An Investigation of DNA Stability and Sequence-dependent Deformability

by

Angela Vivian Toms

A Thesis
Submitted to the Faculty of Graduate Studies in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

Department of Chemistry
University of Manitoba
Winnipeg, Manitoba

THE UNIVERSITY OF MANITOBA

FACULTY OF GRADUATE STUDIES
*****
COPYRIGHT PERMISSION PAGE

AN INVESTIGATION OF DNA STABILITY AND SEQUENCE-DEPENDENT DEFORMABILITY

BY

ANGELA VIVIAN TOMS

A Thesis/Practicum submitted to the Faculty of Graduate Studies of The University

of Manitoba in partial fulfillment of the requirements of the degree

of

Doctor of Philosophy

Abstract

The primary focus of this dissertation is the importance of the structure of DNA in the protein-DNA recognition process. The research described herein centers on two distinct areas of investigation.

In the first series of studies, the stability of a C-C mismatch-containing DNA sequence was investigated in an effort to obtain diffraction quality crystals. Determining the structures of DNA sequences with C-C mismatches is important in order to understand their structure and how they are able to escape correction by repair systems. The effect of pH, ionic strength, cation type and polyamines on the stability of the mismatch containing sequence d(CGGCCGCCG) was investigated with UV and CD thermal melting studies.

In the second area of investigation, the structure of a DNA fragment was investigated through X-ray crystallography. The existence of distinct families of DNA-binding proteins, which employ related structural motifs for recognition, suggests the involvement of common sequence elements in DNA regulatory regions. One sequence that is represented frequently in functionally important sites involving protein-DNA interactions is GTG/CAC. In an attempt to further explore the sequence dependent effects observed in the structure of d(CGGTGG)/d(CCACCG), a DNA fragment from an E. coli regulatory region, the structure of the DNA dodecamer d(CGGTGGCCACCG)/d(CGGTGGCCACCG) was determined. The DNA dodecamer was co-crystallized in the presence of spermine and cobalt hexammine resulting in two structures which were solved to resolutions of 1.95 Å and 1.75 Å respectively. Analysis of the structures revealed a number of distortions from the ideal A-type conformation. No unusual distortions, however, were found in the GTG region.

CHAPTER 1 Introduction 9

CHAPTER 2 Experimental 50

CHAPTER 1 Introduction

CHAPTER 2 Experimental

APPENDIX A DNA Structure and Nomenclature

CHAPTER 1 Introduction

CHAPTER 2 Experimental

CHAPTER 3 Results and Discussion

Acknowledgements

I would like to express my gratitude to my supervisor, Professor Anthony Secco, for his patience and guidance throughout my graduate program. He has ensured that I have the skills necessary to excel in the scientific community. The high standards he works by serve to inspire.

I would like to thank the members of my advisory and/or examining committee, Dr. Frank Hruska, Dr. Lisa Rosenberg, Dr. Scott Kroeker, Dr. John Templeton and Dr. David Rose for their helpful input.

I am grateful to Dr. Stephen Evans for providing the lab with his molecular modelling program SETOR and for allowing me to accompany his group to the NSLS for data collection. I am also grateful to Dr. Leslie Tari for his advice and generous assistance with X-ray data collection at the University of Calgary.

I would also like to thank Brookhaven National Lab and the NSLS for allowing up to participate in the PX data collection program.

I gratefully acknowledge the Natural Sciences Research Council of Canada and the University of Manitoba for financial support throughout my graduate studies.

CHAPTER 1 *Introduction*

1.1 Background and Significance

The study of nucleic acid structure has grown into a vast field which continues to play a key role in furthering our understanding of DNA repair and regulation of gene expression. Nucleic acids are conformationally complex molecules. This complexity is sometimes masked by the simplicity of the classic DNA double helix. The dynamic nature of these molecules is shown by the readiness of oligonucleotides to be structurally responsive to changes in changes in hydration, counter-ions, or by ligand binding, and can be highly sequence dependent (Mazur, 1989; Robinson & Wang, 1996; Rouzina & Bloomfield, 1998; Y. Kim *et al.*, 1993). X-ray crystallography permits investigation at an atomic level of sequence specific effects and structural anomalies in a variety of environments.

1.1.1 The Importance of DNA Structure

Francis Crick summarized the relationships among DNA, RNA and protein in his scheme "the central dogma of molecular biology": DNA directs its own replication and its transcription to RNA which, in turn, directs its translation to proteins (Figure 1-1, Crick, 1970). At each step of replication, transcription and translation interactions between proteins and nucleic acids occur. While much insight has been gained in the last decade as to the nature of these interactions, the mechanism of the recognition processes is still not completely understood. In the following sections the importance of protein-DNA recognition is discussed as it relates to DNA mismatch repair and to the regulation of gene expression.

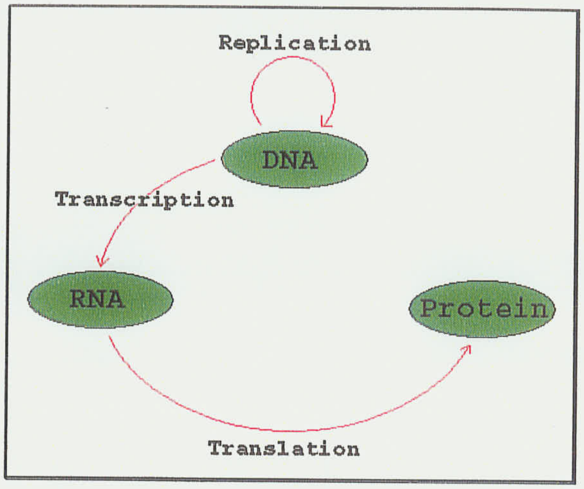

Figure 1-1 Central dogma of molecular biology.

1.1.2 Mismatch Repair

Nucelic acids utilize the specificity of the hydrogen-bonding of the bases to establish the fidelity of replication, transcription and translation. The base-pairing pattern in DNA is called Watson-Crick pairing. In the Watson-Crick base-pairing scheme, adenine pairs with thymine, and guanine with cytosine (Figure 1-2).

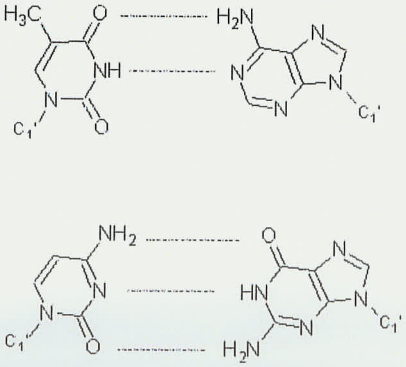

Figure 1-2 Watson-Crick base-pairing scheme.

Watson-Crick base-pairing for T:A (top) and C:G (bottom). Dashed lines depict hydrogen-bonds.

Maintaining the fidelity of the genetic code is crucial for cell survival. Consequently, the replication of DNA must be a carefully controlled process. However, the formation of non-Watson-Crick pairs does occur during DNA replication and recombination. These mistakes, if uncorrected, lead to mutations in the genetic code. The expressed proteins may be ineffective and their malfunction can lead to genetic disease or death of the organism. The correction of mistakes in the genetic code is carried out by a complex set of DNA proofreading and repair enzymes.

The efficiency of recognition and repair of the mismatched bases is dependent both on the type of mismatch and on the sequences flanking the mismatch site (Kramer *et al.*, 1984; Dohet *et al.*, 1985). If a mismatch adopts a geometry similar to a Watson-Crick base pair (W-C b.p.) then it will not be repaired efficiently, but if the helix is perturbed, then it is easily repaired (Hunter *et al.*, 1986). Out of the 8 possible mismatches, C-C mismatches are the least efficiently repaired. However, there is no simple "wobbling" of the C-C base pairs that yields a W-C b.p. hydrogen-bonding envelope. To date, very little structural information is available on the pairing possibilities of C-C mismatches in oligonucleotide anti-parallel duplexes.

1.1.2.1 C-C mismatches

C-C mismatches have been implicated in the disease known as fragile X syndrome (Chen *et al.*, 1995). Fragile X syndrome is a common genetic cause of mental retardation and is so named because of an aberration resulting in triplet expansion in a fragile gene on the X chromosome (Kremer *et al.*, 1991). The mechanism giving rise to this chromosomal abnormality is not yet known. However, current thinking is that during DNA replication, the gene-replicating machinery slips and makes multiple copies of small segments of DNA (Sutherland & Richards, 1994). The resulting tandem repeat can become expanded even further by slippage during chromosomal replication. The repeating segment in the fragile X site is CCG. The individual strands of $(CCG)_n$ can form hairpin loops where the DNA doubles back on itself to form a base-paired helix with every third base pair being a C-C mismatch.

Determining the structures of DNA sequences containing C-C mismatches is important in order to understand their structure and how they are able to escape repair. Work towards obtaining the first X-ray crystal structure of an anti-parallel C-C mismatch containing DNA fragement was carried out by a former student in this laboratory, Mei Chen. Difficulties were encountered in attempting to crystallize the dodecamer d(CGCT-GGCCACCG). UV and CD experimental results on the C-C mismatch-contaning dodecamer indicated that the sequence could adopt a double stranded dodecamer as well as folding back on itself in a hairpin structure (Chen, 1997). The conformational flexibility is likely due to the instability resulting from the incorporation of two C-C mismatches — C-C mismatches are the least thermodynamically stable of all the mismatches (Aboul-Ela *et al.*, 1985). To reduce the possibility of conformational flexibility and thus increase the likelihood of obtaining diffraction quality crystals we designed the 9-mer sequence, d(CGGCCGCCG), which contains a single C-C mismatch in the center of the duplex flanked on both sides by G-C base-pairs.

1.1.3 Regulation of Gene Expression

Most regulatory mechanisms used in bacteria obey the following rule: *a system is turned "on" when it is needed and "off" when it is not needed.* For example, in *E. coli* under conditions of active growth, only about 5% of the genome is actively transcribed at any given time; the remainder is either silent or transcribed at a very low rate. When growth conditions change, some active genes are turned off and some inactive genes are turned on. One of the most effective ways to control the expression of genes in response to environmental change is through a repressor or an activator protein which recognize changes in the concentration of signal molecules. Gene expression is then regulated by the complexation of the regulatory protein with a specific control region on the DNA. However, before specific binding can occur, the protein must locate and recognize its target.

When considering even a simple organism such as *E. coli,* the magnitude of this problem becomes apparent. The *E. coli* genome contains approximately 5 x 10^6 base pairs, which are sufficient to encode about 3000 genes. While the relative length of DNA is near infinite, there are only a few copies of each particular regulatory protein in the cell. How these proteins efficiently locate their relatively small target sites on the DNA remains uncertain. Three theories of searching mechanisms have been proposed (von Hippel & Berg, 1989; Bustamante *et al.*, 1999): (*i*) the sliding, (*ii*) the inter-segment transfer and (*iii*) intra-domain association and dissociation theories, all involving some degree of protein-sliding along the DNA. How then, does the sliding protein recognize its binding site?

Three factors implicated in the recognition process are: (*i*) electrostatic interactions between specific amino acids of the proteins and the phosphate backbone at the target site, (*ii*) hydrogen bonding between the amino acids of the proteins and the bases in the cognate sequence and (*iii*) sequence-dependent conformational inhomogeneities along the DNA helix that could decrease the diffusional mobility of the protein at the target sequence. The truth, likely, embraces each of these to some extent. The work presented in this thesis is an attempt to investigate the last factor, namely, structural inhomogeneity in the DNA helix.

One particular sequence found frequently in sites involving protein-DNA interactions is GTG/CAC. This dimeric trinucleotide appears in the consensus sequence of regulatory regions at a frequency greater than statistically predicted, suggesting a role in the overall regulatory process (Lu *et al.*, 1983). By looking at its structure within the context of a known protein-binding region it is possible to gain insight into how this trimer may function as a signal to searching proteins. The region chosen for structural investigation in this thesis comes from the regulatory region of the *gal* operon.

1.1.3.1 The *gal* Operon

The *gal* operon is a constitutive operon, meaning that it continues to express its genes at a significant level even when the operon is "off". The constitutive nature of the operon and complex control mechanisms are likely results of its dual role in cellular metabolism. It encodes enzymes responsible both for the catabolism of galactose in the absence of glucose and for the conversion of uridinediphosphoglucose to uridinediphosphogalactose (UDP-galactose) in the absence of exogenous galactose. (UDP-galactose is a precursor in *E. coli* cell wall synthesis and is required at all times.)

The three structural genes in the *gal* operon encode the enzymes galactokinase (galK), galactose-1-phosphate uridyltransferase (galT) and uridine diphosphogalactose 4'-

epimerase (galE). Overall, the concerted action of the enzymes convert galactose to glucose 1-phosphate, which is utilized as an energy source in *E. coli* bacteria (Figure 1-3).

The *gal* operon (Figure 1-4) is regulated by a small ensemble of proteins that utilize interdependent pathways of positive and negative transcriptional regulation. Transcription of this operon occurs at two partially overlapping, mutually exclusive promoters, P1 and P2 (Adhya & Miller, 1979), separated by only five base pairs. These two promoters are differentially affected by cAMP-CRP (cyclic adenosine monophosphate - cyclic AMP receptor protein complex). Binding of cAMP-CRP activates transcription from P1 while depressing transcription from P2 (Irani et al., 1989; Goodrich & McClure, 1992). In addition to regulation by cAMP-CRP, transcription of the *gal* operon is also subject to negative control mediated by the repressor protein GalR. GalR, a dimer in solution, binds to the two operator sequences, O_E and O_I, which flank the transcription start sites, thus blocking transcription from P1 and P2 (Choy & Adhya, 1992; Choy *et al.*, 1995). The activity of GalR is modulated by the binding of galactose or fucose, which releases GalR-mediated repression through an allosteric mechanism (Weickert & Adhya, 1993). The concerted effect of the various controlling mechanisms on expression is summarized in Table 1-1.

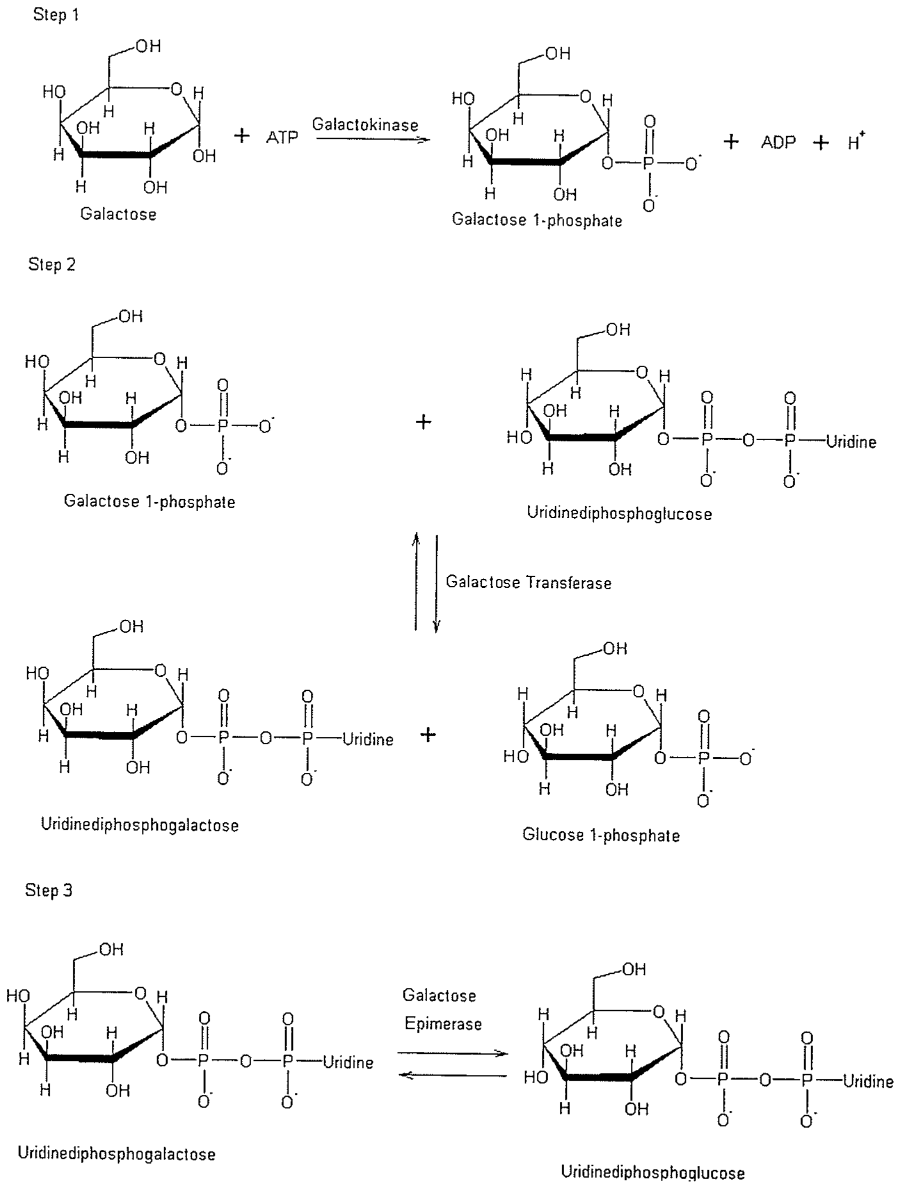

Figure 1-3 The galactose catabolism pathway.

Galactokinase, galactose transferase and galactose epimerase act to convert galactose to glucose 1-phosphate.

TABLE 1-1 **Summary of *gal* expression in response to sugar concentrations.**
Galactose concentration is the primary "switch" for transcription. Glucose modulates expression, to a lesser extent, by altering the level of cAMP.

Galactose	Glucose	Rate of Expression
Low	High	Low
Low	Low	Low
High	High	Moderate
High	Low	High

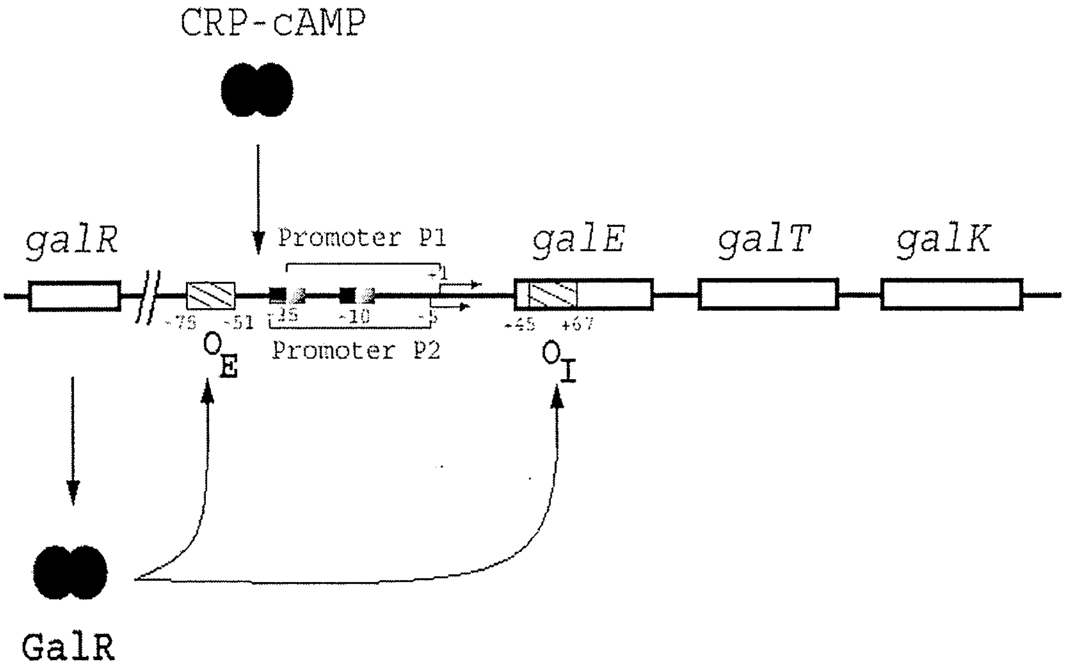

Figure 1-4 Schematic of *gal* operon.

Transcription of the *gal* genes is controlled by two overlapping tandem promoters, designated P1 and P2, which are five bases apart. The *gal* repressor GalR is encoded at some distance upstream. The dimeric GalR binds to two operator regions O_E and O_I flanking the *gal* promoters at the indicated positions. A CRP binding site is located immediately upstream of P1. Binding of CRP to this site activates transcription from P1 and represses transcription from P2.

1.1.3.2 Choice of DNA Sequence

A previous investigation of the structure of a short DNA fragment contained within the interior operator of the *gal* operon, d(CGGTGG)/d(CCACCG), showed significant anomalies at the GTG site (Tari & Secco, 1995). Despite the rationale for the observed A-T base-pair opening, the question remains as to whether the anomaly is inherent in GTG or the result of crystallization conditions, or perhaps, just a fortuitous choice of sequence. To test whether the anomaly will occur in a similar environment, the study of the self-complementary 12-mer CGGTGGCCACCG was undertaken. This 12 base pair (b.p.) sequence was chosen for two reasons. First, the hexamer alone crystallizes with two duplexes related by a 2-fold axis perpendicular to the helix axis such that one duplex stacks on the other to make a continuous, but not covalently linked, 12 b.p. helix. Thus, the expectation is that the 12-mer sequence could take advantage of this packing arrangement with little, if any, structural rearrangement. The second reason relates to ease of crystallization. Self-complementary sequences facilitate crystallization as no purification is required to achieve an exact 1:1 ratio of complementary strands. If an exact ratio is not achieved it can greatly interfere with crystallization.

Formation of poor quality crystals is often encountered with nucleic acids. The poor crystal quality can be thought of a resulting from the shape of the B-type double helix, which approximates that of a cylinder. In the crystal the "B-DNA cylinders" can adopt a side-by-side close hexagonal packing arrangement with the cylinders also stacking end-on. In this type of hexagonal lattice there is a high degree of rotational disorder associated with the molecules. By incorporating structural characteristics that favour the ordered assembly of the DNA molecules it is possible to increase the likelihood of obtaining high quality crystals. For instance, by incorporating overhanging base pairs into the design of the DNA sequence an interlock is created between the end-on stacked helices to prevent rotational disorder in this dimension. The second construct chosen for investigation, GGT-GGCCACCGC, contains the GTG trimer as well as two overhanging base pairs at the termini. The desired effect of the overhanging b.p.'s was to reduce the possibility of rotational disorder between the stacked helices and thus increase the likelihood of obtaining diffraction quality crystals.

1.2 Overview of Macromolecular X-ray Crystallography

1.2.1 The Need For X-rays

The most obvious way to determine the shape of objects around us is to look at them. If the objects are small it becomes necessary to use a microscope. However, there is a limit to how small an object can be and still be visible under a microscope — you cannot image objects that are smaller than the wavelength of light you are using. In order to obtain atomic resolution, it is necessary to use wavelengths that are comparable to the interatomic distances, *i.e.* X-rays. Unfortunately, it is not possible to build an X-ray microscope to look at molecules as X-ray lenses do not exist yet. However, an X-ray lens can be simulated mathematically on a computer. In the case of a microscope, light strikes the object and is diffracted in various directions. The lens then collects the diffracted rays and reassembles them with perfect phase fidelity to form an image. With X-rays, it is possible to detect the diffraction from the molecule, but the loss of all phase information makes it necessary to use a computer to reassemble the image. The image obtained in a crystallographic experiment is not really a picture of the atoms *per se*, but a map of the distribution of electrons in the molecule in the crystal.

