University of Chicago).

Olson, W. K., Bansal, M., Burley, S. K., Dickerson, R. E., Gerstein, M., Harvey, S. C., Heinemann, U., Lu, X. J., Neidle, S., Shakked, Z., *et al.* (2001). A standard reference frame for the description of nucleic acid base-pair geometry. J Mol Biol *313*, 229-237.

Parkinson, G., Vojtechovsky, J., Clowney, L., Brunger, A. T. and Berman, H. M. (1996). New parameters for the refinement of nucleic acid-containing structures. Acta Crystallogr D Biol Crystallogr *52*, 57-64.

Pearson, R. G. (1963). Hard and Soft Acids and Bases. J. Am. Chem. Soc. *85*, 3533-3539.

Pezzano, H. and Podo, F. (1980). Structure of Binary Complexes of Mono- and Polynucleotides with Metal Ions of the First Transition Group. Chem. Rev. *80*, 365-401.

Poojary, M. D. and Manohar, H. (1986). Interactions of Metal Ions with 2'-Deoxyribonucleotides. Crystal and Molecular Structure of Cobalt(II) Complex with 2'-Deoxyinosine 5'-Monophosphate. J. Chem. Soc. Dalton Trans., 309-312.

Portugal, F. H. and Cohen, J. S. (1977). *A Century of DNA* (Cambridge, Massachusetts and London, England: MIT Press).

Rakitin, A., Aich, P., Papadopoulos, C., Kobzar, Y., Vedeneev, A. S., Lee, J. S. and Xu, J. M. (2001). Metallic conduction through engineered DNA: DNA nanoelectronic building blocks. Phys Rev Lett *86*, 3670-3673.

Ravel, B. (2001). ATOMS: crystallography for the X-ray absorption spectroscopist. J Synchrotron Radiat *8*, 314-316.

Ravel, B. and Newville, M. (2005). ATHENA, ARTEMIS, HEPHAESTUS: data analysis for X-ray absorption spectroscopy using IFEFFIT. J Synchrotron Radiat *12*, 537-541.

Rehr, J. J. and Albers, R. C. (2000). Theoretical approaches to X-ray absorption fine structure. Rev. Mod. Phys. *72*, 621 - 652.

Reinert, F. and Hufner, S. (2005). Photoemission spectroscopy: from early days to recent applications. New J Phys *7*.

Rhodes, G. (1993). Crystallography Made Crystal Clear (San Diego: Academic Press Inc.).

Robertson, J. M. and Woodward, I. (1937). An X-ray study of the phthalocyanines. Part III. Quantitative structure determination of nickel phthalocyanine. J. Chem. Soc., 219 - 230.

Roussel, A. and Cambillau, C. (1992). TURBO-FRODO. Biographics, Marseilles, France.

Rozenberg, H., Rabinovich, D., Frolow, F., Hegde, R. S. and Shakked, Z. (1998). Structural code for DNA recognition revealed in crystal structures of papillomavirus E2-DNA targets. Proc Natl Acad Sci U S A *95*, 15194-15199.

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

Scott, W. G., Finch, J. T., Grenfell, R., Fogg, J., Smith, T., Gait, M. J. and Klug, A. (1995). Rapid crystallization of chemically synthesized hammerhead RNAs using a double screening procedure. J Mol Biol *250*, 327-332.

Seeman, N. C. (2003). DNA in a Material World. Nature *421*, 427-431.

Shui, X., McFail-Isom, L., Hu, G. G. and Williams, L. D. (1998). The B-DNA dodecamer at high resolution reveals a spine of water on sodium. Biochemistry *37*, 8341-8355.

Soler-Lopez, M., Malinina, L., Tereshko, V., Zarytova, V. and Subirana, J. A. (2002). Interaction of zinc ions with d(CGCAATTGCG) in a 2.9 A resolution X-ray structure. J Biol Inorg Chem *7*, 533-538.