1.2.2 X-ray Sources

X-rays are produced by accelerating a beam of electrons into a metal anode from which core electrons are ejected. The wavelengths of X-radiation produced (as a result of the cascade of adjacent shell electrons filling in the core shell vacancies) are dependent on the type of anode metal. The resulting radiation is filtered either with a thin metal foil or by using the intense low order diffraction from a graphite crystal to obtain a monochromatic beam. For most macromolecular work, the target is copper and the characteristic wavelength is 1.54 Å. An alternative source of X-rays is obtained when a beam of electrons is bent by a magnet. This is the principle behind synchrotron radiation sources which are capable of producing X-ray beams that are orders of magnitude more intense than those produced by a conventional lab-based rotating anode generator. The high intensity source greatly reduces data collection times and allows for smaller crystals to be used. A further advantage is that the X-ray spectrum is continuous from around 0.5 to 3.0 Å. The use of shorter wavelengths has been found to increase crystal lifetime as well as to decrease the noise due to scattering from air. The main drawback associated with synchrotron radiation is that few facilities exist, access is restricted, and it is often necessary to travel long distances with fragile crystals.

1.2.3 The Crystal Lattice and Unit Cell

A crystal is a regular, repeating array of atoms or molecule in three dimensions. It can be regarded as being a convolution of a molecular fragment and a three-dimensional periodic lattice (Figure 1-5). The lattice is defined by three axes and the angles between them. Along each axis direction there is a repeat distance, referred to as the unit cell translation or unit cell repeat. The unique portion of the crystal is defined as the unit cell. The basis vectors of the unit cell are denoted **a**, **b**, and **c**, and angles between them α, β, and γ (Figure 1-5). In crystals used in experiments the periodicity is never perfect. These imperfect crystals can be regarded as being composed of small mosaic blocks (Figure 1-5).

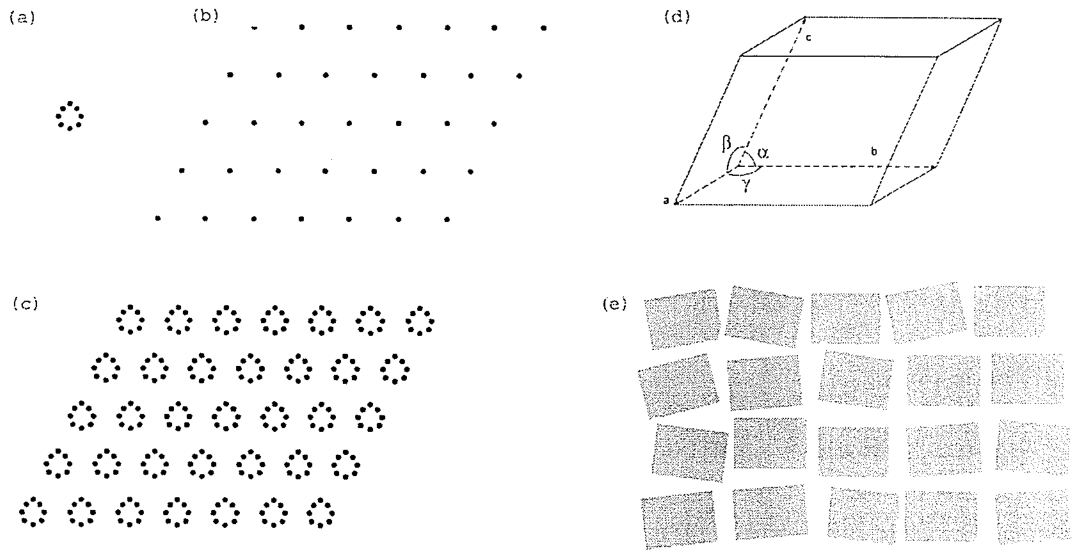

Figure 1-5 The crystal lattice, unit cell and mosaicity.

(a) Representation of a two dimensional molecule, (b) is a two dimensional lattice, and (c) is the convolution of the molecule with the lattice, *i.e.* a two dimensional crystal. A three dimensional crystal is simply the convolution of a molecule (or the unique portion thereof) and a three dimensional lattice. (d) The unique portion of the crystal, the unit cell. (e) An imperfect crystal composed of small mosaic blocks, which can be considered as optically independent fragments.

1.2.4 Diffraction Theory: Interaction of X-rays with a Crystal

The scattering of an X-ray beam by a crystal results from the interaction between the electric component of the beam and the electrons in the crystal. When a monochromatic X-ray beam strikes an electron, the electron starts to oscillate in the direction of the electric vector of the incident beam. In the case of elastic scattering, this oscillating electron acts as transmitter and radiates X-rays with the same frequency. The scattered radiation has the same wavelength as the incident radiation, and this is the radiation responsible for the interference effects in a diffraction experiment. The intensity of the radiation scattered by a free electron is independent of the incident wavelength. Electrons in an atom are bound by the nucleus and are, therefore, not free electrons. However, to a good approximation, they can be regarded as such if the frequency of the incident radiation ν is greater than the natural absorption frequencies, ν_n, at the edges of the scattering atom (*i.e.* the wavelength of the incident radiation is shorter than the absorption-edge wavelength). When $\nu \approx \nu_n$, the electrons in the atom can no longer be treated as free electrons, and the scattering factor becomes complex (the details and importance of this are discussed later). For most purposes the electrons in an atom can be treated as free electrons, as is the case for most atoms in a macromolecular structure. The ability to scatter X-rays is referred to as the "scattering amplitude of an atom", also known as the "atomic scattering factor", f. The scattering factor is dependent on the number of electrons and their positions within the electron cloud. Atomic scattering factors also decrease appreciably as a function of $(\sin \theta)/\lambda$ (Figure 1-6) due to interference effects from the scattering between the electrons surrounding the nucleus.

When monochromatic X-rays strike a collection of atoms in a molecule or crystal, the scattered radiation from the respective atoms forms an interference pattern called a diffraction pattern. X-ray scattering from a single molecule would be extremely weak and nearly impossible to detect. Thus, X-ray crystallography experiments are carried out on crystals containing extremely large numbers of molecules arranged such that the scattered waves add up in phase and raise the scattering amplitude to a measurable level.

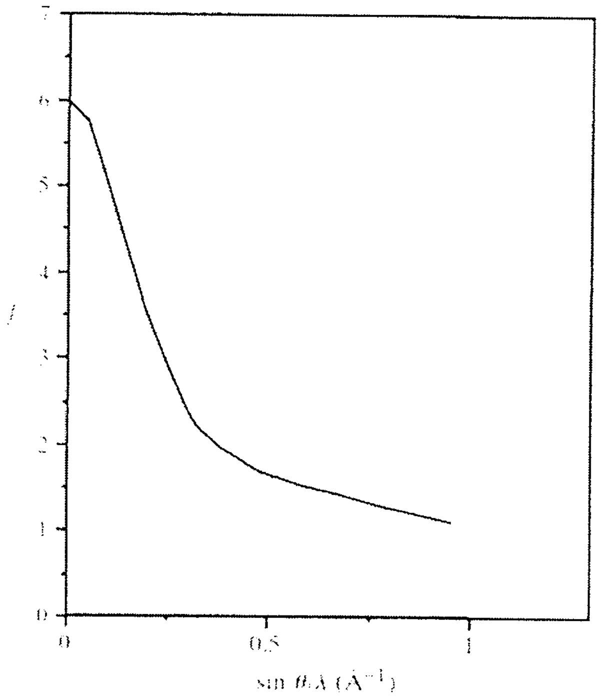

Figure 1-6 The atomic scattering factor as a function sinθ/λ.

The atomic scattering factor *f* for carbon as a function of sinθ/λ, expressed in units of the scattering by one electron. Reproduced from Drenth (1994).

The scattering from a unit cell depends on the type and arrangement of atoms within it. The observed diffraction pattern results from addition of the scattered waves from each of the atoms. The resultant wave scattered in one direction is the composite of the waves scattered by the atoms in that direction. For j atoms in the unit cell at positions r_j with respect to the origin, the amplitude of each wave is given by the atomic scattering factor, f_j. The phase of the wave, with respect to a wave scattered by hypothetical electrons at the origin of the unit cell, is $2\pi r_j \bullet \mathbf{S}$, where **S**, the scattering vector, is perpendicular to the reflecting plane (Figure 1-7). The resultant wave is given by the structure factor, F(S):

$$F(S) = \sum_{j}^{n} f_j \cdot \exp 2\pi i (r_j \cdot S) \qquad \text{(EQ 1-1)}$$

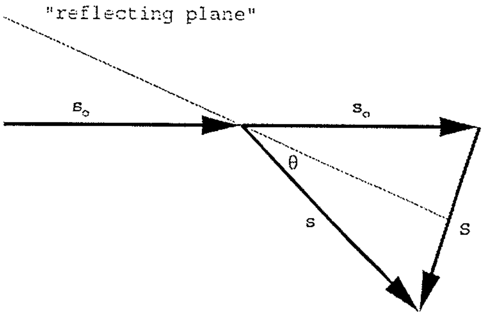

Figure 1-7 The scattering vector and reflecting plane.

The direction of the incident wave is indicated by $\mathbf{s_0}$ and that of the scattered wave as **s**. Both vectors are of length $1/\lambda$. The incident wave can be regarded as being reflected against the plane that makes equal angles with $\mathbf{s_0}$ and **s**, with θ as the reflecting angle. The vector S is perpendicular to this plane and $|S| = 2\sin\theta/\lambda$.

To calculate the scattering from the entire crystal, it becomes necessary to take into account the scattering from all of the unit cells in the crystal. Relative to the first unit cell, another unit cell has its origin at t x **a**, u x **b**, and v x **c**, where t, u, and v are whole numbers, and **a**, **b**, and **c** are the basis vectors of the unit cell. With respect to the first origin, scattering due to this cell is

$$F(S)\exp(2\pi i t\mathbf{a} \bullet S)\exp(2\pi i u\mathbf{b} \bullet S)\exp(2\pi i v\mathbf{c} \bullet S)$$

The waves scattered by a crystal is then the sum of the waves scattered by all unit cells. Assuming that the crystal has a very large number of unit cells (n_1 x n_2 x n_3), the wave scattered by the crystal is as follows:

$$K(S) = F(S) \sum_{t=0}^{n_1} \exp[2\pi it(a\cdot S)] \sum_{u=o}^{n_2} \exp[2\pi iu(b\cdot S)] \sum_{v=0}^{n_3} \exp[2\pi iv(c\cdot S)] \quad \text{(EQ 1-2)}$$

For an infinitely large crystal, the three summations over the exponential function in Equation 1-2 are zero unless the following conditions are met:

$$a \cdot S = h; \qquad b \cdot S = k; \qquad c \cdot S = l \quad \text{(EQ 1-3)}$$

where *h, k,* and *l* are whole numbers. These conditions are known as the Laue conditions.

Bragg's law is another way of looking at the conditions for diffraction. The diffraction spots can be considered as reflections, because the crystal can be thought of a being composed of thousands of mirrors that reflect the X-rays. These mirrors are called the Bragg planes. When light is reflected from a mirror, the angle of incidence is equal to the angle of reflection. The same is true of Bragg planes: when the angle of incidence is equal to the angle of reflection, the light rays hitting the plane in phase, exit in phase. Rays from one plane have identical path-lengths and rays reflected from different planes have different path-lengths. In order for the scattered radiation from multiple planes to be in phase, the difference in path-length must be equal to an integral number of wavelengths. The difference in path-length depends on the angle of incidence and the distance between the planes. The relationship is known as Bragg's law:

$$2d\sin\theta = \lambda \qquad \text{(EQ 1-4)}$$

where d is the distance between reflecting lattice planes, θ is the reflecting angle and λ is the wavelength (Figure 1-8).

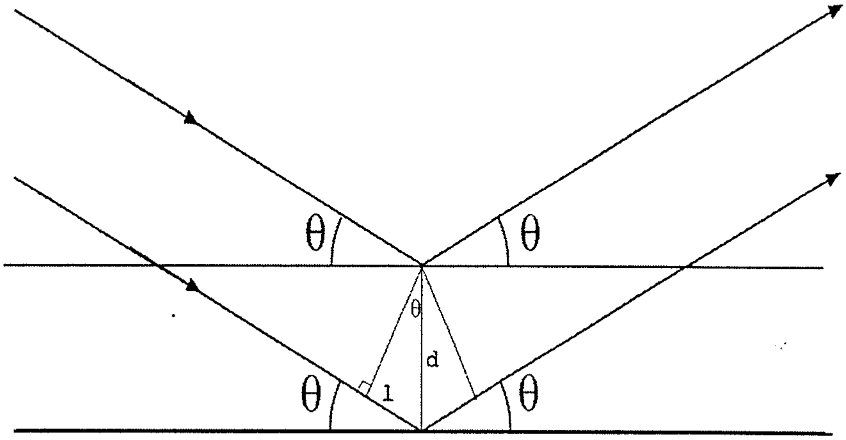

Figure 1-8 Bragg's law.

The difference in path-length between the rays reflected form the two planes is twice the distance *l*. Simple geometry tells us that the upper angle in the little triangle must be θ. The distance *l* is equal to d$\sin\theta$. For the two rays to be in phase, twice *l* must be equal to the wavelength.

In considering diffraction from a crystal, the planes that divide the unit-cell vectors **a**, **b**, and **c** into an equal number of parts are called lattice planes. If the lattice planes divide the **a** vector of the unit cell into *h* equal parts, the first index for this set of planes in *h*. The second index, *k,* is related to the division of **b** and the third index, *l*, to the division of **c**. These are called the indices of the reflections from the crystal, or Miller indices. Thus, the scattering by a crystal can be regarded as reflections by the lattice planes and the spacing between these planes, $d(hkl)$, is equal to $1/|S(hkl)|$. The structure factor for n atoms can now be written as

$$F(hkl) = \sum_{i}^{n} f_j \exp[2\pi i(hx + ky + lz)] \qquad \text{(EQ 1-5)}$$

or

$$F(hkl) = |F(hkl)| \cdot \exp i\alpha(hkl) \qquad \text{(EQ 1-6)}$$

where x, y, and z are the fractional coordinates along the lattice axes **a**, **b**, and **c** of each atom in the unit cell. |**F**(*hkl*)| is the structure factor amplitude and α(*hkl*) is the phase (Figure 1-9).

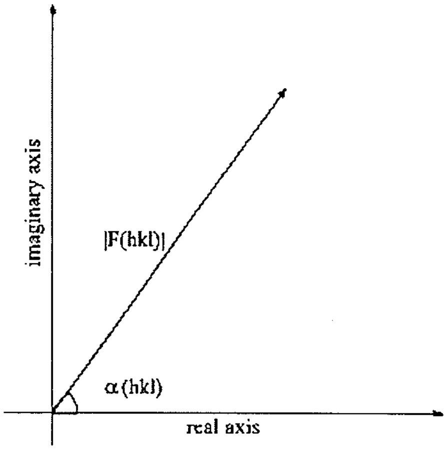

Figure 1-9 Argand diagram.
The composite scattered wave can be represented as a vector with amplitude |**F**(*hkl*)| and phase α(*hkl*).

Up to this point the size of the electron density cloud has been considered to be independent of temperature, which is true under most conditions. If this were to hold in the definition of the atomic scattering factors, the scattering of X-rays by a crystal would also be independent of temperature. However, the scattering of X-rays is not temperature independent. The atoms in the crystal vibrate about an equilibrium point, thus effectively smearing out the electron cloud and increasing the size of the atom. The greater the thermal vibrations, the weaker the scattered intensity becomes, especially at high scattering angles, due to destructive interference (*i.e.* a result of differing path-lengths of the scattered radiation). Thus a correction factor needs to be applied to the atomic scattering factors to account for this diminished scattering power. In the case of isotropic vibrations, the correction factor is given by

$$T(isotropic) = \exp-\left[\frac{(B\sin\theta^2)}{\lambda}\right] \quad \text{(EQ 1-7)}$$

where $B=8\pi^2\overline{u}^2$, where $\overline{u}^2$ is the root mean square amplitude of atomic vibration. The structure factor is then given by

$$F(hkl) = \sum_{j}^{n} f_j \exp-\left[\frac{(B\sin\theta^2)}{\lambda}\right] \exp[2\pi i(hx+ky+lz)] \quad \text{(EQ 1-8)}$$

As it is the distribution of electrons in the crystal that are observed in a crystallographic experiment, an expression relating the electron density to the diffraction pattern is required. The electron density distribution, $\rho(xyz)$, in a crystal is obtained by taking the inverse Fourier transform of the diffraction pattern as shown below

$$\rho(xyz) = \frac{1}{V} \cdot \sum_{h,k,l}^{\infty} |F(hkl)| \exp(i\alpha(hkl)) \exp[-2\pi i(hx+ky+lz)] \quad \text{(EQ 1-9)}$$

where V is the volume of the unit cell and |**F**(*hkl*)| is the structure factor amplitude and $\alpha(hkl)$ the phase, $2\pi(hx + ky + lz)$.

In a single crystal diffraction experiment the observed quantities, **I**(*hkl*), are proportional to the square of the structure factors, $|\mathbf{F}(hkl)|^2$. **I**(*hkl*) provides only the magnitude of the structure factor, and not the phase. Therefore, electron density maps cannot be generated directly from the intensities. This is known as the "phase problem" and is discussed in a later section.

1.2.5 Crystallization and Crystal Mounting

Crystallographic structure determination begins with the growth of a suitable crystal, and the production of diffracting crystals is often the bottleneck in the process. The hanging-drop vapor diffusion technique is the most popular method for crystallization of macromolecules. The principle of vapor diffusion is straightforward. A drop composed of a mixture of sample and reagent solution is placed in vapor equilibrium with a liquid reservoir of reagent. Typically the droplet contains a lower reagent concentration than the reservoir. To achieve equilibrium, water leaves the droplet and eventually ends up in the reservoir. As water leaves the droplet, the sample undergoes an increase in supersaturation. Both the sample and reagent increase in concentration as the water leaves the droplet for the reservoir. The supersaturated state is a thermodynamically unstable state which can develop either in a crystalline or an amorphous phase when it returns to equilibrium. Equilibrium is reached when the reagent concentration in the droplet is approximately the same as that in the reservoir. The formation of crystals depends critically on a number of factors including temperature, pH, ionic strength and the molecular properties of the sample to name a few. The combination of the factors required to obtain diffraction quality crystals must be determined empirically by screening a wide range of conditions.

A macromolecular crystal usually consists of between 30% and 80% solvent of crystallization and as such is an unusual state of crystalline material. As a result there are generally relatively few intermolecular interactions between the macromolecules. The macromolecules are tenuously held together by a series of hydrogen-bonds in a network of solvent.

Crystal mounting methods are designed to maintain the high solvent content of the crystals. The crystals can be mounted in one of two ways: (1) sealed in a thin-walled glass capillary with a small amount of the crystallization solution to prevent dehydration, or (2) suspended by surface tension in a thin film of cryo-solution and flash-frozen to cyro-temperatures (around 100 K) on a small nylon loop. The purpose of the cryo-solution is to prevent the formation of ordered ice, which would interfere with the scattering due to the sample. The cryo-solutions are designed to approximate the equilibrium conditions of the crystallization reagents in the droplet with the exception of an added cryo-protectant, such as glycerol, ethylene glycol, or MPD (2-methyl-2,4-pentanediol). Sufficient cryo-protectant concentration is indicated by transparent glass formation on flash cooling.

1.2.6 Data Collection

Most data collection is done by the oscillation method with an area detector, usually an image plate or CCD detector. The purpose of the oscillation method is to stimulate a reflection fully via an angular rotation. In this technique the crystal is rotated through a small angle, ϕ (0.1 to 1.5 °) about an axis perpendicular to the X-ray beam. As the crystal rotates, successive reciprocal lattice planes satisfy the diffraction condition and the diffraction pattern is recorded on the area detector (Figure 1-10). Provided the rotation angle is not too large, adjacent levels will not overlap and data from many layers can be collected on each image. The rotation angle is chosen depending on the detector, crystal cell dimensions, Bravais lattice type, properties of the incident X-ray beam, and crystal mosaic spread.

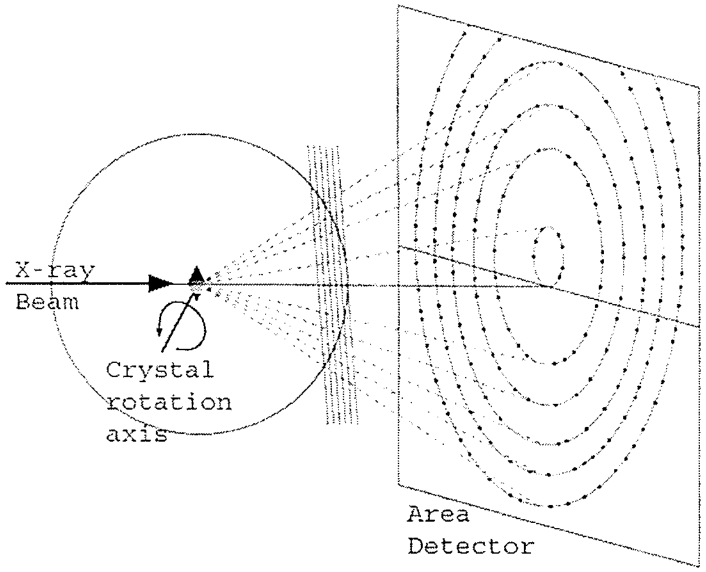

Figure 1-10 The oscillation geometry.

The plane of the reflections in the reciprocal sphere that is approximately perpendicular to the X-ray beam gives rise to an ellipse of reflections on the detector.

The raw data images consist of a two-dimensional array of spots (*i.e.* reflections) on the detector. There are a number of steps which must be carried out in order to obtain a set of scaled and indexed structure factor intensities from the raw data images. The major steps involved are: (1) visualization and preliminary analysis of the diffraction pattern (spot identification), (2) indexing of the diffraction pattern, (3) refinement of the crystal and detector parameters, (4) integration of the diffraction maxima, (5) relative scaling between images, (6) precise refinement of crystal parameters using the whole data set, and

(7) merging and statistical analysis of the measurements related by space-group symmetry (Otwinowski & Minor, 1997). Once a data set has been obtained the only problem remaining is to solve the structure.

1.2.7 Structure Solution Methods

There are two main approaches to solving macromolecular structures or equivalently solving the phase problem. The first approach involves perturbing the structure (*e.g.* by the addition of a heavy atom) and/or the diffraction pattern and using the differences to gain phase information. One application of this approach, made possible by the broad spectrum emission of synchrotron radiation, is known as multiple-wavelength anomalous dispersion (MAD) phasing. The second approach is to guess at the phases. This technique is known as molecular replacement and requires knowledge of what the molecular structure is expected to look like.

1.2.7.1 Patterson Function

As noted earlier, the structure factor phases cannot be determined experimentally. While the electron density is not directly accessible from the measured structure factor intensities, all of the information about the structure is contained in the intensities. The Fourier transform of the intensities (amplitudes squared), known as the Patterson function (Equation 1-10), yields a map of the vectors between atoms (Figure 1-11). For relatively small numbers of atoms, it is possible to work out the original positions of the atoms that would give rise to the observed Patterson peaks. For large numbers of atoms the Patterson map quickly becomes too complex to solve directly. The Patterson function is useful in both MAD and molecular replacement approaches to structure solution.

$$P(u, v, w) = \frac{1}{V} \sum_{hkl} I(h, k, l) \exp\{-2\pi i(hu + kv + lw)\} \tag{EQ 1-10}$$

P(u, v, w) is called the Patterson function, V is the unit cell volume, u, v, w is a vector in real space, and *h, k, l* is a reciprocal space vector with indices *h, k,* and *l.*

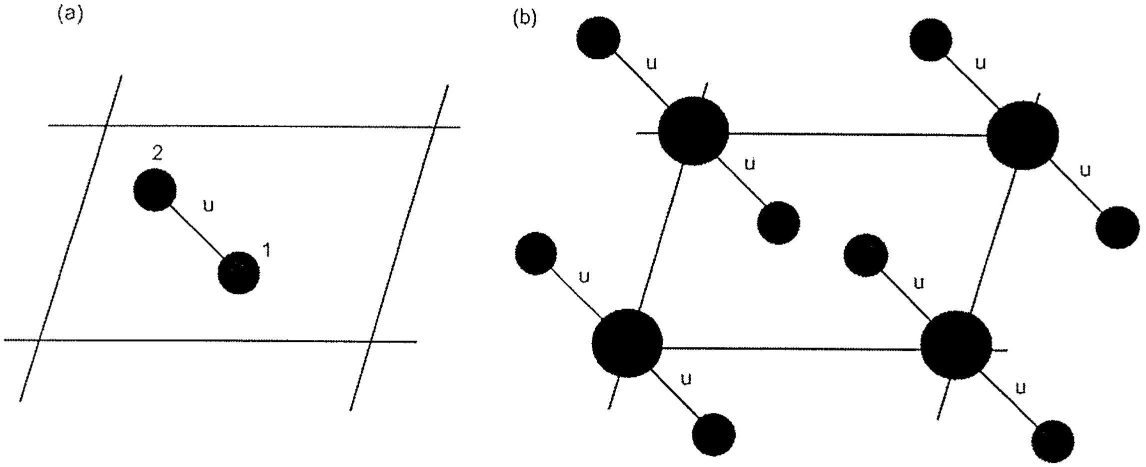

Figure 1-11 Patterson function.
(a) A two-dimensional unit cell with only two atoms. (b) The corresponding Patterson map.

1.2.7.2 MAD Phasing

The theoretical basis for MAD phasing lies in the atomic scattering factors. In the calculation of atomic scattering factors (see section "1.2.4"on page 22) it was assumed that the frequency of the incident wave was large compared to the resonance frequency of the atom. For wavelengths normally used in crystallographic studies, this assumption holds approximately only for light atoms, but is not generally true for heavier atoms whose absorption edges fall between 2 Å and 0.3 Å (Table 1-2). When the absorption edge of an atom is approached, anomalous scattering occurs. The atomic scattering factor which takes into account these resonance effects, $f_{anomalous}$, is separated into three parts:

$$f_{anomalous} = f_o + f' + if''$$

where f_o is the contribution to the scattering if the electrons are free electrons and is a real number, and f' and f'' are the correction terms. The quantity f' is in phase with f_o and f'', the imaginary part, is always π/2 ahead of the phase of the real part (Figure 1-12).

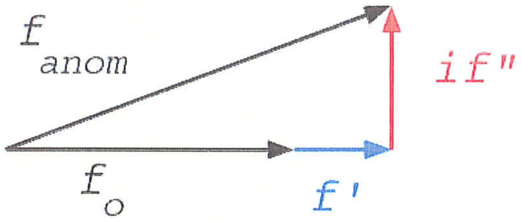

Figure 1-12 Anomalous scattering factor.

The anomalous contribution to f_{anom} consists of two parts, the real part, f', in phase with f_o and an imaginary part, if'', π/2 out-of-phase f_o.

TABLE 1-2 **Elemental absorption edges.**

Atomic No.	Element	Edge (Å)
6	C	43.7
7	N	30.9
8	O	12.4
26	Fe	1.743
27	Co	1.608
30	Zn	1.284
34	Se	0.979
35	Br	0.9202

The presence of f'' in the structure factor leads to a breakdown in Friedel's law. Friedel's law refers to the relationship between a reflection *hkl* and its partner $\overline{hkl}$. It can be shown that the magnitude of **F**(*hkl*) and **F**($\overline{hkl}$) are equivalent and that the phases are equal but opposite in sign when there is no anomalous scattering (Figure 1-13). The π/2 phase shift in the imaginary portion of the atomic structure factor correction results in an anomalous difference between |**F**(*hkl*)| and |**F**($\overline{hkl}$)| (Figure 1-13). The presence of f' results in differences in the structure factor amplitudes at different wavelengths. These differences can be exploited in order to gain phase information.

In a MAD experiment an anomalous scatterer (a heavy atom) is incorporated into the macromolecular structure and complete data sets are collected at multiple wavelengths. In order for a MAD phasing experiment to be successful, the signal from the anomalous scatterer must be sufficiently large relative to the signal from anomalous plus non-anomalous scatterers. The anomalous signal from each heavy atom depends on the absorption edges of the atom and the incident wavelength. The strength of the anomalous scattering effect depends on the wavelength of the X-rays, which will change both the normal scattering, *f'*, and the out-of-phase scattering, *if''*, of the anomalous scatterer. (Figure 1-14).

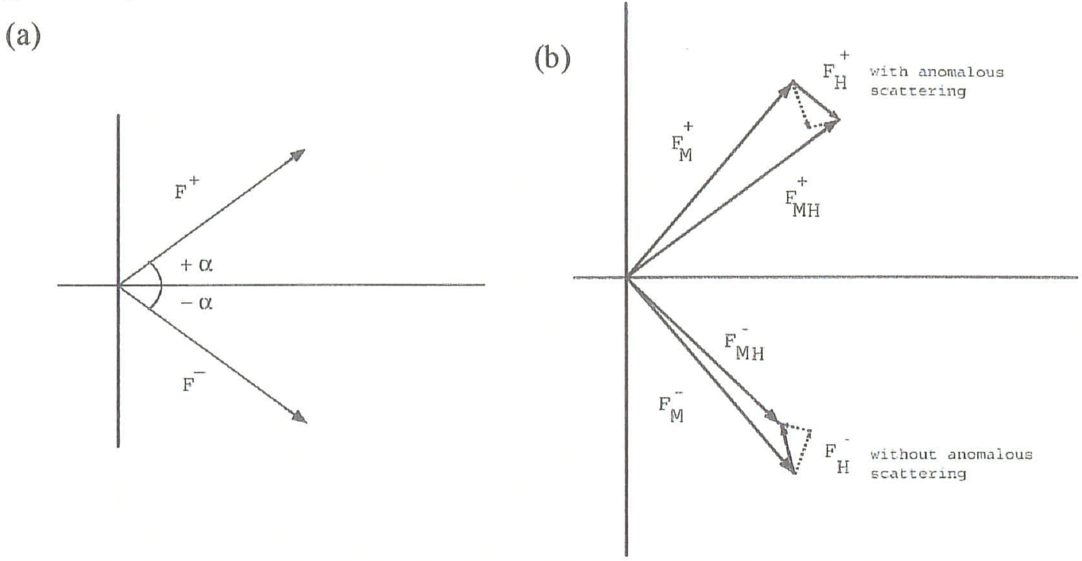

Figure 1-13 Friedel's law.

(a) In the absence of anomalous scattering, Friedel's law is obeyed and |F(hkl)| = |F($\overline{hkl}$)| and the phases are equal but opposite in sign. (b) By adding a heavy atom with an anomalous scattering component to (b) the symmetry between the Friedel pair's is broken due to the imaginary component of the anomalous scattering which is +π/2 out-of-phase regardless of which of the Friedel mates is considered.

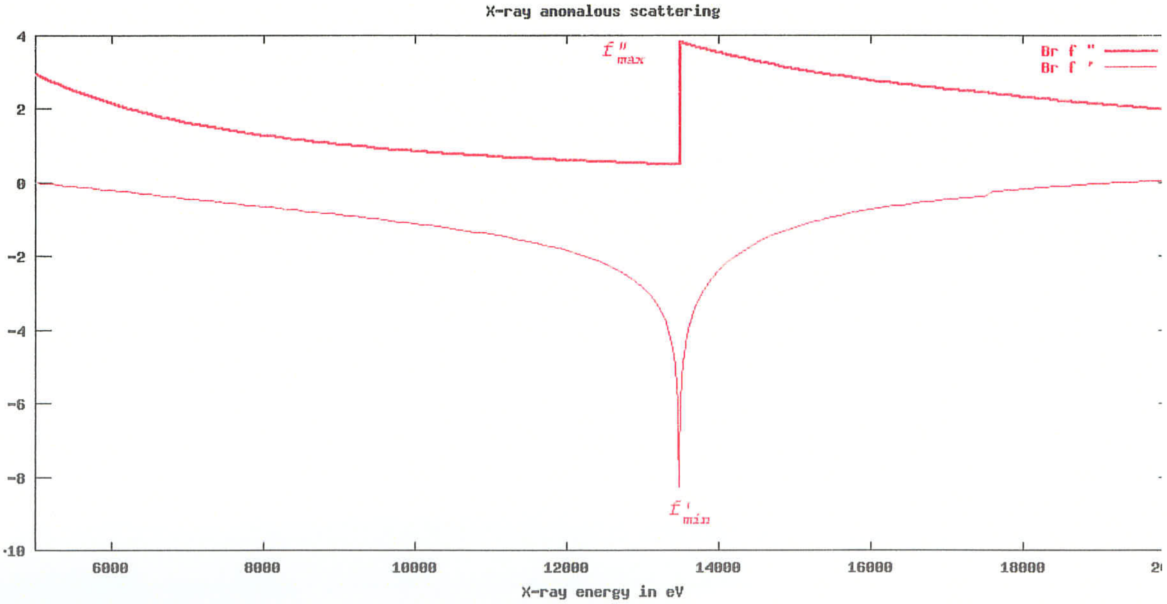

Figure 1-14 *f'* and *f''* as a function of X-ray energy.

Theoretical plot of *f'* and *f''* for bromine as a function of energy (Brennan & Cowan, 1992).

The expected signal from a MAD experiment can be calculated using the following equations:

perturbation due to $f''= (N_A/2N_T)^{1/2}(2f_A''/Z_{eff})$

perturbation due to $f'= (N_A/2N_T)^{1/2}(|f_A'(\lambda_i) - f_A'(\lambda_j)|)/Z_{eff}$

where N_A = number of anomalous scatterers, with anomalous scattering factors f'' and f'.

N_T = total number of atoms in the structure.

Z_{eff} = effective normal scattering power for all atoms.

Since the perturbation due to f'' and f' are orthogonal, the net expected signal can be taken to be the root-mean-square sum of these two quantities. Thus in order to maximize the expected signal, knowledge of the anomalous scattering factors in the crystal environment is required. The imaginary component of the anomalous scattering is proportional to the atomic absorption coefficient for the atom and can be obtained from X-ray absorption spectra. The relationship between f'' and f' is known as the Kramers-Kronig dispersion relation (James, 1948) and allows for the calculation of f' from the experimentally determined f''. The maximum perturbation due to f'' is achieved at the energy of peak absorption, f''max. The maximum perturbation due to f' is achieved between structure amplitudes at the energy of the inflection point of the edge (f'min) and at a remote energy (f'remote, typically on the high-energy side of the absorption edge), and is proportional to $|f'\text{min} - f'\text{remote}|$. In a three-wavelength experiment, diffraction data are collected at X-ray energies corresponding to f''max, f'min and f'remote.

A remarkable feature of anomalous scattering is that it does not fall off with $(\sin\theta)/\lambda$. This means that the contribution of anomalous scattering to the total measured intensity actually increases at higher resolution. Thus the phasing power of a high resolution data set is much higher than that of a lower resolution data sett. The MAD phasing signal is relatively small in both cases so great care must be taken during data collection to minimize errors in the intensity measurements.

To understand MAD phasing it is necessary to return to the structure factor equation, now including the wavelength-dependent anomalous scattering terms, for an individual reflection **h** at wavelength λ:

$$ {}^{\lambda}F(h) = {}^{o}F_T(h) + \sum_{k}^{\text{number of kinds}} {}^{o}F_{A_k}(h)\left[\frac{f'}{f_o} + \frac{f''}{f_o}\right]_k \qquad \textbf{(EQ 1-11)} $$

where ${}^{o}F_T$ is the structure factor representing the normal scattering component for all the atoms. ${}^{o}F_A$ is the normal scattering component for all the anomalous scatterers. This factorization separates out the wavelength dependence so that it is contained solely in the anomalous scattering factors f' and f'', which are independent of atomic positions.

In the simplest case of only one type of anomalous scatterer, the magnitude of the measured intensity $|{}^{\lambda}F(h)|^2$ can be expressed in terms of its components:

$$ \begin{aligned} \left|{}^{\lambda}F(\pm h)\right|^2 = {} & \left|F_T(h)\right|^2 + \\ & a(\lambda)\left|F_A(h)\right|^2 + \\ & b(\lambda)\left|F_T(h)\right|\left|F_A(h)\right|\cos[\Phi_T(h) - \Phi_A(h)] \pm \\ & c(\lambda)\left|F_T(h)\right|\left|F_A(h)\right|\sin[\Phi_T(h) - \Phi_A(h)] \end{aligned} \qquad \textbf{(EQ 1-12)} $$

where:

$$a(\lambda) = \frac{f'^2 + f''^2}{f_o^2};$$

$$b(\lambda) = \frac{2f'}{f_o};$$

$$c(\lambda) = \frac{2f''}{f_o}$$

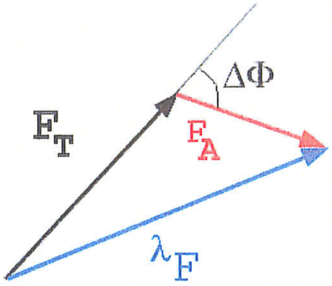

Figure 1-15 MAD phasing.

The measured $^{\lambda}F(h)$ amplitudes combined with the MAD phase equations make it possible to solve for the structure factor amplitudes of F_A and F_T as well as the phase difference, $(\Phi_T - \Phi_A)$.

Figure 1-15 illustrates the vector relationships among the three quantities that the MAD phasing equations solve for each Bragg reflection **h** by measuring the scattering amplitude at wavelength λ. Each measurement of $|F(\mathbf{h})|$ at some wavelength gives one instance of the above equation, and the separate instances may be treated as a system of simultaneous equations from which the quantities $|F_A|$, $|F_T|$ and $(\Phi_T - \Phi_A)$ can be obtained.

In actuality there are two measurements for each wavelength, because the observation for $|F(\bar{h})|$ may be treated as an observation of $|F(\mathbf{h})|$ with the value of f'' inverted. To obtain the values for the three unknowns, at least three observations are necessary and data collection over two wavelengths is required. However, MAD experiments are generally carried out at three wavelengths resulting in over-determination of the system.

The essence of MAD phasing is that if the anomalous scatterers can be located in the unit cell then the corresponding phase angle $\Phi_A$can calculated. From the MAD phasing equations values for $|F_T|$ and $(\Phi_T - \Phi_A)$ are obtained. Thus, if $\Phi_A$is known Φ_T can be calculated and a Fourier transform of $|F_T|$ and Φ_T (Equation 1-9) will yield an electron density map corresponding to all the atoms in the structure. For most MAD structures the anomalous scatterer sites are located using a Patterson map based on the anomalous differences (*i.e* differences between Bijvoet pairs, $|\mathbf{F}(hkl)|$ and $|\mathbf{F}(\overline{hkl})|$). The resulting Patterson map contains only the interatomic vectors between the anomalous scatterers and their surroundings and is generally simple enough for direct solution. The correct enantiomorph for the anomalous-scatterer partial structure must be determined (Φ_A versus $-\Phi_A$) in order to obtain an electron density image of the macromolecule. The choice of enantiomorph cannot be determined directly from the MAD data. In order to determine the correct choice electron density maps must be calculated with both enantiomorphs. The correct choice of enantiomorph is indicated by an interpretable electron density map. At this point it is possible to fit a starting model to the electron density and to begin refinement.

1.2.7.3 Molecular Replacement

When identical or similar structures exist in different crystallographic environments, similarities between their diffraction patterns are expected. The technique of molecular replacement exploits this similarity for phase determination. The basis for the method involves finding the orientation and position of a model structure within the unit cell of the unknown structure. From the model structure in the unknown unit cell a set of starting phases can be generated for use with the observed structure factor amplitudes for calculation of a starting electron density map. The relationship between two structures can be expressed by six parameters, three rotational and three translational. It is possible to split this search into two three-dimensional searches: a rotation search followed by a translation search (Figure 1-16).

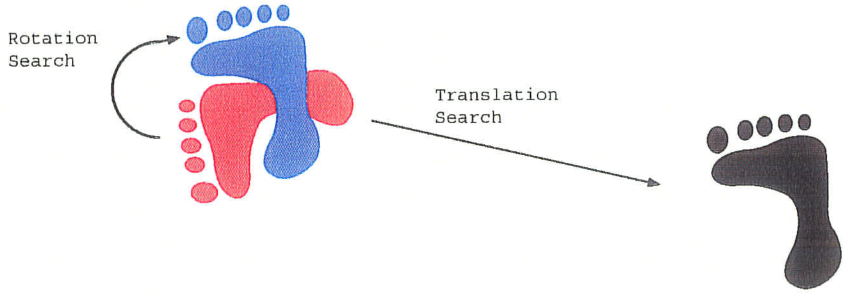

Figure 1-16 Generalized molecular replacement.

The six-dimensional search required to correctly orient and position the search model in the target unit cell can be broken down into two three-dimensional searches: a rotation search to correctly orient the model, followed by a translation search to correctly position the search model in the unit cell.

The rotation search is carried out by looking for agreement between the Patterson functions of the search model and the unknown structure as a function of their relative orientation. A function to evaluate this agreement was originally defined by Rossmann and Blow (Rossmann & Blow, 1962) as:

$$R(C) = \int_{U} P_1(x) P_2(C_x) dV \qquad \text{(EQ 1-13)}$$

where P_1 and P_2 are Patterson functions, C is a rotation operator that rotates the coordinate system of P_2 with respect to P_1, and U is a volume of integration, usually spherical, centered at the origin. The rotation function exploits the fact that the intramolecular vectors depend only on the orientation of the molecule and not on its position in the unit cell. Thus in the evaluation of the rotation function only the intramolecular vectors in Patterson map need be considered. Typically the search model is considered within a large unit cell of P1 symmetry to eliminate intermolecular vectors in its Patterson map. Since all vectors in a Patterson map start at the origin, the vectors closest to the origin in the observed Patterson map will generally correspond to the intramolecular vectors. Thus by choosing an appropriate integration volume, U, it is possible to confine the rotation search to just the intramolecular vectors. A maximum of the rotation function R(C) indicates a potential orientation for the search model in the target crystal.

Once the orientation of the search model is known, the position in the unit cell must be found. The only difference between correctly and incorrectly positioned molecules is in the intermolecular vectors. The intermolecular vectors are dependent on the position within the unit cell and the symmetry of the cell. Traditional translation functions compare sets of intermolecular vectors from the model in the unit cell of the unknown structure with the observed Patterson. The simplest translation function is similar to a correlation function and is given by:

$$T(x, y, z) = \sum_{h, k, l} F^2_{calc}(h, k, l, x, y, z) F^2_{obs}(h, k, l) \qquad \text{(EQ 1-14)}$$

where the F^2_{calc} are the calculated structure factor intensities of the model for each reflection, *h, k,* and *l*, at all positions, x, y, and z, within the unit cell and F^2_{obs} are the observed structure factor intensities. When T(x,y,z) is a maximum the model structure should in best agreement with the observed data. An alternative to the translation function is a translation search whereby the search model is moved on a grid within the unit cell and at each point the calculated structure factor amplitudes and the observed amplitudes are compared by calculating an R-factor (Equation 1-15) as a function of the molecular position. The R-factor is a measure of the percent difference between the observed and the calculated structure factor amplitudes. Thus the molecular position with the lowest R-factor is likely the best solution.

$$R = \frac{\sum_{h, k, l} \left| |F_{obs}(h, k, l)| - |F_{calc}(h, k, l)| \right|}{\sum_{h, k, l} |F_{obs}(h, k, l)|} \qquad \text{(EQ 1-15)}$$

1.2.8 Refinement

The goal in macromolecular crystallography is to obtain an accurate model based on the observed diffraction data. Once an initial model has been obtained, the positions of atoms in the crystal and their thermal parameters are optimized based on the agreement between the observed and calculated diffraction data. The refinement of a macromolecular structure is a difficult optimization problem because it is poorly determined (low observation to parameter ratio) and highly linear (the atomic coordinates are highly interdependent). The target functions for optimization are complicated functions of the parameters (positions and B-factors of the atoms), with many local minima in addition to the desired global minimum, which often results in the model adopting conformations other than the desired "best fit". Some of these problems have been addressed in the design of crystallographic refinement techniques. The apparent ratio of data-to-parameters can be increased by incorporating chemical information in the form of restraints or by reducing the conformational freedom of the model with constraints. The careful choice of the refinement target can reduce the significance of local minima, but does not completely eliminate them. If the starting model is poor, the significance of local minima is still very high, but can be overcome with the use of molecular dynamics-based simulated annealing. In a crystallographic refinement, the goal is to find a global minimum of an energy target (Equation 1-16) as a function of the atomic coordinates.

(EQ 1-16)

$$E = E_{\mathrm{chem}} + w_{\mathrm{X\text{-}ray}} E_{\mathrm{X\text{-}ray}}$$

where E_{chem} contains empirical information about the restraints on the system, $E_{\mathrm{X\text{-}ray}}$ is related to the difference between observed and calculated data, and $w_{\mathrm{X\text{-}ray}}$ is a weight chosen to balance the gradients arising from the two energy terms such that the model is chemically reasonable and agrees with the diffraction data.

The empirical energy term is generally based on harmonic restraints. Typical restraints include bond lengths, angles, torsion angles, and as well as a non-bonded contact interactions. The restraints are entered as energy terms and the parameters are derived from the average geometry and deviation observed in high resolution structures found in the Nucliec Acid Database which contain bases, sugars or the phosphodiester linkage.

$$E_{chem} = \sum_{bonds} k_b(r - r_o)^2 + \sum_{angles} k_\theta(\theta - \theta_o)^2 + \sum_{dihedrals} k_\phi \cos(n\phi + d) + \sum_{ij}(ar^{-12} + br^{-6}) \quad \text{(EQ 1-17)}$$

The X-ray energy term is dependent on the optimization method. Several methods have been developed for the refinement of macromolecules, including partial and full matrix least-squares (Hendrickson, 1985; Tronrud *et al.*, 1987), conjugate gradient minimization (Jack & Levitt, 1978) and molecular dynamics-based simulated annealing (Brünger *et al.*, 1987). One method implemented recently is based on a maximum likelihood function which has been found to reduce the significance of local minima (Pannu & Read, 1996; Murshudov *et al.*, 1997; Adams *et al.*, 1999). The maximum likelihood target can be used alone in first order gradient descent minimizations or incorporated into molecular dynamics-based simulated annealing approaches.