Storoni, L. C., McCoy, A. J. and Read, R. J. (2005). Likelihood-enhanced fast rotation functions. Acta Cryst *D60*, 432-438.

Stura, E. A. and Gleichmann, T. (1999). Soading Techniques, In Crystallization of Nucleic Acids and Proteins, A. Ducruix, and R. Giegé, eds. (New York: Oxford University Press), pp. 366.

Subirana, J. A. and Soler-Lopez, M. (2003). Cations as hydrogen bond donors: a view of electrostatic interactions in DNA. Annu Rev Biophys Biomol Struct *32*, 27-45.

Teo, B. K. (1980). Chemical applications of extended X-ray absorption fine structure spectroscopy. Acc. Chem. Res., 412 - 419.

The International Tables of Crystallography. (1968). International Tables for X-ray Crystallography, Volume I, Symmetry Groups (Edited by Henry, N.F.M. and Lonsdale, K.), The Kynock Press, Birmingham, England.

Timsit, Y. and Moras, D. (1992). Crystallization of DNA. Methods Enzymol *211*, 409-429.

Valls, N., Wright, G., Steiner, R. A., Murshudov, G. N. and Subirana, J. A. (2004). DNA variability in five crystal structures of d(CGCAATTGCG). Acta Crystallogr D Biol Crystallogr *60*, 680-685.

Watson, J. D. and Crick, F. H. C. (1953). A Structure for Deoxyribose Nucleic Acid. Nature *171*, 737-738.

Wettig, S. D., Li, C. Z., Long, Y. T., Kraatz, H. B. and Lee, J. S. (2003). M-DNA: a self-assembling molecular wire for nanoelectronics and biosensing. Anal Sci *19*, 23-26.

Wilkins, M. H. F., Stokes, A. R. and Wilson, H. R. (1953). Molecular Structure of Deoxypentose Nucleic Acids. Nature *171*, 738-740.

Wood, D. O., Dinsmore, M. J., Bare, G. A. and Lee, J. S. (2002). M-DNA is stabilised in G*C tracts or by incorporation of 5-fluorouracil. Nucleic Acids Res *30*, 2244-2250.

Wood, D. O. and Lee, J. S. (2005). Investigation of pH-dependent DNA-metal ion interactions by surface plasmon resonance. J Inorg Biochem *99*, 566-574.

Woods, K. K., Maehigashi, T., Howerton, S. B., Sines, C. C., Tannenbaum, S. and Williams, L. D. (2004). High-resolution structure of an extended A-tract: [d(CGCAAATTTGCG)]2. J Am Chem Soc *126*, 15330-15331.

Appendix A –EXAFS Analysis Input Files and Math Expressions

Table A1: M-DNA theoretical input file created in Atoms 3.0.

space = P 1
a = 68.990 b = 68.990 c = 56.180
alpha = 90.0 beta = 90.0 gamma = 90.0
core = Ni100 edge = K rmax = 7.0
shift 0.00000 0.00000 0.00000