In essence, the maximum likelihood is quite simple: the best model is most consistent with the observations. To say that a model is consistent with an observation means that, if the model were correct, there would be a reasonably high probability of making an observation with that value. With the observations taken as a set, the probability of making the entire set of observations is an excellent measure of the quality of the model. If the observations are assumed to be independent, the joint probability of making the set of observations is the product of the probabilities of making each individual observation. In the case of crystallographic refinement, it is not strictly true that the diffraction observations are independent. However, this approximation has been implemented in crystallographic refinements with success (Pannu & Read, 1996). The probabilities of making an observation have to include the effects of all sources of error, including not just measurement errors, but also errors in the model itself. A function, σ_A, is used to obtain an estimate of the model errors and incompleteness. To make this estimate as unbiased as possible, cross-validation is used by setting aside a randomly selected set of typically 10% of the diffraction data (often referred to as the test set) for the calculation of σ_A, while the remaining 90% (or working data set) is used in the refinement of the target function. The σ_A function along with the model structure are used to compute the expected value of $<|F_{obs}|>^{CV}$ and the corresponding variance $\sigma^2_{ML}{}^{CV}$(which takes into account all errors, not just those associated with the model). The X-ray energy term is then given as:

$$E_{X-ray} = \sum_{hkl \in working} \left(\frac{1}{\sigma^2_{ML}{}^{CV}}\right)(|F_{obs}| - <|F_{obs}|>^{CV})^2 \qquad \textbf{(EQ 1-18)}$$

The crystallographic refinement proceeds iteratively with successive optimization calculations followed by model validation. The overall progress of a refinement is monitored by the R-factor (Equation 1-15) which can be calculated from the working data set, $R_{working}$, and the test data, R_{free}. Large discrepancies between $R_{working}$ and R_{free} are generally indicative of problems with the model. Deviations from ideal geometry and poor correlation with electron density maps are also indicators that further refinement is necessary.

As indicated earlier, if the starting model is extremely poor, the model structure may become trapped in a local-minimum conformation. To fully optimize the structure it becomes necessary to resort to other optimization methods such as molecular dynamics-based simulated annealing which explores a greater region of conformational space. Simulated annealing requires the definition of a target function based on the parameters of the system, a generation mechanism to create a Boltzmann distribution of conformations at a given temperature, and a sequence of temperatures at which the Boltzmann distribution is computed. It is important to note that the temperature in a simulated-annealing refinement has no physical meaning. It does, however, denote the likelihood of overcoming barriers in the target function. With the minimization target, E, (Equation 1-16), defined as the potential energy of the system, it is possible to use molecular dynamics to generate changes in atomic coordinates consistent with the Boltzmann distribution at a given temperature. Molecular dynamics consists of the numerical integration of Newton's equations of motion:

$$m_i \frac{\partial^2 r_{i,u}}{\partial t^2} = -\frac{\partial E}{\partial r_{i,u}} \qquad \textbf{(EQ 1-19)}$$

where $\mathbf{r}_{i,\mathrm{u}}$ and m_i are the coordinates and mass of atom *i* respectively, and E is the potential energy given by Equation 1-16 . The success of simulated annealing depends highly on the annealing schedule. A common annealing schedule is to heat the system briefly to a high temperature and then allow it to cool slowly. At each step in the cooling process, molecular dynamics are performed; the annealing schedule is complete when the system has returned in its starting temperature. The temperature needs to be high enough to allow conformational transitions, but not so high that the model moves too far away from the correct structure. While seemingly unreasonable from a biological point of view, the annealing process allows the model to escape local minima in the target function and reach a better fit with the data.

CHAPTER 2 Experimental

2.1 The Reagents and DNA Molecules

In crystallizing macromolecules it is important to control the quality of all reagents used. Any variations in the reagents can lead to reproducibility problems in the crystallization experiments. The reagents used for all solution studies and crystallization trials are listed in Table 2-1. All reagents were dissolved in nanopure filtered water (dH_2O, Milli Q filter) to give a set of stock working solutions. The choice of reagents and how they interact with the DNA will be described later.

TABLE 2-1 **Reagents for solution studies and crystallization trials.**

Reagent	Form	Source	Concentration
Barium chloride	anhydrous	Science Borealis Limited	200 mM
Cacodylic acid	sodium salt hydrate 98% purity	Aldrich	250 mM
Calcium chloride	anhydrous	Damon Educational Inc.	200 mM
Cobalt hexammine	chloride salt	Eastman Kodak Company	100 mM
Cobalt nitrate	certified A.C.S grade	Fisher Scientific	200 mM
Cupric chloride		Damon Educational	200 mM
Magnesium chloride	hydrate 99.9% purity	Aldrich	200 mM
Manganous chloride	hydrate 99.3% purity	Baker Chemical Company	
2-methyl-2,4-pentanediol	99+% purity	Aldrich	50% (v/v)
Nickel chloride	hydrate 98% purity	British Drug Houses Ltd.	200 mM
Potassium chloride	certified A.C.S. grade	Fisher Scientific	800 mM
Sodium chloride	analytical reagent grade	Mallinckrodt	800 mM
Spermindine	99% purity	Aldrich	100 mM
Spermine	tetrahydrochloride 95% purity	Sigma	100 mM
Strontium chloride		Macalaster Scientific	200 mM
Zinc chloride	certified A.C.S. grade	Fisher Scientific	200 mM

The DNA molecules, including the brominated derivatives, were synthesized by Midland Certified Reagent Company (Midland, Texas) using phosphoramidite chemistry. The DNA oligomers were provided in pure form as the ammonium salt, after purification by high performance liquid chromatography. Table 2-2 contains the DNA sequences studied and some of their physical properties. Working solutions of the DNA were prepared by dissolving 5-15 mg of the DNA in 50-150 μL of dH_2O to give a concentration of 10-15 mM. The actual DNA concentrations were determined by diluting 1 μL of the DNA stock with 999 μL of dH_2O and taking UV absorption readings at 260 nm (Shimazu UV-2101PC scanning spectrophotometer with matching quartz SUPRASIL UV cells, pathlength of 1 cm and approximate capacity of 1.4 mL). The DNA concentrations were then calculated using Beer's law and the extinction coefficients (SS values used) in Table 2-2.

$$A_{260} = \varepsilon_{260} c l \tag{EQ 2-1}$$

where A and ε, respectively, are the absorbance and extinction coefficient at 260 nm, c is the concentration of the DNA and *l* is the pathlength of the UV cell.

TABLE 2-2 The DNA sequences.
The extinction coefficients at 260 nm were calculated according to Fasman (Fasman, 1976) for both the single stranded (SS) and double stranded (DS) DNA species.

DNA Sequence	Abbrev.	Extinction Coefficient (260 nm)		Molecular Weight (g/mol)
CGGCCGCCG	9mer	103 (SS)	144.2 (DS)	2692.79
CGGTGGCCACCG	12mer	108 (SS)	151.2 (DS)	3636.4
CGGTGG[^{5}BrC]CACCG	Br12mer	108 (SS)	151.2 (DS)	3726.3
GGTGGCCACCGC	GG12mer	107.3 (SS)	150.2 (DS)	3636.4

2.2 UV and CD Solution Studies

The absorption of UV light by molecules occurs when the energy of the incident radiation is the same as that of a possible electronic transition in the molecule. Although electronic transitions arise between ground and excited states of the entire molecule, the majority of the action can usually be assigned to parts of the molecule where electrons are loosely bound (chromophores). The UV absorbance of DNA is almost entirely due to its aromatic bases which exhibit a π-π* electronic transition around 260 nm (Table 2-3).

When the bases are incorporated into the DNA double helix, there is a reduction in the UV absorbance due to electronic interactions between the stacked bases. This phenomenon is known as the hyperchromic effect. Thus by monitoring the UV absorbance of DNA as a function of temperature it is possible to monitor the denaturation of duplex DNA and gain information about its thermal stability. The plot of absorbance versus temperature is referred to as a melting curve and the temperature at the midpoint of the transition is referred to as the melting temperature, T_m. If the DNA is denatured in a cooperative process, the denaturation will occur over a narrow temperature range. If the DNA is slowly unzippered, however, the thermal denaturation will occur over a broader temperature range. Thus the shape of the thermal denaturation profile can also yield some information as to the stability of the natured DNA conformation.

TABLE 2-3 **Light absorption characteristics for the nucleoside monophosphates (Bush, 1974).**

Nucleoside	λ_{max} (nm)	10^{-4} x ε ($Lmol^{-1}cm^{-1}$)
Adenosine	259.5	1.49
Guanosine	276	0.9
Cytidine	271	0.91
Thymidine	267	0.97

Information about the local conformational environment of the chromophores can be gained if a plane-polarized light source is used. Plane-polarized light is the sum of two beams of left- and right-circularly polarized light of the same frequency and intensity. The different interaction of the left- and right- circularly polarized light with an asymmetric chromophore results in circular birefringence and circular dichroism. The bases of DNA, being planar, are not optically active (*i.e.* they do not show circular birefringence or circular dichroism), but once they are substituted by a chiral moiety like a furanose, optical activity is induced.

When the plane-polarized light is passed through a solution containing an asymmetric chromophore, one of the components is slowed down relative to the other. Thus, the two components of the plane-polarized light emerge from the sample out-of-phase and the angle of polarization is altered (Figure 2-1b). This phenomena is known as circular birefringence. Circular dichroism results from the differential absorption of the left- and right-circularly polarized component in the region of an electronic transition. The two components of the plane polarized light emerge from the sample not only out-of-phase, but having different intensities. The emergent beam becomes elliptically polarized, with its major axis tilted through the angle caused by the circular birefringence (Figure 2-1c). Circular dichroism is measured by determining the ellipticity of the emergent light. Ellipticity, Θ, is defined by:

$$\Theta(\text{degrees})=32.98\ (\varepsilon_L-\varepsilon_R)lC \qquad \text{(EQ 2-2)}$$

where ε_L and ε_R are the extinction coefficients for the left- and right-circularly polarized light, respectively, l is the pathlength and C is the concentration of the sample.

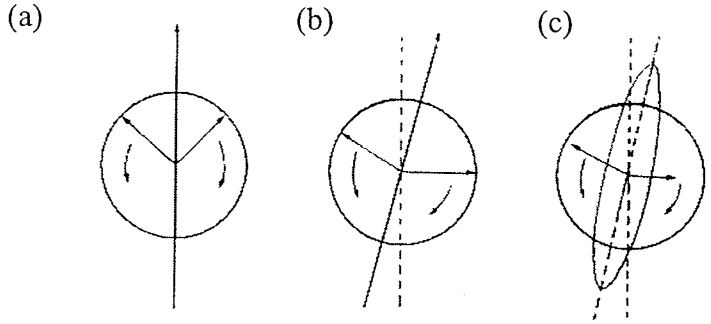

Figure 2-1 The effect of an optically active medium on plane-polarized light. (a) illustrates the original plane-polarized light is the sum of two circularly polarized beams, in (b) the optical activity is caused by the different indices of refraction, while in (c) both optical rotation and CD are active, giving the elliptical polarization.

2.2.1 C-C Mismatch Solution Studies

It was desirable to characterize the stability and behaviour of the 9mer sequence under a variety of conditions, prior to beginning crystallization trials, given the propensity of the C-C mismatch dodecamer sequence d(CGCTGGCCACCG) to adopt both hairpin and duplex structures(Chen, 1997). Factors known to affect the conformation and stability of DNA include: pH, cation concentration and type, and polyamine binding. The effect of these factors on the 9mer were investigated in a series of UV and CD experiments.

Solutions were prepared for UV and CD experiments by direct dilution from the DNA stock solution into the desired buffers. Samples were allowed to equilibrate for 24 hours before readings were taken. Concentrations were calculated from optical density readings at 260 nm, using the theoretical extinction coefficients in Table 2-2.

All DNA absorption spectra were measured using a Shimazu UV-2101PC scanning spectrophotometer and a HAKKE G water bath. Thermal denaturation studies were carried out in Teflon-stoppered quartz microcuvettes with 1 cm pathlengths. Sample heating was carried out in a step-wise manner, allowing the sample to equilibrate for 5 minutes before readings were taken. Melting curves were followed at 275 nm, the wavelength at which the greatest change in absorbance occurred throughout melting (Figure 2-2)

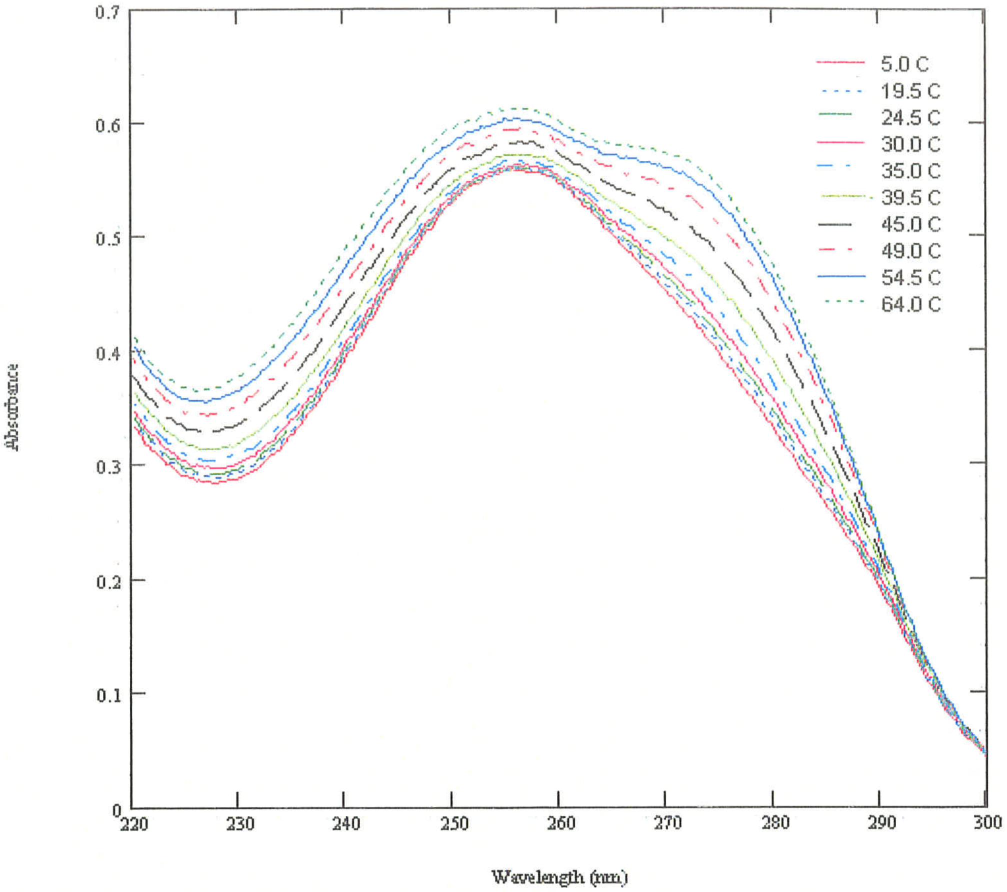

Figure 2-2 UV thermal denaturation profile.
UV absorbance spectra for the 0.008mM 9mer in 25mM Na cacodylate, 5mM $MgCl_2$ pH 5 buffer as a function of temperature.

All circular dichroism measurements were made by using a Jasco J500A spectropolarimeter with a thermostated cell connected to a NESLAB water bath. All spectra were measured in a Teflon-stoppered quartz SUPRASIL cell with a 1.0 cm pathlength. The scanning rate was 10nm/min. and the instrument sensitivity was 2. The spectra were collected between 220 and 300 nm as voltage readings. For each set of conditions, spectra were collected on a blank buffer and on the DNA containing solution. The spectra were converted to molar ellipticity (degrees M^{-1} cm^{-1}), [Θ], as follows:

$$[\Theta] = \frac{(4.0198)(\text{sample voltage - blank voltage})(\text{sensitivity})}{lc} \qquad \text{(EQ 2-3)}$$

where 4.0198 is a conversion factor, *l* is the sample pathlength (cm) and c is the DNA concentration (M).

2.2.1.1 Characterization of the 9mer

The 9mer sequence is capable of forming either a self-complementary double stranded duplex or a hairpin structure. Only the duplexed form of the 9mer will contain the C-C mismatch of interest. Since the thermal denaturation of DNA is an equilibrium process, information about the strand stoichiometry can be gained by investigation of the T_m dependence on DNA concentration. If the DNA forms some sort of single stranded structure, for example a hairpin structure, its melting temperature will be DNA concentration independent; if the DNA forms a duplexed structure its T_m will show a DNA concentration dependence (Figure 2-3). To determine which species is present in solution, the dependence of the melting temperature on the 9mer strand concentration was determined.

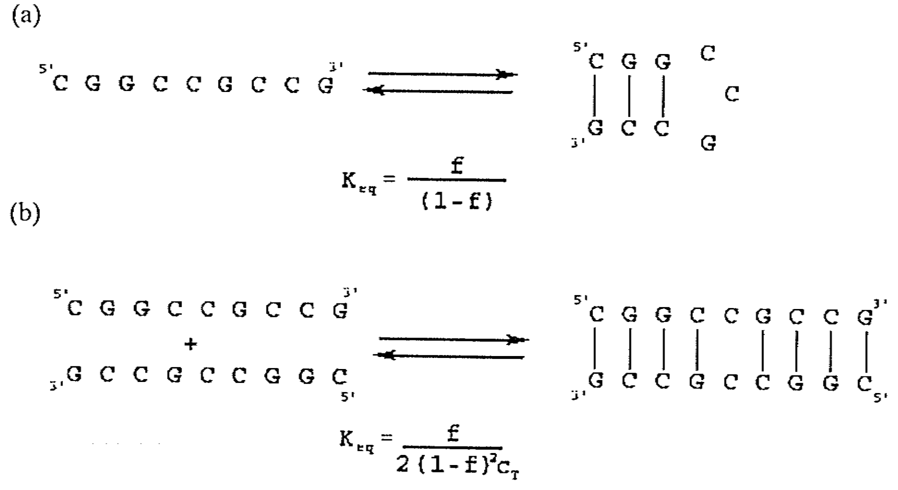

Figure 2-3 Possible thermal denaturation equilibria.
Structures and denaturation equilibria for the 9mer sequence. f is the fraction of total strand concentration on the left-hand side of the equilibrium and c_T is the total DNA strand concentration.

The thermal denaturation of the 9mer was followed in a pH 5 buffer of 25mM sodium cacodylate and 16 mM $MgCl_2$ for three DNA concentrations, 0.006, 0.010, and 0.015 mM. The DNA concentration dependence of the melting temperature confirms that the duplexed 9mer is the dominant species present in solution (Figure 2-4).

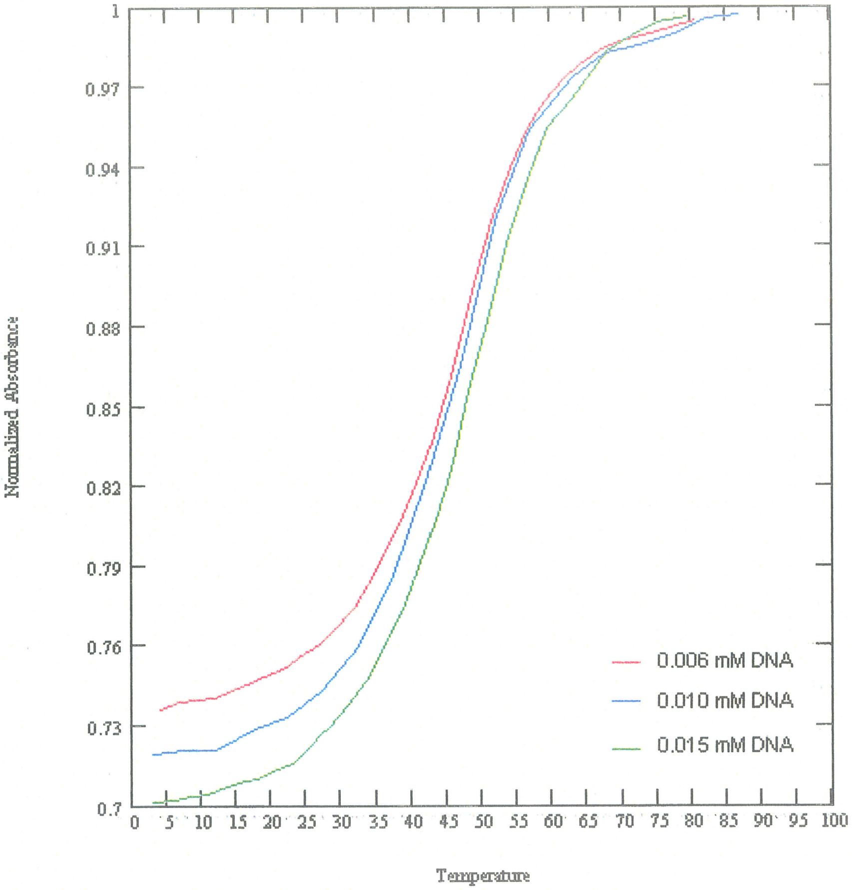

Figure 2-4 Effect of DNA concentration on T_m.
Thermal denaturation profile of the 9mer in 25mM Na cacodylate, 16mM $MgCl_2$, pH 5 buffer.

The effect of the pH, cation type and polyamine on the global conformation of the 9mer sequence were investigated using CD spectroscopy. Table 2-4 summarizes the buffers investigated. The DNA concentration in all cases was 0.008 mM and data was collected at 10°C. Figure 2-5 to Figure 2-7 summarizes the effects of pH, cation type and polyamine type on the DNA conformation. Comparison with CD spectra of DNA in an A- and B-type conformation (Figure 2-8) reveals that the 9mer has adopted a B-type conformation under all conditions investigated.

TABLE 2-4 **CD buffers.**

pH	[NaCacodylate]	[cation]	[polyamine]
7	25 mM	5 mM $MgCl_2$	0
5	25 mM	5 mM $MgCl_2$	0
5	25 mM	5 mM $ZnCl_2$	0
5	25 mM	50 mM NaCl	0
5	25 mM	5 mM $MgCl_2$	0.25 mM Spermine
5	25 mM	5 mM $MgCl_2$	0.25 mM Spermidine
5	25 mM	5 mM $MgCl_2$	0.25 mM Cobalt hexammine

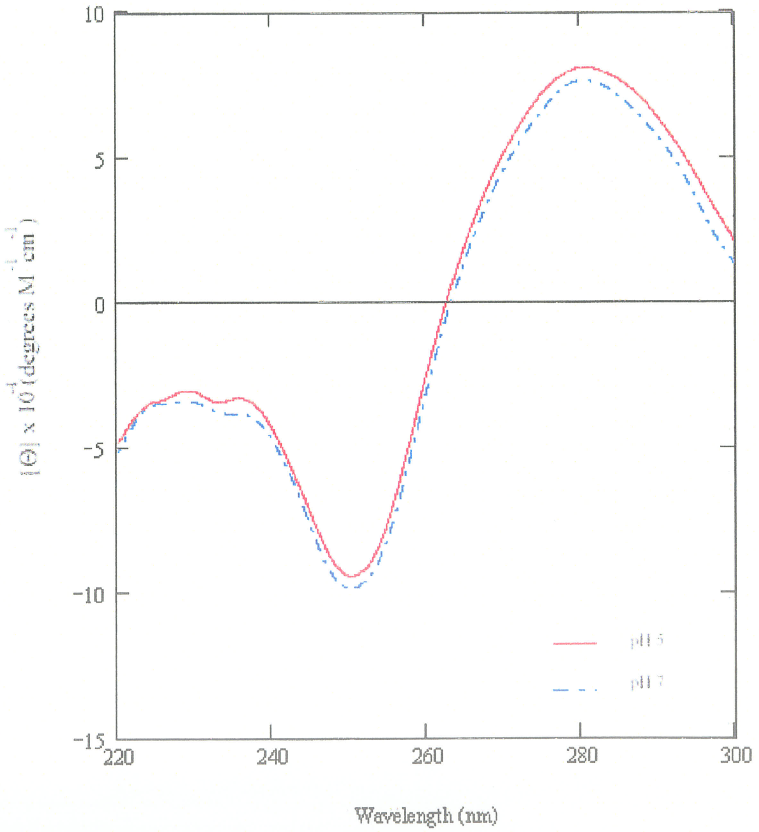

Figure 2-5 CD spectra for the 9mer at pH 5 and pH7.
Spectra of 0.008 mM 9mer in 25 mM Na cacodylate, 5mM $MgCl_2$ buffers of pH 5 and 7.

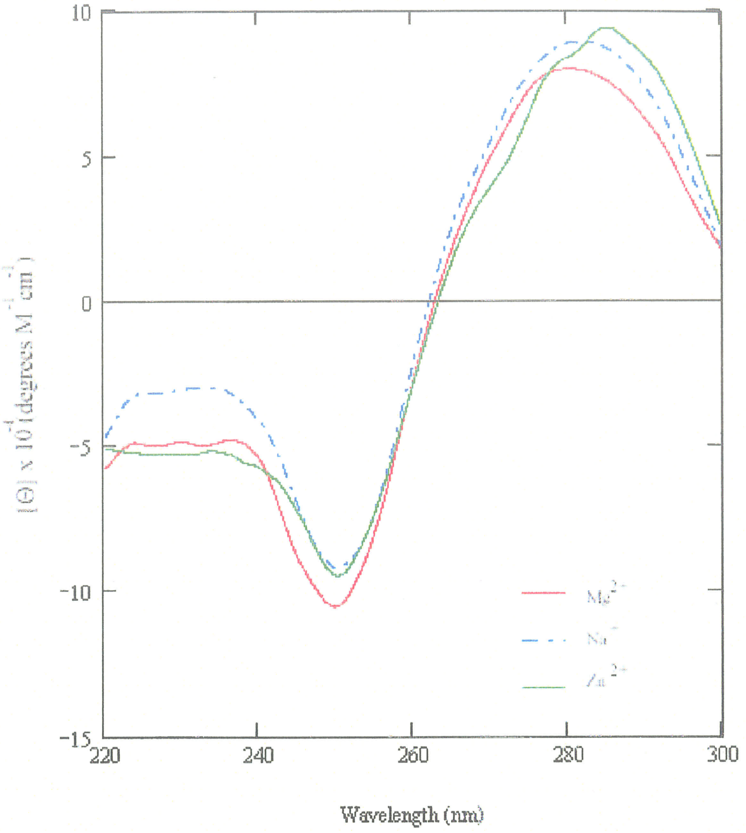

Figure 2-6 Effect of cation type on the CD spectra of the 9mer.
Spectra of 0.008 mM 9mer in 25mM Na cacodylate pH 5 buffer with 5 mM $MgCl_2$/$ZnCl_2$ or 50 mM NaCl.

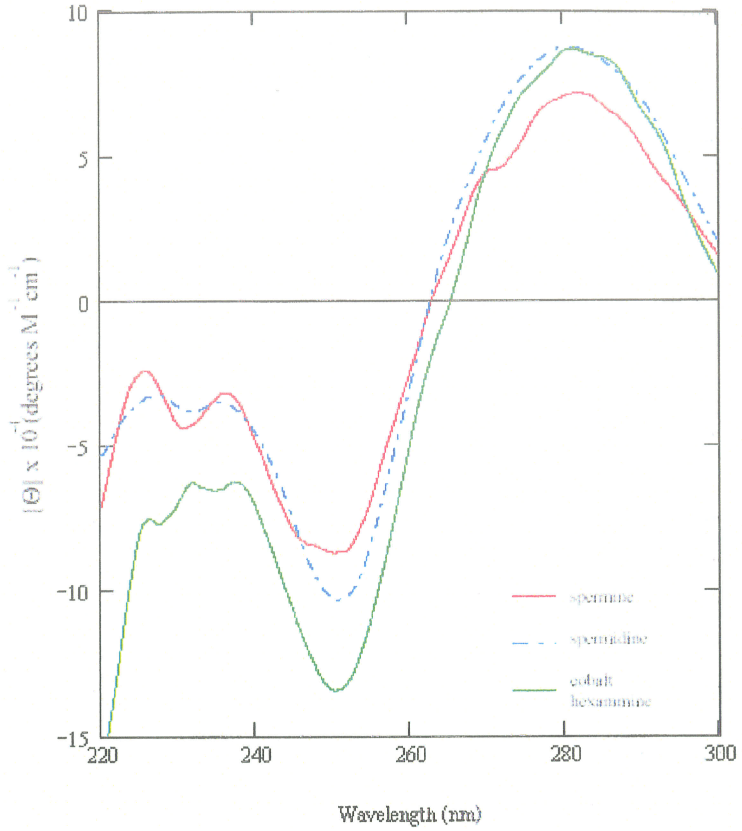

Figure 2-7 CD spectra of the 9mer in the presence of polyamines.
Spectra of 0.008 mM 9mer in 25mM Na cacodylate, 5mM $MgCl_2$ pH 5 buffer with 0.25mM spermine/ spermidine/cobalt hexammine.

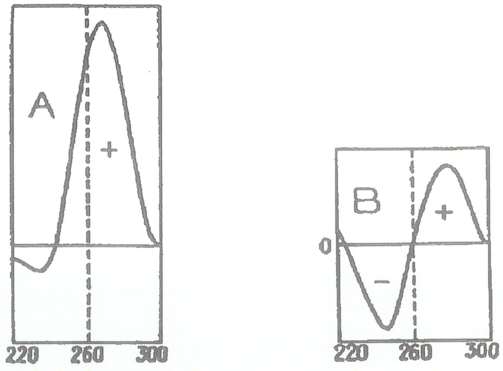

Figure 2-8 Schematic representation of the CD spectra for A- and B-DNA
Reproduced from Ivanov *et al.*, 1973.

2.2.1.2 Stabilization of the 9mer Sequence

UV thermal denaturation studies were carried out in order to find conditions under which C-C mismatch is most stable. This information provides a starting point for the crystallization trials. Table 2-5 summarizes the conditions investigated.

TABLE 2-5 **UV buffers.**

pH	[DNA]	[NaCacodylate]	[cation]	[polyamine]
7	0.005 mM	25 mM	16 mM $MgCl_2$	-
5	0.006 mM	25 mM	16 mM $MgCl_2$	-
5	0.008 mM	25 mM	-	-
5	0.008 mM	25 mM	1 mM NaCl	-
5	0.008 mM	25 mM	10 mM NaCl	-
5	0.008 mM	25 mM	50 mM NaCl	-
5	0.008 mM	25 mM	1 mM $MgCl_2$	-
5	0.008 mM	25 mM	5 mM $MgCl_2$	-
5	0.008 mM	25 mM	1 mM $ZnCl_2$	-
5	0.008 mM	25 mM	5 mM $ZnCl_2$	-
5	0.008 mM	25 mM	5 mM $MgCl_2$	0.25 mM Spermine
5	0.008 mM	25 mM	5 mM $MgCl_2$	0.25 mM Spermidine
5	0.008 mM	25 mM	5 mM $MgCl_2$	0.25 mM $Co(NH_3)_6Cl_3$

The stability of a C-C mismatch is dependent on the solution environment as well as its sequence context. Peyret *et al.* have shown that in the sequence context ACT/TCA the mismatch exerts the greatest destabilization on the duplex and in the context CCG/GCC the mismatch displays its greatest stability (*i.e.* destabilizes the duplex to a lesser extent) (Peyret *et al.*, 1999). It has also been shown that C-C mismatches are considerably more stable at lower pH (Avizonis & Kearns, 1995; Battacharyya & Lilley, 1989; Brown *et al.*, 1990), presumably due to protonation at the N3 atom of cytosine, allowing for the formation of a more stable hemiprotonated C-C pairing interaction (Figure 2-9).

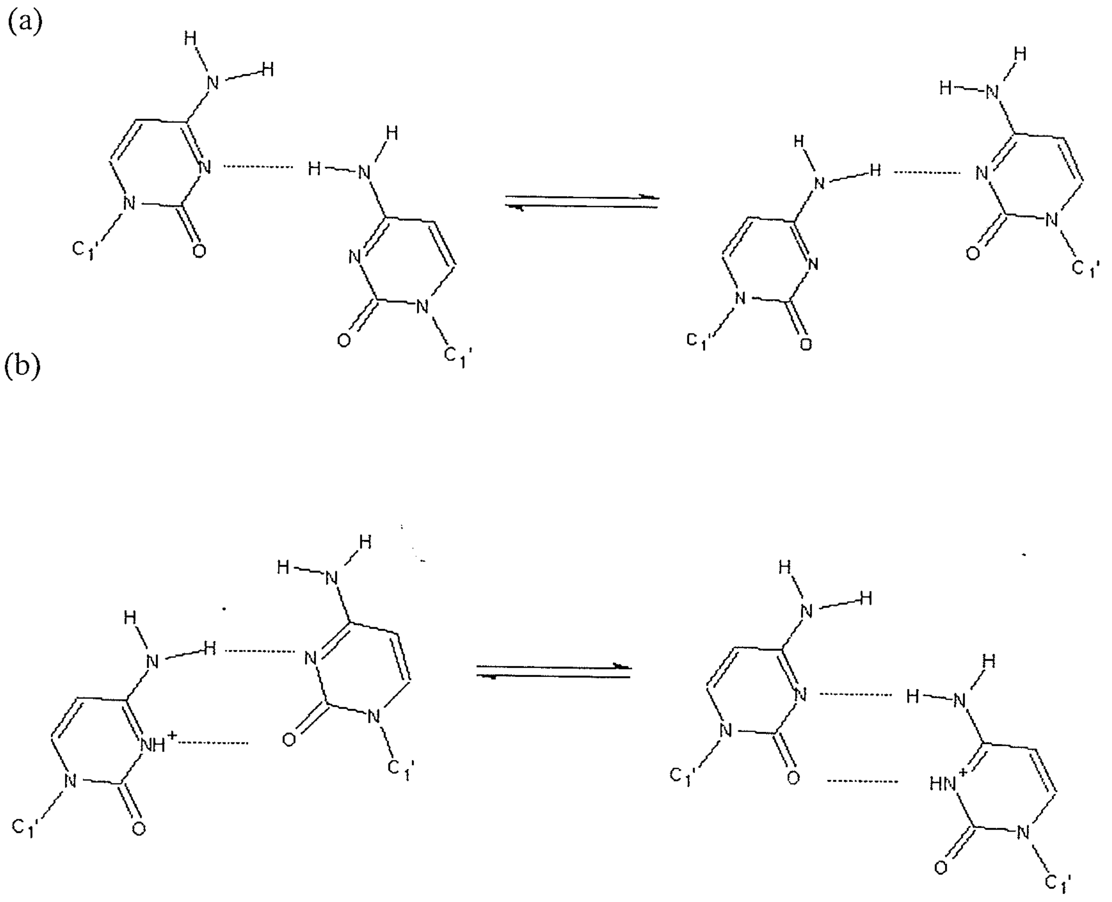

Figure 2-9 Possible C-C mismatch H-bonding as a function of pH.
(a) the neutral pH C-C mismatch and (b) low pH form of the C-C mismatch

The thermal denaturation of the 9mer was followed using UV spectroscopy. The UV thermal denaturation was carried out in a 25mM Na cacodylate, 16 mM $MgCl_2$ buffer at pH 5 and pH 7. The melting temperature was increased by approximately 3°C at the lower pH. This increase in stability is not as great as observed in other C-C mismatch containing sequences (CCA, Boulard *et al.*, 1997; CCA, Brown *et al.*, 1990); however, given the sequence context this was not unexpected. The effect of pH on the thermal denaturation profile of the 9mer sequence can be observed in Figure 2-10. The stabilization by 3°C is accompanied by a greater cooperativity of the denaturation as indicated by a sharper melting profile.

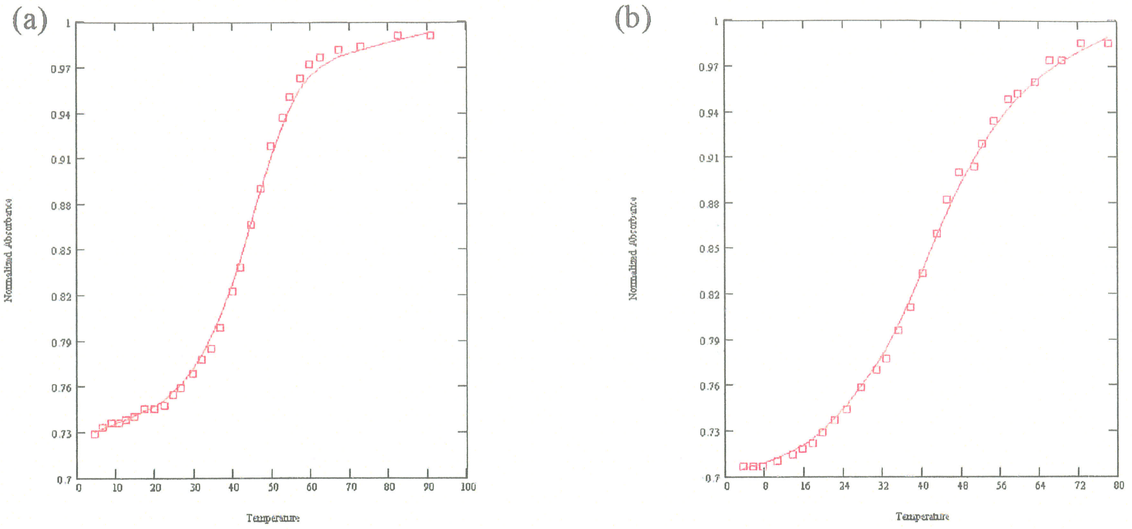

Figure 2-10 The effect of pH on the 9mer thermal stability.
(a) Thermal profile for 0.006 mM 9mer in 25 mM Na cacodylate, 16mM $MgCl_2$ pH 5 buffer and (b) thermal profile for 0.005 mM 9mer in 25 mM Na cacodylate, 16mM $MgCl_2$ pH7 buffer.

UV absorbance spectroscopy is the most common method for detecting nucleic acid structural transitions, however, it has been shown that some critical transitions are invisible to UV absorbance spectroscopy (Davis *et al.*, 1998). In order to ensure no other transitions were occurring during the thermal denaturation of the 9mer its thermal denaturation was also followed using CD spectroscopy. The CD thermal denaturation experiments were conducted in 25mM Na cacodylate 5mM $MgCl_2$ buffers of pH 5 and pH 7 with a DNA concentration of 0.008 mM. Spectra were collected from 10 to 60 ^oC, allowing the sample to equilibrate for 10 minutes at each temperature prior to data acquisition. Figure 2-11 shows only a single conformational transition as the 9mer duplex is denatured at both pH 5 and 7.

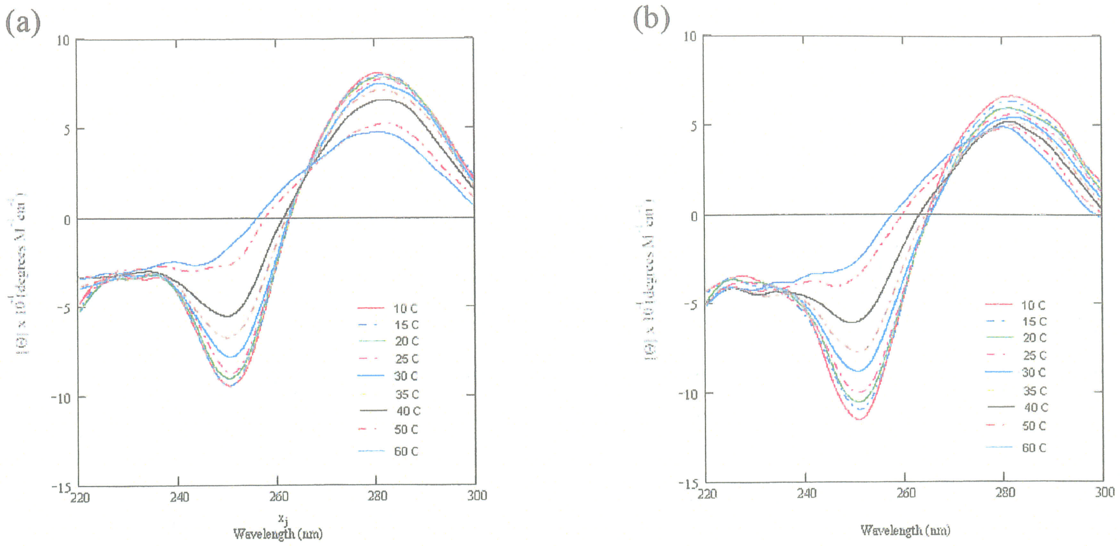

Figure 2-11 Temperature dependent CD spectra.
Spectra of 0.008 mM DNA in 25mM Na cacodylate, 5 mM $MgCl_2$ buffer (a) pH 5 and (b) pH 7.

Metal ions affect the conformational and structural behaviour of DNA. These conformational perturbations depend on type of metal involved, the metal ion concentration and, to a lesser extent, the DNA sequence. Thermal stability studies of other DNA sequences indicate that divalent cations confer stability at much lower concentrations than monovalent cations. It has also been shown that divalent cations that have an affinity for the DNA bases as opposed to the backbone can destabilize the DNA duplex. The effect of Na^+, Mg^{2+}, and Zn^{2+} binding on the C-C mismatch containing sequence were investigated.

The effect of increasing Na^+ can be observed in Figure 2-12. The base level of stability was obtained by melting the 9mer in a 25mM Na cacodylate pH 5 buffer in the absence of additional salts (mono- or divalent). A substantial increase in the melting temperature is achieved only at high Na^+ concentrations.

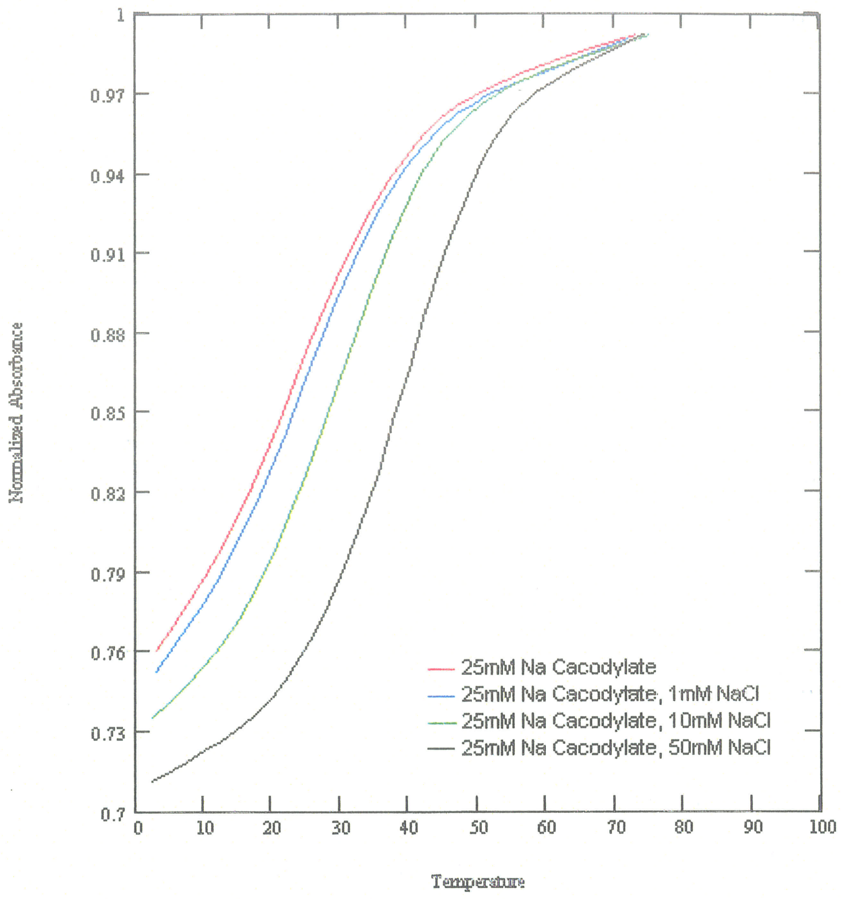

Figure 2-12 Melting curves for the 9mer in the presence of sodium cations. Normalized UV absorbance plots illustrating the effect of increasing sodium ion concentrations on the melting temperature of the 9mer at pH 5.

The effect of increasing the Mg^{2+} and Zn^{2+} concentrations can be observed in Figure 2-13 and Figure 2-14, respectively. Increasing concentrations of both of these inorganic cations stabilized the 9mer duplex. A substantial increase in the thermal stability of the 9mer is achieved even at low M^{2+} concentrations. A comparison of Figure 2-13 and Figure 2-14 reveals that at the concentrations of M^{2+} investigated the relative affinity of the cation for the DNA bases does not have an appreciable effect on the stabilization of the 9mer duplex.

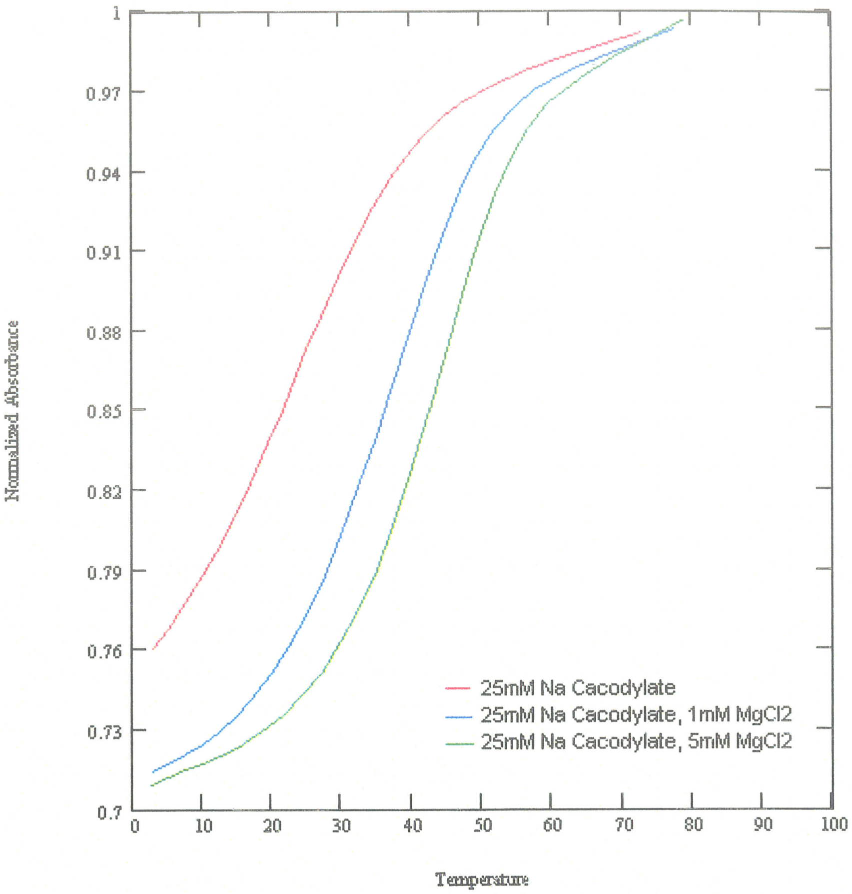

Figure 2-13 Melting curves of the 9mer in the presence of magnesium ions. Normalized UV absorbance plot illustrating the effect of increased magnesium concentrations on the melting temperature of the 9mer at pH 5.