atoms

! elem	x	y	z	tag	occ.
H	0.00406	-0.11477	-0.02447	G1H5T	1.00000
O	0.00206	-0.10993	-0.04021	G1O5'	1.00000
C	0.02102	-0.11119	-0.05096	G1C5'	1.00000
C	0.03866	-0.10335	-0.03690	G1C4'	1.00000
O	0.03672	-0.08293	-0.03282	G1O4'	1.00000
C	0.03296	-0.08006	-0.00806	G1C1'	1.00000
N	0.01581	-0.06836	-0.00847	G1N9	1.00000
C	0.01549	-0.04938	-0.00625	G1C4	1.00000
N	0.03103	-0.03838	-0.00283	G1N3	1.00000
C	0.02640	-0.02051	-0.00073	G1C2	1.00000
N	0.03995	-0.00831	0.00271	G1N2	1.00000
H	0.05392	-0.01174	-0.00308	G1H21	1.00000
H	0.03750	0.00519	0.01120	G1H22	1.00000
N	0.00636	-0.01405	-0.00212	G1N1	1.00000
O	0.04757	0.02154	-0.01687	G1OZ	1.00000
H	0.06114	0.02121	-0.02277	G1HZ	1.00000
C	-0.00957	-0.02664	-0.00582	G1C6	1.00000
O	-0.02695	-0.02090	-0.00708	G1O6	1.00000
C	-0.00387	-0.04517	-0.00787	G1C5	1.00000
N	-0.01538	-0.06089	-0.01098	G1N7	1.00000
C	-0.00235	-0.07398	-0.01112	G1C8	1.00000
C	0.02761	-0.09984	0.00226	G1C2'	1.00000
C	0.04014	-0.11300	-0.01232	G1C3'	1.00000
O	0.05979	-0.11347	-0.00395	G1O3'	1.00000
P	0.06681	-0.11245	0.02350	C2P	1.00000
O	0.08622	-0.12115	0.02577	C2O1P	1.00000
O	0.05095	-0.11883	0.03952	C2O2P	1.00000
O	0.06894	-0.08965	0.02510	C2O5'	1.00000
C	0.08290	-0.07916	0.01072	C2C5'	1.00000
C	0.09269	-0.06184	0.02312	C2C4'	1.00000
O	0.07855	-0.04769	0.02828	C2O4'	1.00000
C	0.07231	-0.05544	0.05025	C2C1'	1.00000
N	0.05265	-0.05208	0.05157	C2N1	1.00000
C	0.03961	-0.06718	0.05164	C2C6	1.00000

C	0.04709	-0.03356	0.05271	C2C2	1.00000
O	0.06023	-0.02016	0.05258	C2O2	1.00000
N	0.02853	-0.02995	0.05392	C2N3	1.00000
C	0.01536	-0.04398	0.05404	C2C4	1.00000
N	-0.00252	-0.03940	0.05530	C2N4	1.00000
H	-0.01274	-0.04979	0.05659	C2H41	1.00000
H	-0.00638	-0.02521	0.05504	C2H42	1.00000
C	0.02066	-0.06349	0.05287	C2C5	1.00000
C	0.08262	-0.06831	0.06454	C2C2'	1.00000
C	0.10029	-0.06807	0.04799	C2C3'	1.00000
O	0.11584	-0.05617	0.05653	C2O3'	1.00000
P	-0.08597	0.09845	0.09630	G23P	1.00000
O	-0.08811	0.11964	0.09397	G23O1P	1.00000
O	-0.09741	0.08562	0.08038	G23O2P	1.00000
O	-0.06347	0.09275	0.09464	G23O5'	1.00000
C	-0.04928	0.10271	0.10899	G23C5'	1.00000
C	-0.02982	0.10632	0.09649	G23C4'	1.00000
O	-0.01887	0.08885	0.09277	G23O4'	1.00000
C	-0.02050	0.08400	0.06808	G23C1'	1.00000
N	-0.02742	0.06443	0.06833	G23N9	1.00000
C	-0.01660	0.04922	0.06627	G23C4	1.00000
N	0.00233	0.04962	0.06283	G23N3	1.00000
C	0.00922	0.03238	0.06023	G23C2	1.00000
N	0.02755	0.03113	0.05667	G23N2	1.00000
H	0.03640	0.04150	0.06326	G23H21	1.00000
H	0.03319	0.01968	0.04717	G23H22	1.00000
N	-0.00299	0.01531	0.06118	G23N1	1.00000
O	0.05092	0.00980	0.07091	G23OZ	1.00000
H	0.06062	0.02090	0.07451	G23HZ	1.00000
C	-0.02326	0.01593	0.06495	G23C6	1.00000
O	-0.03393	0.00133	0.06591	G23O6	1.00000
C	-0.02969	0.03435	0.06753	G23C5	1.00000
N	-0.04822	0.03986	0.07031	G23N7	1.00000
C	-0.04543	0.05811	0.07059	G23C8	1.00000
C	-0.03563	0.09690	0.05701	G23C2'	1.00000
C	-0.03343	0.11499	0.07172	G23C3'	1.00000
O	-0.01748	0.12674	0.06369	G23O3'	1.00000
P	-0.01154	0.13021	0.03617	C24P	1.00000
O	-0.00120	0.14882	0.03396	C24O1P	1.00000
O	-0.02829	0.12603	0.02045	C24O2P	1.00000
O	0.00394	0.11305	0.03380	C24O5'	1.00000
C	0.02131	0.11255	0.04836	C24C5'	1.00000
C	0.03938	0.10428	0.03617	C24C4'	1.00000
O	0.03631	0.08461	0.03117	C24O4'	1.00000
C	0.02684	0.08661	0.00899	C24C1'	1.00000
N	0.01310	0.07243	0.00815	C24N1	1.00000
C	-0.00661	0.07732	0.00803	C24C6	1.00000
C	0.01944	0.05425	0.00744	C24C2	1.00000
O	0.03750	0.05146	0.00767	C24O2	1.00000
N	0.00612	0.04011	0.00657	C24N3	1.00000
C	-0.01318	0.04415	0.00641	C24C4	1.00000
N	-0.02539	0.03027	0.00561	C24N4	1.00000
H	-0.03979	0.03324	0.00383	C24H41	1.00000