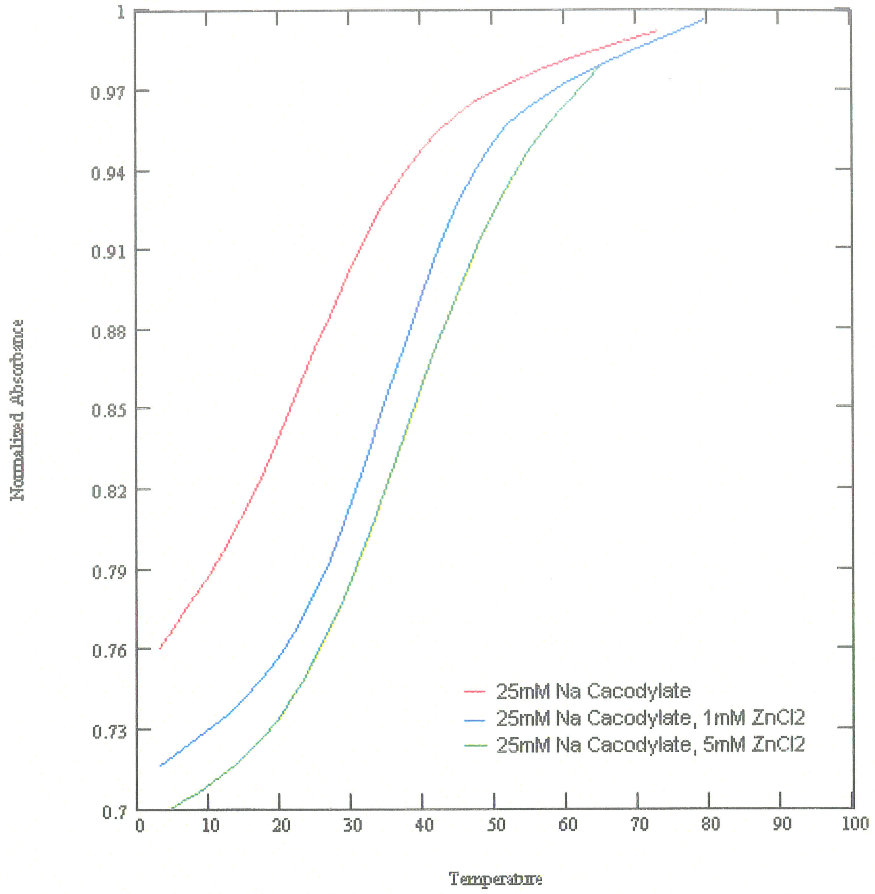

Figure 2-14 Melting curves for the 9mer in the presence of zinc ions.
Normalized UV absorbance plots showing the effect of increasing zinc concentrations on the thermal stability of the 9mer at pH 5.

Figure 2-15 shows the additive effects of the mono- and divalent salts on the thermal stability of the 9mer. The addition of low concentrations of magnesium or zinc salts to the 50 mM NaCl buffer results in a slight increase in the overall duplex stability.

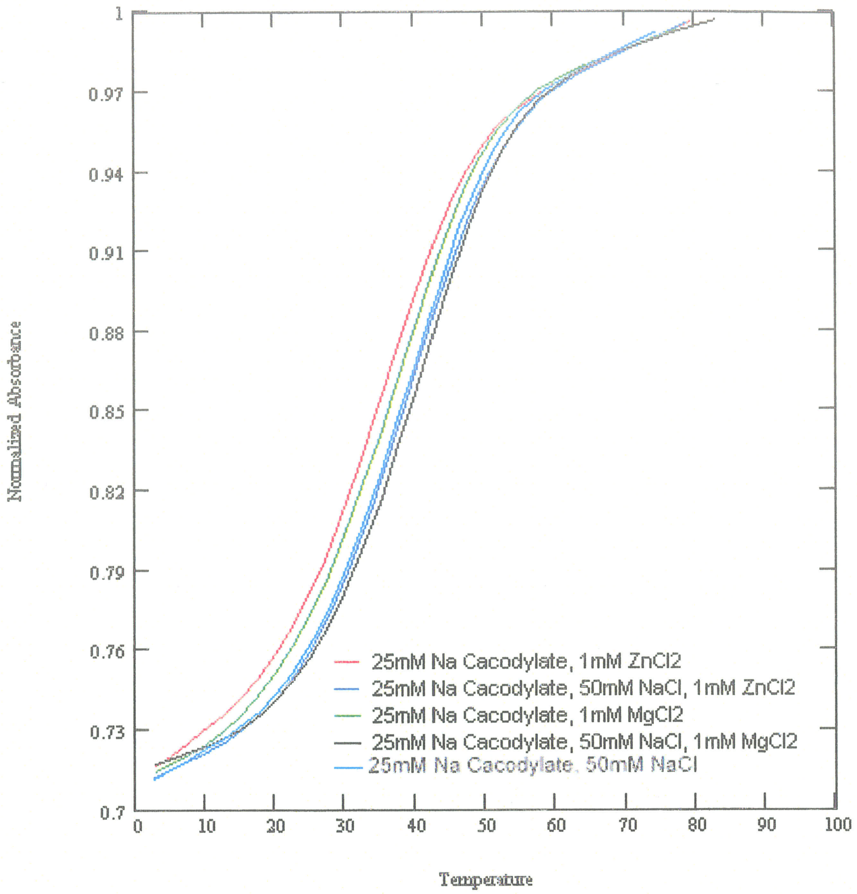

Figure 2-15 Combined cation effects.
Normalized UV absorbance plots illustrating the effect on the thermal stability of the 9mer sequence of the addition of 1mM $MgCl_2$ or 1mM $ZnCl_2$ to the 25mM Na cacodylate, 50mM NaCl pH 5 buffer.

Spermine and spermidine are polyamines that exist under biological conditions as polycations which can interact with and stabilize anionic cellular components like nucleic acids. The polyamines play a crucial, though not fully understood, role in a number of cellular processes, including cell growth and regulation, DNA replication, and chromatin condensation. *In vitro* they have been shown to increase the melting temperature of DNA and facilitate conformational changes, including DNA bending, aggregation, and induction of B → A and B→ Z transitions. Cobalt hexammine behaves similarly to the biologically relevant polyamines and is capable of promoting similar conformational transitions.

The effect of polyamines on the stability of duplex DNA is governed by ionic charge, pH, and the ionic concentration of the medium. The effects of spermine, spermidine and cobalt hexammine on the stability of the 9mer were investigated at pH 5.

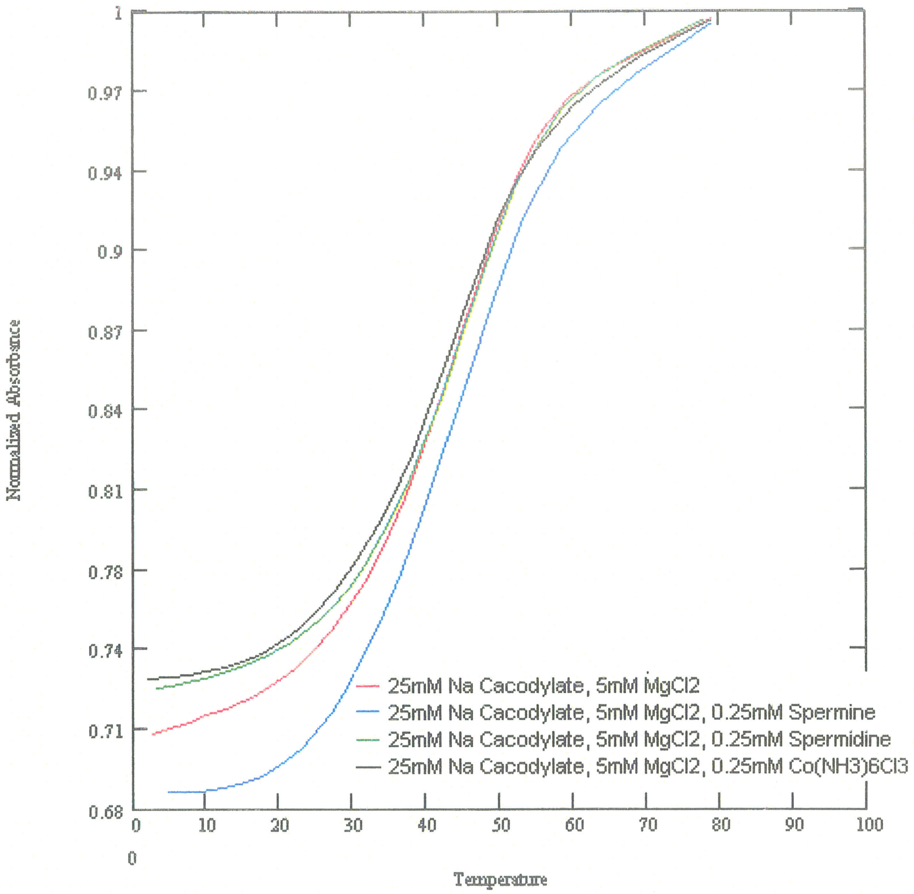

Figure 2-16 Effect of polyamine.
Normalized UV absorbance plots illustrating the effect on the thermal stability of the 9mer sequence of the addition spermine, spermidine and cobalt hexammine.

The effect of the addition of low concentrations of polyamine to the stability of the 9mer is illustrated in Figure 2-16. The polyamine concentration investigated was restricted due to aggregation of the DNA with spermine concentrations of greater than 1mM. From the thermal profile, it is clear that spermine is the most effective at stabilizing the 9mer.

2.2.2 12mer Solution Studies

UV thermal stability studies carried out by Mei Chen, a former student in this laboratory, on the 12mer in the presence of $MgCl_2$ indicated that the sequence adopts a duplex form with a melting temperature of ~70°C in a 25mM Na cacodylate, 16mM $MgCl_2$ pH 6.9 buffer (Chen, 1997). Crystals were obtained of the 12mer grown in the presence of zinc. It was unknown what, if any, effect zinc would have on the conformation of the 12mer. Preliminary analysis of the diffraction data indicated that the12mer had adopted a B-type conformation with the helix axis parallel to the *c*-axis. The length of the *c*-axis was, however, too short to accommodate the symmetry related helices stacking upon each other. In order to investigate the possibility that the zinc was inducing a hairpin conformation in the DNA the effect of zinc on the thermal stability of the 12mer was investigated.

Attempts to carry out the thermal denaturation of the 12mer in presence of zinc under conditions similar to the initial studies done by Mei Chen proved unsuccessful. When the 12mer is heated in a 25 mM Na cacodylate pH 7 buffer with zinc concentrations above 5mM, the DNA begins to aggregate at temperatures above 35°C. The thermal denaturation was successfully carried out in 25mM Na cacodylate, pH 7, with 1mM $ZnCl_2$ and $MgCl_2$ (Figure 2-17). The addition of zinc results in a slight destabilization of the helix, but the decrease in T_m is not sufficiently large enough to suggest a drastic change in DNA conformation.

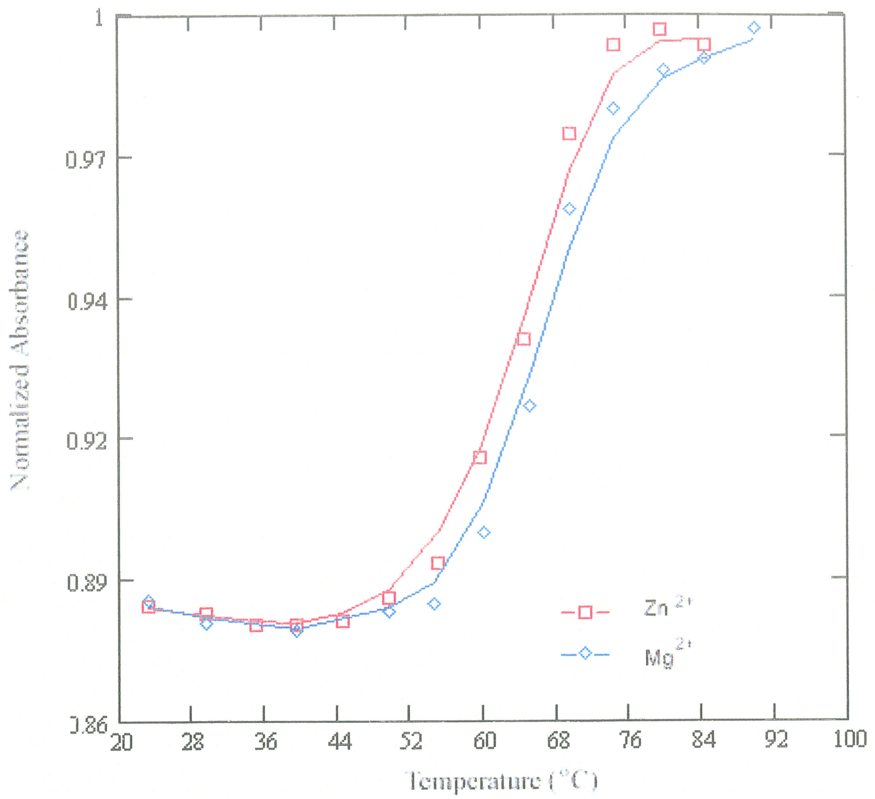

Figure 2-17 12mer thermal profile.
Normalized absorbance plot for the 12mer in 25mM Na cacodylate with 1mM $ZnCl_2/MgCl_2$

2.3 Crystallization Trials

All crystallization trials were conducted using the vapour diffusion method discussed in section "1.2.5"on page 30. In the crystallization of DNA oligomers the sample is an aqueous solution of DNA and the crystallization reagents consist of various aqueous cofactors and a water soluble non-volatile precipitant. The formation of crystals depends critically on a number of factors, including the relative concentrations of co-factors and DNA, the molecular properties and purity of the DNA and co-factors, rate of evaporation, temperature and pH to name a few.

The initial crystallization screening process is carried out over a large coarse grid of conditions. The reagents used in the screening process for all of DNA sequences studied are described below.

Buffers: The choice of buffer is influenced by a number of considerations — the buffer must be able to maintain the proper charged state of the DNA, it must be stable and it should not interact strongly with the DNA molecules. The pH of the crystallization drop is one of the variables that affects crystallization, thus it is critical that the pH of the droplet is controlled throughout the lifetime of the droplet. It is also critical that the pH of the crystallization media is near neutral otherwise the ionic form of the DNA bases may be altered or chemical degradation of the DNA may occur. Sodium cacodylate (dimethyl arsenate) was used as it is stable, only weakly interacts with DNA and has a favourable buffering strength in the pH range of 5 to 7.4 (pKa=6.19). It also has the added benefit of being an anti-microbial agent.

Metal Ions: Metal ions play an important role in biological functions associated with nucleic acids. Mono- and divalent cation salts are used to stabilize the negative charge on the phosphate backbone and are essential for maintaining the electrostatic neutrality of nucleic acids in solution. How the metal interacts with the DNA depends on its molecular properties, which is correlated to its location in the periodic table (see Figure 2-18). The salts used in the crystallization trials were sodium chloride, potassium chloride, barium chloride, calcium chloride, magnesium chloride, strontium chloride, copper chloride, cobalt(III) hexammine, cobalt(II) nitrate, manganous chloride, nickel chloride, and zinc chloride.

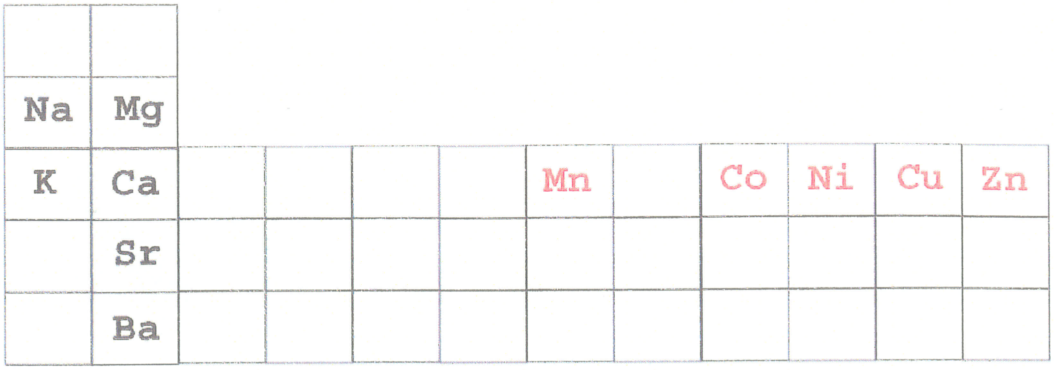

Figure 2-18 Metal ions interacting with nucleic acids.
The alkaline and alkali earth cations (shown in bold) interact with the sugar hydoxyls, phosphate oxygens on the DNA backbone, the base keto oxygen and ring nitrogen atoms. The transition metal cations (shown in red) interact (in their divalent state, except for Co which also interacts in its trivalent state) with the phosphate oxygens, the base keto oxygens, and ring nitrogen atoms. The preferential affinity for the base atoms versus the phosphate oxygens varies for transition metal cations as follows: $Cu^{2+} > Zn^{2+} > Mn^{2+} > Ni^{2+} > Co^{2+/3+}$, *i.e.* Cu has the greatest affinity for the base atoms and Co has the lowest affinity within this series of cations (Saenger, 1984).

Polyamines: Polyamines also play an important role in the function of DNA and can be used to stabilize the negatively charged DNA. Polyamines are able to stabilize compact nucleic structures with their multiple positive charges distributed over the surface of the molecule. The extended polycationic structure of the polyamines, unlike the compact nature of metal cations, allows them to interact with several DNA molecules at once. The polyamine used in the crystallization trials was spermine (see Figure 2-19).

Figure 2-19 Spermine structure.

Precipitating Agents: The precipitating agent used in the crystallizations was 2-methyl-2,4-pentanediol (MPD). MPD is an ideal precipitant because it is non-volatile and water-soluble, but sufficiently hydrophobic so as not to solvate DNA.

The crystallizations were set up using the hanging drop technique on VDX 24-well crystallization plates, with 1.7 x 1.6 cm wells, and siliconized 22 mm circle glass cover slides (Figure 2-20). The general technique consisted of placing a droplet (2 to 15 μL) of the DNA sample mixed with crystallization reagent on a siliconized glass cover slide which was then inverted and sealed with Dow Corning high vacuum silicone grease over a 500 μL reservoir solution (see Figure 2-21). Droplets were made by mixing an appropriate volume of a crystallization stock (containing mixtures of the crystallization reagents) and an aqueous solution of the DNA oligomer. In preparing crystallization experiments, it is critical that the results be reproducible. This is a very difficult task when working with the small volumes typically encountered in setting up a crystallization droplet. Reproducibility was improved by preparing large volumes (100 to 500 μL) of the crystallization stock from the stock reagent solutions (see Table 2-1). The volume of stock reagents required to achieve the desired crystallization stock were calculated using a spreadsheet program (EXCEL). Aliquots of the crystallization stock were then used in the crystallization set-ups. Typical DNA concentrations encountered in the crystallization droplets is 0.1-1 mM. The working DNA stock of 10-15 mM was diluted with dH_2O to obtain a concentration of 0.5-1.5 mM. This allowed reproducible volumes of ≥ 1 μL to be used in the crystallization set-up.

Figure 2-20 Crystallization supplies.

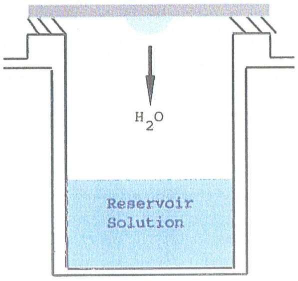

Figure 2-21 Vapour diffusion.
A schematic diagram of a vapour diffusion crystallization setup. A DNA-containing droplet is suspended from a glass cover slip and sealed over the reservoir. The concentration of precipitant is higher in the reservoir than in the droplet, causing the diffusion of water out of the droplet.

In the screening phase a large number of conditions needed to be surveyed. In order to reduce the amount of DNA material required, small (2 to 3 μL) hanging drops were utilized. The screening process involved setting up two-dimensional grids on the VDX plates with one crystallization parameter being varied in each direction. In the initial.stages of the screening process the concentration increments of variables were generally quite large (± 10-30% on the mean concentration). The results of the crystallization screens were observed under a stereo-zoom microscope fitted with a cross-polarizing attachment. DNA crystals examined under the microscope have true edges, while precipitate or dust particles do not. Crystals can appear as needles, blades, plates and various other geometric shapes. The use of crossed polarizers can indicate whether a material is crystalline. In a typical crystallization droplet there are often multiple crystals with different orientations. When these crystals are observed in the dark field, *i.e.* with the polarizers at 90° to each other, they will exhibit extinction as the crystal is rotated through 90° (this is true as long as the crystals are not cubic). Figure 2-22 shows typical examples of what one might observe in a crystallization experiment.

Conditions from the initial screen that resulted in mircocrystalline precipitate or phase changes (*e.g.* oil formation) were subjected to a second coarse grid search involving two new crystallization parameters (see Figure 2-23). Conditions resulting in crystalline material were subjected to finer grid searches (± 1-10% of the mean concentration) in order to optimize the conditions and obtain the best possible single crystals. After optimization of the crystal-yielding conditions, larger droplets (10 to 15 μL) were utilized for obtaining crystals for data collection.

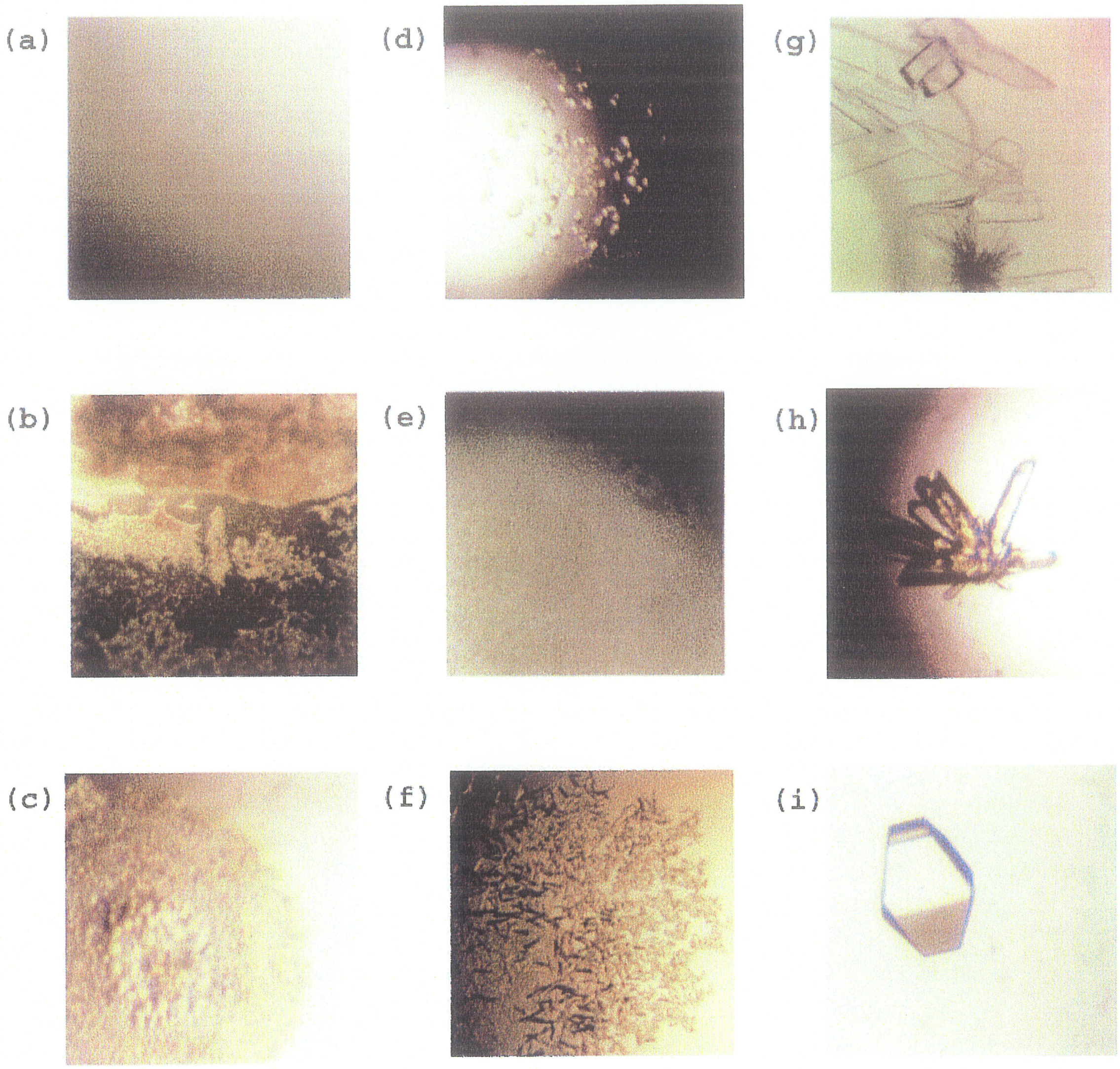

Figure 2-22 Examples of crystallization experiment results.
(a) clear droplet (b) precipitate (c) precipitate with oil/phase change (d) quasi crystals (e) microcrystalline precipitate (f) needles (g) plates (h) rod cluster (i) single crystal

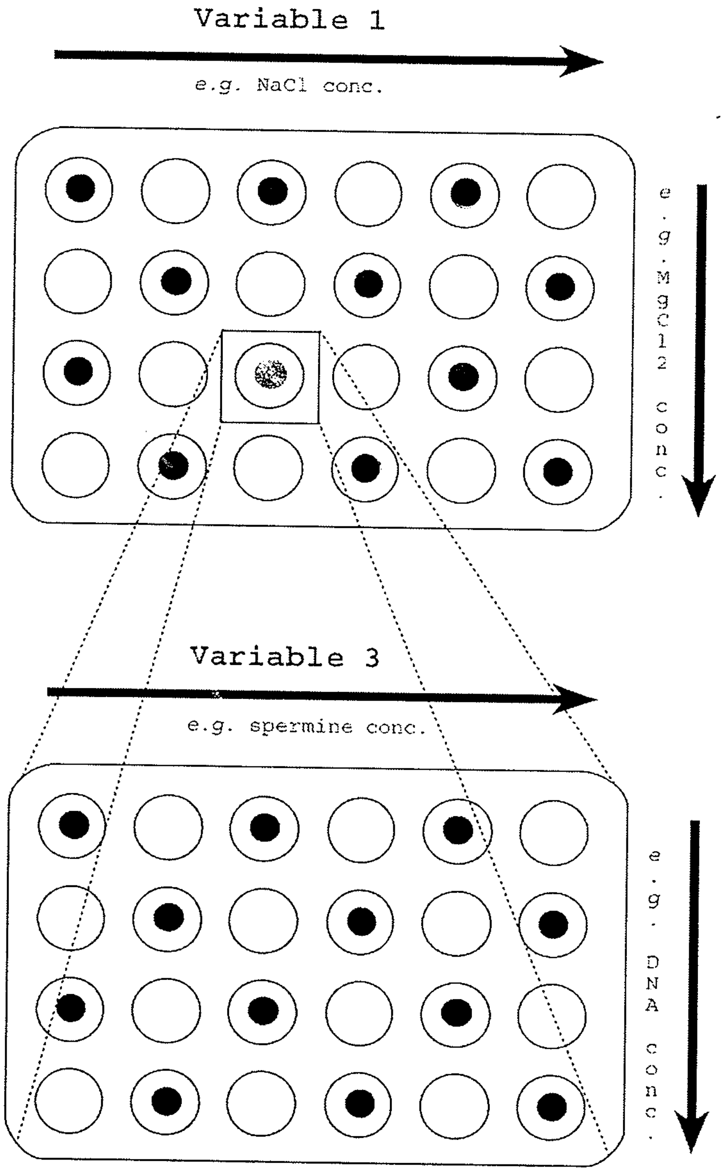

Figure 2-23 Crystallization grid screening.
A schematic representation of a crystallization grid screen on 24-well crystallization plates. Twelve crystallization trials are contained on a single plate.

2.3.1 Crystallization of d(CGGCCGCCG)

From the UV and CD results it was determined that the nonamer was stable over only a very narrow temperature range and at a low pH. As a result all crystallization trials were carried out at 9 °C and pH 5. Low concentrations of a divalent cation appear to be critical for stabilization of the duplex. The affinity of the divalent cation for the backbone or base atoms does not seem to affect the stabilization at the low concentrations encountered in a crystalization droplet. The addition of small quantities of spermine were found to increase the temperature range over which the duplex is stable as well as increasing the overall stability (as indicated by the sharper melting profile and higher melting temperature in Figure 2-16). It was also found that the duplex stability could be further increased, without threat of precipitation often encountered with the addition of high concentrations of divalent salts, by increasing the ionic strength with the addition of sodium salts.

The range of crystallization conditions investigated are summarized in Table 2-6. The crystallization droplets were set up using 1 µL of 0.25mM or 0.5 mM 9mer with 1 and 2 µL of the crystallization stocks. The droplets were equilibrated over a 17.5% MPD reservoir. To date no useable crystals have been obtained. The most promising results so far have yielded a microcrystalline precipitate (Figure 2-24).

TABLE 2-6 **C-C mismatch crystallization conditions.**

Variable	Ranges Investigated	Best result
[Spermine]	0.16 - 0.83 mM	0.5 mM
[Na Cacodylate]	16.7 mM	16.7 mM
[NaCl]	20 - 53 mM	40 mM
[$MgCl_2$]	0 - 20 mM	13.3 mM
[$ZnCl_2$]	0 - 11 mM	0 mM
droplet MPD	3 - 6.7%	6.7%
[DNA]	0.08 - 0.25 mM	0.08 mM
reservoir	17.5% MPD	17.5% MPD

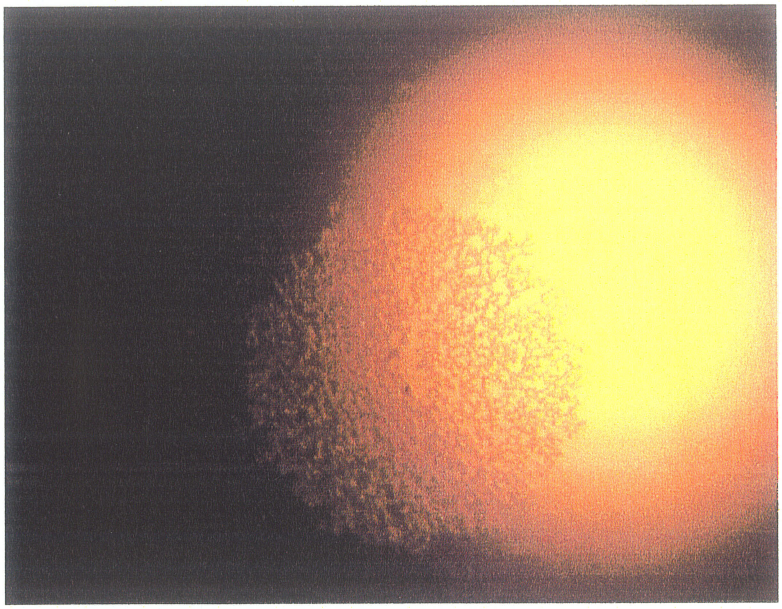

Figure 2-24 Photograph of microcrystalline precipitate of the 9mer.

2.3.2 Crystallization of the Dodecamer Sequences

2.3.2.1 Zinc Crystal Form

The thermal stability of the 12mer duplex, as indicated by its UV melting profile, allowed the crystallization trials to be carried out between 15 and 22 °C. The initial conditions for the 12mer sequence were based on results obtained by a former student in the laboratory, Mei Chen (Chen, 1997). The crystallization droplets were set up by mixing 6 μL of a 1.5 mM solution of DNA with 4.4 μL of crystallizations stock (12 mM spermine, 80 mM NaCl, 40 mM NaCac pH 7.0, 8% MPD, 20 mM $ZnCl_2$) and 2 μL of dH_2O. The droplet concentrations were 0.73 mM DNA, 4.26mM spermine, 28.4 mM NaCl, 14.2 mM NaCac, 2.8% MPD and 7.1 mM $ZnCl_2$. The droplet was equilibrated at 20 °C against 500 μL of 17.5% MPD. Small crystals (approximately 0.3 mm) of a triangular, prism-shaped habit were obtained within 2-4 days (see Figure 2-25). Despite their size and appearance, these crystals diffracted only to 3.4 Å. In an effort to improve the diffraction resolution limit better crystallization conditions were sought. Optimization of the crystallization conditions began with the variation of the zinc-to-DNA phosphate ratio (initial conditions had a ratio of 0.65). Ratios ranging from 0.5 to 2.0 were investigated with the best quality crystals being grown with a Zn:P ratio of 1:1 (corresponding to a $ZnCl_2$ crystallization droplet concentration of 11 mM).

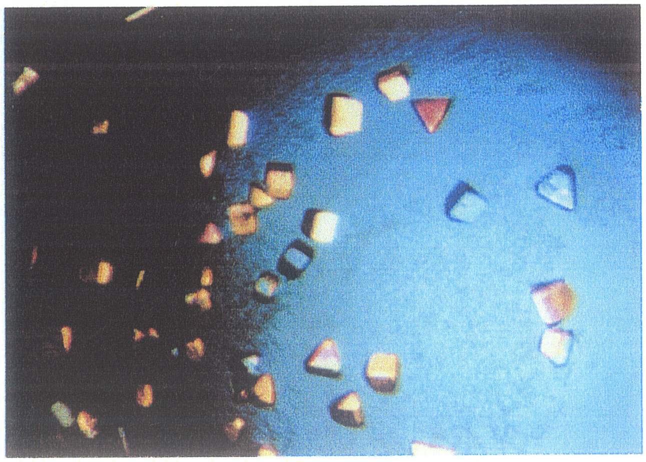

Figure 2-25 Photograph of crystals of the 12mer grown in the presence of zinc.

In order to facilitate crystal structure determination a 5-bromocytosine residue was incorporated into the DNA sequence, CGGTGG[Br^5C]CACCG (Br12mer). Crystals were grown of both the 12mer and Br12mer sequences under the 1:1 Zn:P crystallization conditions for data collection. The improved conditions did not result in an increase in the diffraction resolution which indicated further optimization of the crystallization conditions was required.

The interaction of zinc and other transition metal cations with DNA is believed to be pH dependent (Kejnovsky and Kypr, 1998; Lee & Latimer, 1993). It was reasonable, therefore, to investigate the effect of pH on the crystallization of the 12mer in the presence of zinc. Crystallizations were carried out using the 1:1 Zn:P crystallization conditions with pH values ranging from 6 to 8. The pH of the sodium cacodylate buffer was adjusted (with dilute NaOH or HCl) to achieve the desired pH prior to preparing the crystallization stock solution. While the crystals obtained at pH 7.5 had the best morphology, diffraction quality was unchanged. It was believed that the poor crystal quality was either a result of rotational disorder between the stacked DNA helices or the low MPD concentration (2.8%) in the droplet (high droplet precipitant concentrations have been linked with better quality crystals). The effect of the MPD concentration could be investigated simply by increasing its droplet and/or reservoir concentration. The rotational disorder problem required the design of a new DNA construct. The alternate DNA construct, GGTGGCCACCGC (GG12mer), was designed to have two overhanging base pairs in order to interlock the stacked helices and reduce the possibility of rotational disorder (Figure 2-26). With the 12mer sequence it was found that simply increasing the MPD concentration in the droplet lead to mass precipitation. In order to increase the MPD concentration a new coarse grid search of the zinc-containing crystallization conditions was required.

$^{5'}$GGTGGCCACCGC GGTGGCCACCGC$^{3'}$
$_{3'}$CGCCACCGGTGG CGCCACCGGTGG$_{5'}$

Figure 2-26 Schematic representation of the GG12mer construct.

The crystallization droplets were prepared by mixing 1-2 μL of crystallization stock with 1 μL of 0.5-2.0 mM DNA (either 12mer or GG12mer). The range of parameters investigated are listed in Table 2-7 & Table 2-8 as well as the optimal conditions found. Crystals of the 12mer grown in the presence of zinc under higher MPD concentrations had the same crystal habit as the initial zinc conditions, but took considerably longer to grow (1-2 weeks) and were of a larger size. Extended growth time indicates crystals likely have a greater internal order and thus are expected to diffract to a higher resolution. Data collection was attempted at BNL, however, the crystals did not freeze well and no usable data could be obtained. Room temperature data collection could not be carried out as all crystals were lost due to ambient temperature instability problems in the laboratory. Crystals of the GG12mer exhibited a different habit as shown in Figure 2-27. Unfortunately, these crystals were not stable in the droplet over extended periods of time. It is likely that further optimization of the conditions will be required before data can be collected.

TABLE 2-7 **12mer crystallization conditions in the presence of zinc.**

Variable	Ranges Investigated	Optimized Conditions	
[Spermine]	0.6 - 8.0 mM	4 mM	0.75 mM
[Na Cacodylate]	10 - 18 mM	12.5 mM	13.7 mM
pH	7.5 - 8.1	8.1	8.1
[NaCl]	14 - 41 mM	26 mM	28 mM
[$ZnCl_2$]	1.25 - 18 mM	6 mM	5 mM
[$MgCl_2$]	0 - 11 mM	7.5 mM	-
droplet MPD	3 - 13%	7.25%	10%
[DNA]	0.25 - 0.75 mM	0.54 mM	0.75 mM
reservoir	17 - 35% MPD 0 - 10% *iso*-propanol	17.5% MPD 5% *iso*-propanol	25% MPD

TABLE 2-8 **GG12mer crystallization conditions in the presence of zinc.**

Variable	Ranges Investigated	Optimized Conditions	
[Spermine]	2 - 5 mM	4.12 mM	3.5 mM
[Na Cacodylate]	12 - 17 mM	12.5 mM	12.5 mM
pH	8.1	8.1	8.1
[NaCl]	0 - 40 mM	25 mM	36 mM
[$ZnCl_2$]	3 - 15 mM	10 mM	7 mM
droplet MPD	3 - 13%	3%	3%
[DNA]	0.25 - 0.5 mM	0.25 mM	0.5 mM
reservoir	17 - 25% MPD	17.5% MPD	17.5% MPD

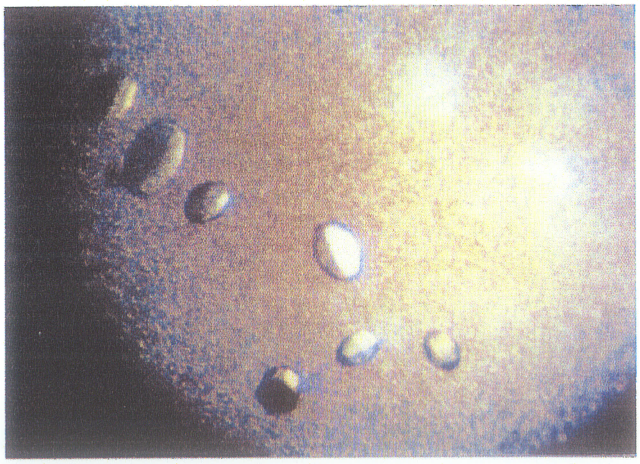

Figure 2-27 **Photograph of GG12mer crystals grown in the presence of zinc.**

2.3.2.2 Cobalt Nitrate Crystal Form

The possibility existed that the crystals grown in the presence of zinc would not yield high quality crystals, thus alternate starting conditions were sought. Specifically the effect of the divalent-cation type was investigated. The starting point for these trials consisted of droplet concentrations 0.73mM 12mer, 4.26 mM spermine, 28.4 mM NaCl, 14.2 mM NaCac pH 7.01, 2.8% MPD and 11 mM divalent cation salt (based on the 1:1 Zn:P conditions). The divalent cation salts used were $BaCl_2$, $MnCl_2$, $NiCl_2$, $CuCl_2$, $SrCl_2$, and $Co(NO_3)_2$. Promising results (*i.e.* oil/phase change, microcrystalline precipitate or quasi crystal formation) were observed for $BaCl_2$, $MnCl_2$, $NiCl_2$, and $Co(NO_3)_2$. Optimizations were carried out by varying the pH as well as the NaCl, divalent salt, spermine, DNA and MPD concentrations. Attempts to optimize the $NiCl_2$ resulted in oils, quasi crystals, or very poorly shaped crystals, none of which could be used for data collection. The $BaCl_2$ and $MnCl_2$ conditions provided very small needle-like crystals which were difficult to reproduce and unsuitable for data collection. The only conditions that provided suitably sized, reproducible crystals were the $Co(NO_3)_2$ conditions. The results of these crystallization trials are summarized in Table 2-9. The two crystal forms obtained are shown in Figure 2-28. The crystals obtained from the low MPD concentration conditions, despite their appearance, did not diffract past 8 Å. Attempts were made to collect data on the higher MPD conditions at BNL, however, the crystals did not freeze well and no usable data was obtained. Room temperature data were not collected as all crystals were lost due to ambient temperature instability problems in the laboratory.

TABLE 2-9 **12mer crystallization conditions in the presence of cobalt nitrate.**

Variable	Ranges Investigated	Optimized Conditions	
[Spermine]	0.25 - 6.5 mM	1.0 mM	1.0 mM
[Na Cacodylate]	14 - 18 mM	14.2 mM	18.3 mM
pH	6.5 - 8.1	6.5	8.1
[NaCl]	10 - 60 mM	28.4 mM	37.3 mM
[$Co(NO_3)_2$]	4 - 45 mM	16 mM	8 mM
droplet MPD	1.5 - 15%	5%	13.3%
[DNA]	0.25 - 0.72 mM	0.72 mM	0.5 mM
reservoir	17 - 25% MPD	17.5% MPD	25% MPD

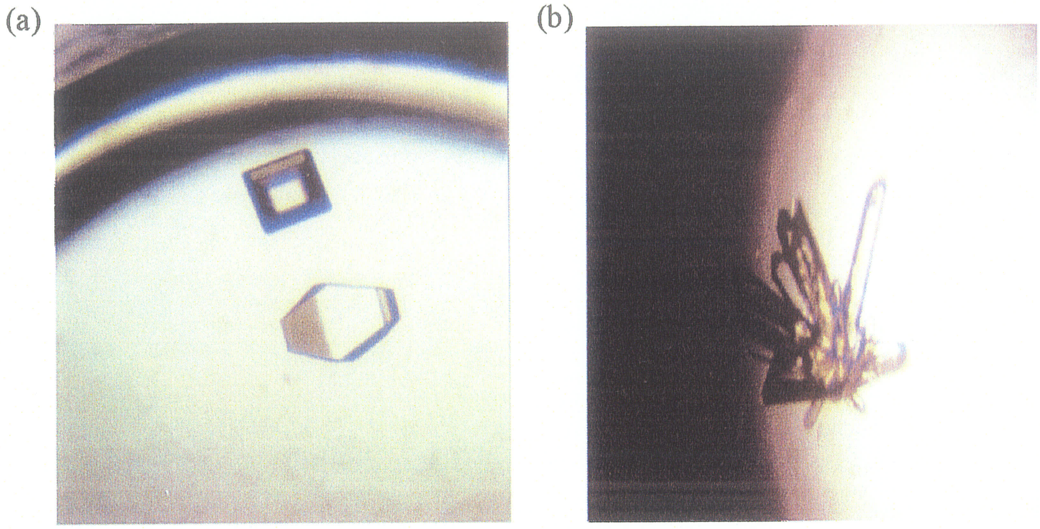

Figure 2-28 **Photographs of the 12mer co-crystallized with cobalt nitrate.** (a) low MPD crystal form and (b) high MPD crystal form.

2.3.2.3 Spermine and Cobalt Hexammine Crystal Forms

In order to increase the possibility of obtaining high quality crystals, an entirely new set of starting conditions was considered. The conditions were based on a 24-condition matrix designed for nucleic acids in general (Berger *et al.*, 1996) and the crystallization condition for the CGCGAATTCGCG B-DNA dodecamer (Wing *et al.*, 1980).

The 24-condition matrix is summarized in Table 2-10. The crystallization droplets were set up using 1 μL of 1.0 mM 12mer with 1 and 2 μL of the crystallization stocks. The droplets were equilibrated over a 25% MPD reservoir. Small crystals grew, over a period of 2 months, from the 1:1 mixtures with condition numbers 1, 3, 7, 21, 23, and 24 and from the 2:1 mixtures for condition numbers 3, 11 and 13. The crystal yielding conditions were broken down into two separate optimization matrices, those with spermine and those with cobalt hexammine. The variables investigated and the optimal conditions are summarized in Table 2-11 for the spermine conditions and in Table 2-12 for the cobalt hexammine conditions. Photographs of the crystals grown in the presence of spermine and those grown in the presence of cobalt hexammine are shown in Figure 2-29. Both crystal forms had hexagonal bar-like habits. The crystals grown in the presence of cobalt hexammine also exhibited spike-like growths on the ends of the crystal, similar to the "unicorn's horn" observed in the crystals of the non-self-complementary hexamer (Tari, 1995).