H	-0.02077	0.01644	0.00673	C24H42	1.00000
C	-0.02013	0.06340	0.00717	C24C5	1.00000
C	0.03040	0.10280	-0.00675	C24C2'	1.00000
C	0.04219	0.11384	0.01129	C24C3'	1.00000
O	0.06160	0.11481	0.00271	C24O3'	1.00000
H	0.06927	0.10393	0.00756	C24H3T	1.00000
Ni	0.01889	0.01283	-0.00733	Ni100	1.00000
Ni	0.02303	0.00075	0.05831	Ni903	1.00000

Table A2: Ni^{2+}-N7 and Ni^{2+}-OP theoretical input file created in Atoms 3.0.

Nickel – N7

```
space = P 1
a = 3.660                    b = 3.660                    c = 3.660
alpha = 90.0                 beta = 90.0                  gamma = 90.0
core = Ni1                   edge = K                     rmax =  6.0
shift   0.00000   0.00000   0.00000

atoms
! elem     x          y          z          tag        occ.
  Ni      0.00000    0.00000    0.00000    NiN7       1.00000
  N       0.00000    0.00000    0.50000    N7         1.00000
```

Nickel - OP

```
space = P 1
a = 4.720                    b = 4.720                    c = 4.720
alpha = 90.0                 beta = 90.0                  gamma = 90.0
core = Ni1                   edge = K                     rmax =  6.0
shift   0.00000   0.00000   0.00000

atoms
! elem     x          y          z          tag        occ.
  Ni      0.00000    0.00000    0.00000    NiN7       1.00000
  O       0.00000    0.00000    0.50000    N7         1.00000
```

Table A3: NiPC theoretical input file created in Atoms 3.0.

space = P 21/c
a = 19.90 b = 4.710 c = 14.90
alpha = 90.0 beta = 121.90 gamma = 90.0
core = Ni edge = K rmax = 6.0
shift 0.00000 0.00000 0.00000

atoms

! elem	x	y	z	tag	occ.
Ni	0.00000	0.00000	0.00000	Ni	1.00000
N	0.09690	0.19960	0.06850	N9	1.00000
C	0.12620	0.40130	0.02950	C10	1.00000
N	0.08690	0.49680	-0.07250	N11	1.00000
C	0.01100	0.38430	-0.14500	C12	1.00000
N	-0.03040	0.18050	-0.12420	N13	1.00000
C	-0.10100	0.15290	-0.22420	C14	1.00000
C	-0.14190	0.58170	-0.46850	CH15	1.00000
C	-0.06700	0.73250	-0.41410	CH16	1.00000
C	-0.01260	0.67940	-0.30670	CH17	1.00000
C	-0.03140	0.48190	-0.25370	C18	1.00000
C	0.34320	0.65390	0.28790	CH1	1.00000
C	-0.10520	0.33550	-0.30740	C19	1.00000
C	-0.16070	0.38430	-0.41480	CH20	1.00000
C	0.31960	0.78980	0.19130	CH2	1.00000
C	0.25710	0.71970	0.10130	CH3	1.00000
C	0.20630	0.51800	0.10670	C4	1.00000
C	0.23090	0.38430	0.20200	C5	1.00000
C	0.30680	0.45220	0.29260	CH6	1.00000
N	0.16700	0.02550	0.25300	N7	1.00000
C	0.16180	0.19110	0.17380	C8	1.00000

Table A4: Mathematical expressions used in Artemis.

Nickel – M-DNA

Math expressions

guess	amp	1
guess	enot_N	0
guess	drN1	0
guess	ssN1	0.003
guess	drOZ	0
guess	ssOZ	0.003
guess	drN3	0
guess	ssN3	0.003
guess	drN7	0
guess	ssN7	0.003
guess	drOP	0
guess	ssOP	0.003
define	sN1	abs(ssN1)
define	sOZ	abs(ssOZ)
define	sN3	abs(ssN3)
define	sN7	abs(ssN7)
define	sOP	abs(ssOP)

Path parameter math expressions

Path G1N1_Ni	N	:	1
	S02	:	amp
	delE0	:	enot_N
	delR	:	drN1
	sigma^2	:	ssN1
Path G1OZ_Ni	N	:	1
	S02	:	amp_1
	delE0	:	enot_N
	delR	:	drOZ
	sigma^2	:	ssOZ
Path C24N3_Ni	N	:	1
	S02	:	amp
	delE0	:	enot_N
	delR	:	drN1
	sigma^2	:	ssN1
Path N7_Ni	N	:	1
	S02	:	amp_1

	delE0	:	enot_N
	delR	:	drN7
	sigma^2	:	ssN7
Path OP_Ni	N	:	2
	S02	:	amp
	delE0	:	enot_N
	delR	:	drOP
	sigma^2	:	ssOP

Nickel – B-DNA

Math expressions

guess	amp	1
guess	enot	0
guess	dr	0
guess	ss	0.003
guess	amp_1	1
guess	enot_1	0
guess	dr_1	0
guess	ss_1	0.003

Path parameter math expressions

Path N7_Ni	N	:	2
	S02	:	amp
	delE0	:	enot
	delR	:	dr
	sigma^2	:	ss
Path OP_Ni	N	:	2
	S02	:	amp_1
	delE0	:	enot_1
	delR	:	dr_1
	sigma^2	:	ss_1

NiPC

Math expressions

guess	amp	1
guess	enot_N	0
guess	dr	0
guess	ss	0.003
guess	ss3	0.003

guess	ss5	0.003
guess	enot_C	0
guess	dr2	0
guess	dr2b	0
guess	dr1	0

Path parameter math expressions

Path N13_Ni	N	:	2
	S02	:	amp
	delE0	:	enot_N
	delR	:	dr1
	sigma^2	:	ss
Path N9_Ni	N	:	2
	S02	:	amp
	delE0	:	enot_N
	delR	:	dr
	sigma^2	:	ss
Path C12_Ni	N	:	2
	S02	:	amp
	delE0	:	enot_C
	delR	:	dr2
	sigma^2	:	ss3
Path C14_Ni	N	:	2
	S02	:	amp
	delE0	:	enot_C
	delR	:	dr2
	sigma^2	:	ss3
Path C10_Ni	N	:	2
	S02	:	amp
	delE0	:	enot_C
	delR	:	dr2b
	sigma^2	:	ss5