The crystallization trials based on the results of Drew and Dickerson (Wing *et al.*, 1980) were prepared by mixing 1.5 μL of 1.0 mM 12mer with 1.5 μL of crystallization stock (60 mM NaCac pH 7.0, 1.2 mM spermine, 30 mm $MgCl_2$, 20% MPD) which was then equilibrated over a 35% MPD reservoir. No precipitate was obtained even after an extended period of time (3 weeks) indicating a lack of supersaturation. The MPD concentration in the droplet was increased until precipitation occurred. After two weeks small crystals were found in droplets with an MPD concentration of 19%. The range of conditions investigated and the crystal-yielding conditions are listed in Table 2-13. The crystals grown had a crystal habit similar to those of the 12mer grown in the presence of spermine, but were much smaller (typically 0.1 x 0.05 x 0.05 mm^3).

TABLE 2-10 **24 condition matrix.**

Condition	pH	Polyamine	Monovalent Salt	Divalent Salt
1	7.0	12 mM Spermine	80 mM KCl	20 mM $MgCl_2$
2	7.0	12 mM Spermine	80 mM KCl	-
3	7.0	12 mM Spermine	80 mM NaCl	20 mM $MgCl_2$
4	7.0	12 mM Spermine	80 mM NaCl	-
5	7.0	12 mM Spermine	80 mM NaCl 12 mM KCl	20 mM $MgCl_2$
6	7.0	12 mM Spermine	12 mM NaCl 80 mM KCl	-
7	6.0	12 mM Spermine	80 mM KCl	20 mM $MgCl_2$
8	6.0	12 mM Spermine	80 mM KCl	-
9	6.0	12 mM Spermine	80 mM NaCl	20 mM $MgCl_2$
10	6.0	12 mM Spermine	80 mM NaCl	-
11	6.0	12 mM Spermine	80 mM NaCl 12 mM KCl	20 mM $MgCl_2$
12	6.0	12 mM Spermine	12 mM NaCl 80 mM KCl	-
13	7.0	12 mM Spermine	80 mM NaCl	20 mM $BaCl_2$
14	7.0	12 mM Spermine	80 mM KCl	20 mM $BaCl_2$
15	6.0	12 mM Spermine	80 mM NaCl	20 mM $BaCl_2$
16	6.0	12 mM Spermine	80 mM KCl	20 mM $BaCl_2$
17	7.0	12 mM Spermine	40 mM LiCl	80 mM $SrCl_2$ 20 mM $MgCl_2$
18	7.0	12 mM Spermine	40 mM LiCl	80 mM $SrCl_2$
19	7.0	12 mM Spermine	-	80 mM $SrCl_2$ 20 mM $MgCl_2$
20	6.0	12 mM Spermine	-	80 mM $SrCl_2$
21	5.5	20 mM $Co(NH_3)_6Cl_3$	80 mM NaCl	20 mM $MgCl_2$
22	5.5	20 mM $Co(NH_3)_6Cl_3$	80 mM KCl	20 mM $MgCl_2$
23	5.5	20 mM $Co(NH_3)_6Cl_3$	12 mM NaCl 80 mM KCl	-
24	5.5	20 mM $Co(NH_3)_6Cl_3$	40 mM LiCl	20 mM $MgCl_2$

TABLE 2-11 **12mer crystallization conditions in the presence of spermine.**

Variable	Ranges Investigated	Optimized Conditions	
[Spermine]	5 - 19 mM	16 mM	16 mM
[Na Cacodylate]	13 - 27 mM	26.7 mM	26.7 mM
pH	7.1	7.1	7.1
[NaCl]	30 - 66 mM	53.3 mM	46.7 mM
[$MgCl_2$]	6 - 20 mM	13.3 mM	13.3 mM
droplet MPD	7.5 - 16%	13.3%	13.3%
[DNA]	0.3 - 1.0 mM	0.5 mM	0.5 mM
reservoir	30 - 50% MPD	50% MPD	50% MPD

TABLE 2-12 **12mer crystallization conditions in the presence of cobalt hexammine**

Variable	Ranges Investigated	Optimized Conditions	
[$Co(NH_3)_6Cl_3$]	3 - 17 mM	10 mM	13.3 mM
[Na Cacodylate]	20 - 27 mM	26.7 mM	26.7 mM
pH	5.5 - 7.1	7.1	7.1
[KCl]	0 - 53 mM	34.7 mM	40 mM
[$MgCl_2$]	0 - 10 mM	3.3 mM	3.3 mM
droplet MPD	6.7 - 13.3%	6.7%	6.7%
[DNA]	0.17 - 0.5 mM	0.17 mM	0.33 mM
reservoir	20 - 35% MPD	25% MPD	25% MPD

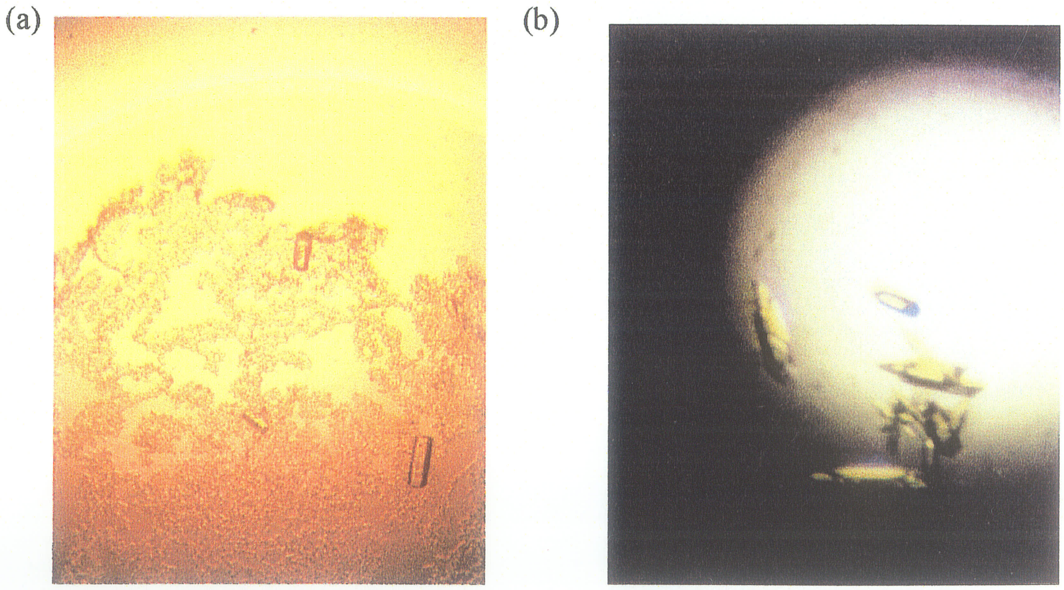

Figure 2-29 **12mer co-crystallized with spermine and cobalt hexammine.** Photographs of (a) the spermine crystal form and (b) the cobalt hexammine crystal form.

TABLE 2-13 **Drew and Dickerson based crystallization conditions.**

Variable	Ranges Investigated	Optimized Conditions	
[Spermine]	0.2 - 0.9 mM	0.28 mM	0.28 mM
[Na Cacodylate]	12 - 30 mM	8 mM	8 mM
pH	7.1	7.1	7.1
[NaCl]	0 - 8 mM	0 mM	0 mM
[$MgCl_2$]	2 - 18 mM	3 mM	4 mM
droplet MPD	10 - 25%	19%	19%
[DNA]	0.2 - 0.75 mM	0.2 mM	0.2 mM
reservoir	20 - 35% MPD	35% MPD	35% MPD

In order to facilitate an eventual crystal structure determination of the 12mer, crystals of the bromo derivative DNA, Br12mer, were also sought. The crystal producing conditions from the 24-condition matrix results of the 12mer sequence were used as a starting point. Only the spermine-based conditions yielded suitable crystals, see Table 2-14 and Figure 2-30 for results.

TABLE 2-14 **Br12mer crystallization conditions in the presence of spermine.**

Variable	Ranges Investigated	Optimized Conditions
[Spermine]	5 - 19 mM	6 mM
[Na Cacodylate]	13 - 27 mM	15 mM
pH	7.1	7.1
[NaCl]	30 - 66 mM	33 mM
[$MgCl_2$]	6 - 20 mM	8.25 mM
droplet MPD	7.5 - 16%	7.5%
[DNA]	0.3 - 1.0 mM	0.238mM
reservoir	30 - 50% MPD	50% MPD

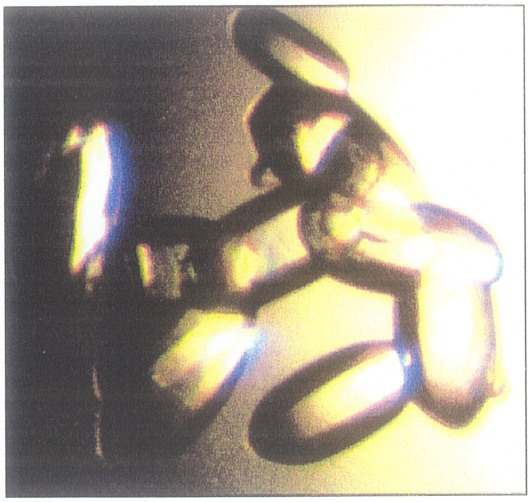

Figure 2-30 Crystals of the Br12mer grown in the presence of spermine.

2.3.3 Cryoprotection and Crystal Mounting

DNA crystals contain a very high percentage of solvent. With the loss of the crystal lattice solvent, the diffraction pattern deteriorates and no useful data can be collected. Mounting methods are therefore designed to maintain the solvated environment of the crystal as it is in the drop from which it grew (referred to as the mother liquor).

For room temperature-data collection the crystal can be mounted and sealed in a glass capillary with a small amount of the mother liquor. The glass capillaries (glass type 50, Hampton Research) are designed to have a low absorption coefficient for copper Kα radiation. These glass capillaries are extremely thin walled (0.01 mm), the length of the capillary (80 mm) has a well defined diameter (available in 0.1-1.0 mm diameters), and one end is sealed while the other has a funnel shape.

In crystal mounting the first step is to choose an appropriately sized diameter of capillary. This is achieved by measuring the crystal under the microscope and choosing the smallest diameter capillary that will accommodate the crystal. The capillary is then cut with a capillary cutting stone (25 x 25 x 0.65 mm, Hampton Research) so that both ends of the capillary are open. The funnel end of the capillary is then attached to a flexible piece of tubing which is attached to a syringe. The free end of the capillary is then placed into the crystallization droplet and a small amount of mother liquor is drawn up. The capillary is then removed from the droplet and a small amount of air is drawn in. The result is a small "slug" of mother liquor in the capillary which serves to keep the crystal in the droplet environment. The crystal can then be drawn into the capillary. The excess mother liquor surrounding the crystal must be wicked away (with sterile paper wicks, Hampton Research) to reduce the amount of background scattering due to solvent and to help prevent crystal slippage during data collection. The open end of the capillary is sealed with epoxy before the funnel end is trimmed and also sealed. The sealed capillary is secured in a brass mounting pin (3 mm in diameter an 12.5 mm in length, Hampton Research) with the crystal at an appropriate height above the pin (determined by the for particular X-ray diffractometer). If the crystal droplet volume is small, the mother liquor may evaporate before the crystal can be drawn up into the capillary. To increase the time available, mother liquor or reservoir solution can be added to the droplet. Sometimes the crystal requires dislodging from the cover slip, other crystals, or precipitate in the droplet. DNA crystals are very fragile, therefore any manipulation must be carried out with a very steady hand and fine precision tipped tools. The crystal manipulation is accomplished with acupuncture needles or rabbit whiskers as they are strong yet flexible enough so as not to damage the crystal.

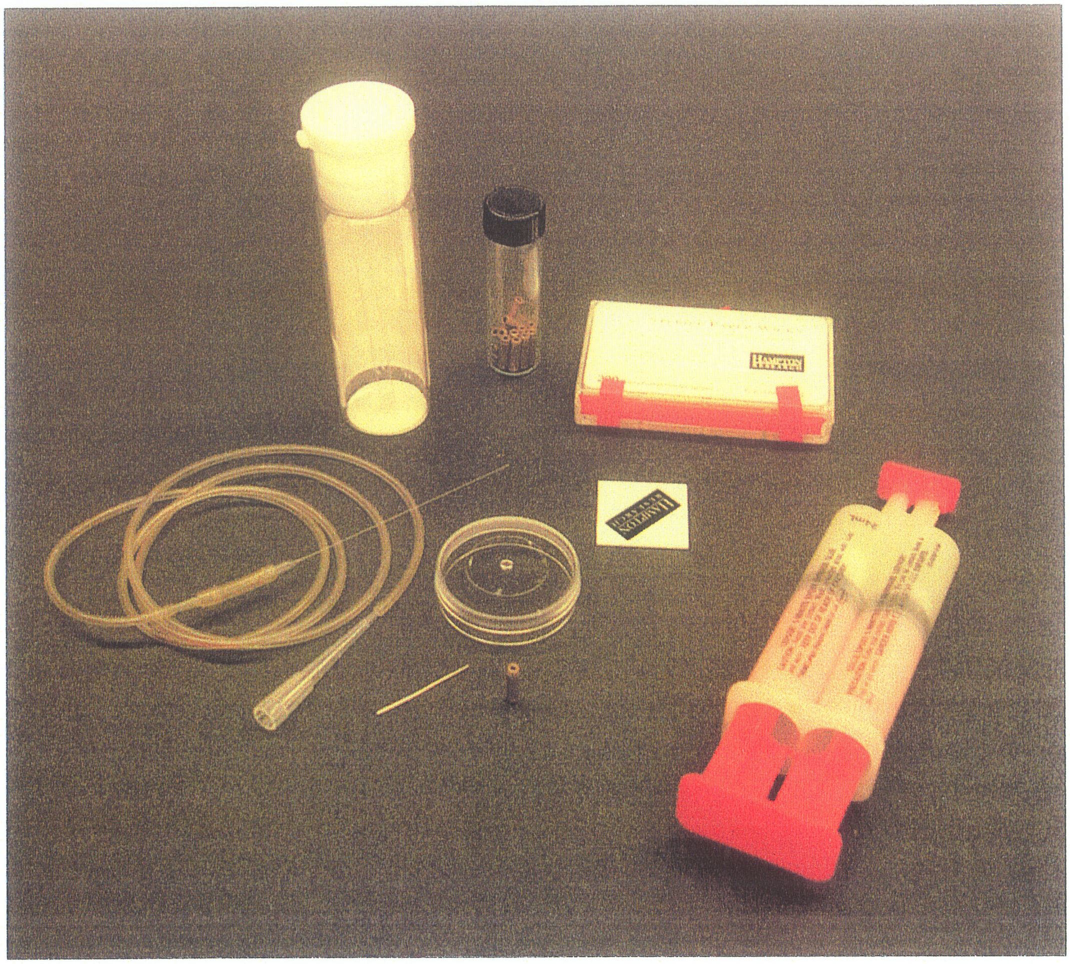

Figure 2-31 Glass capillary mounting supplies.

Data collection at cryogenic temperature (100 K) reduces the amount of radiation data and increases the lifetime of the crystal in the beam. However, simple freezing of the crystal results in ordered ice formation in the crystal which will interfere with the diffraction pattern or result in crystal cracking. To prevent this the crystal must first be cryoprotected. The cryosolutions were made by mixing a solution of twice the concentration of all droplet reagents (DNA excluded) with an equal volume of a cryoprotectant solution of appropriate concentration. The cryoprotectant used for all crystals was MPD. The concentration of MPD was chosen based on the crystal stability (more stable crystals are able to withstand higher MPD concentrations) and the ability to prevent ice formation. The cryosolutions used for the various crystal forms are summarized in Table 2-15.

TABLE 2-15 **Cryoprotection Conditions**

Crystal Form	Polyamine	NaCacodylate	Monovalent Salt	Divalent Salt	MPD
1:1 Zn:P 12mer/Br12mer	4.26 mM Spermine	14.2 mM pH 7.1	28.4 mM NaCl	11 mM $ZnCl_2$	30%
high MPD zinc 12mer	1 mM Spermine	18.3 mM pH 8	37.3 mM NaCl	5.3 mM $ZnCl_2$	33%
high MPD $Co(NO_3)_2$ 12mer	1 mM Spermine	18.3 mM pH 8	37.3 mM NaCl	8 mM $Co(NO_3)_2$	33%
Spermine 12mer	16 mM Spermine	26.7 mM pH 7.1	46.7 mM NaCl	13.3 mM $MgCl_2$	55%
Spermine Br12mer	6 mM Spermine	15 mM pH 7.1	33 mM NaCl	8.25 mM $MgCl_2$	55%
$Co(NH_3)_6^{3+}$ 12mer	13.3 mM $Co(NH_3)_6Cl_3$	26.7 mM pH 7.1	40 mM KCl	3 mM $MgCl_2$	47%
Drew&Dickerson 12mer	0.28 mM Spermine	8 mM pH 7.1	-	4 mM$MgCl_2$	48%

For cryogenic temperature data collection the crystals are suspended in a thin film of cryosolution on a thin nylon loop (20 μm fiber diameter) and flash frozen in liquid nitrogen. The nylon loops are mounted in hollow stainless steel microtubes (0.65 mm diameter, 24 mm in length with snap notches, Hampton Research). The microtubes, after being snapped off at the 10 mm mark, are glued into a copper pin.The copper pin is attached to a metal base/vented-cap system which allows the copper pin to be secured to a magnetic platform on the goniometer head or to be screwed into a 1.8 mL cryovial for storage and shipping (Hampton Research). Immediately prior to crystal mounting the crystals are cryoprotected by adding a small volume (20-40 μL) of the cryosolution to the crystal droplet. To facilitate flash freezing of the mounted crystal, the copper-pin assembly is attached to the end of a crystal wand (chrome plated steel wand 9.6 mm in diameter with a magnet molded inside one end, Hampton Research). The crystal is then fished out (with surface tension) of the cryosolution with the nylon loop of the copper-pin assembly and immediately plunged into liquid nitrogen —care must be taken to have as little cryosolution as possible on the loop to reduce the amount of background scattering. To protect the crystal from being disturbed the copper-pin assembly is secured in the cryovial (still under liquid nitrogen). The cryovials are mounted into cryocanes (aluminium holders, 28.7 cm in length, which will accommodate 5 cryovials) which are then placed in a nitrogen dry-shipper dewar for transport to the synchrotron. After the initial flash freezing of the crystal it is imperative that it remain at cryogenic temperatures or crystal deterioration will occur.

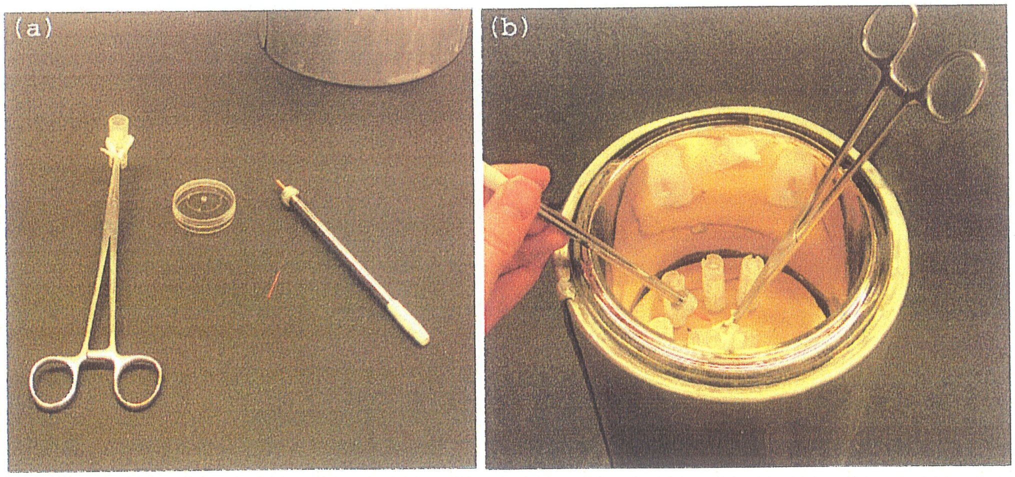

Figure 2-32 Cryogenic crystal mounting.
(a) The crystal is scooped out of the droplet with the nylon-loop-copper-pin-magic wand assembly and then (b) plunged into the liquid nitrogen where it is secured in the pre-cooled cryovial.

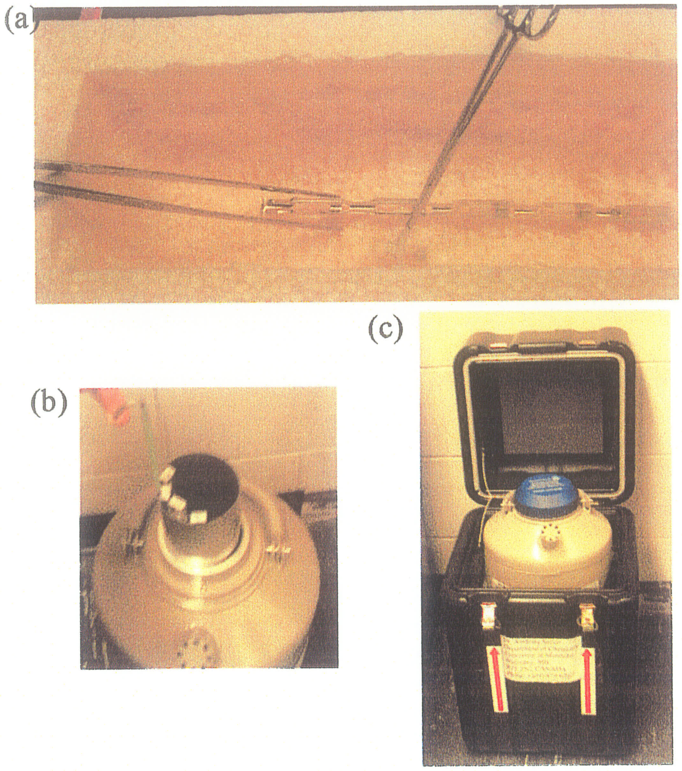

Figure 2-33 Cryocane loading and dry shipper.
(a) The crystal containing cryovials are loaded into the cryocane under liquid nitrogen, (b) the crystal containing cryocanes are loaded into a canister in the dryshipper, and (c) the dry shipper is loaded into the labelled shipping case.

2.4 Data Collection and Processing

2.4.1 Single Wavelength Data Collection

For all crystal forms the initial screening was carried out at room temperature on two separate trips to Calgary. Data sets collected at room temperature were carried out (by the author) in Dr. Leslie Tari's lab at the University of Calgary on a Rigaku RU-H3RHB 18kW rotating copper anode source (running at 50kV and 100mA, $\lambda = 1.54$ Å) and a MAR rotation camera with MAR345 image plate detector. The brass-pin mounted crystals were inserted in a Huber goniometer which was subsequently mounted on the diffractometer (Figure 2-34). The crystal was centered in the X-ray beam manually using a CCD video device. The data were collected using the oscillation method with a $\Delta\phi$ angle of 1° or 1.5°, an exposure time of 10-15 minutes and a crystal to detector distance of 180 mm. Data collection parameters for the crystals forms are listed in Table 2-16. Samples of the raw data (*i.e.* a single oscillation image) for each of the crystals forms are shown in Figure 2-35 to Figure 2-40.

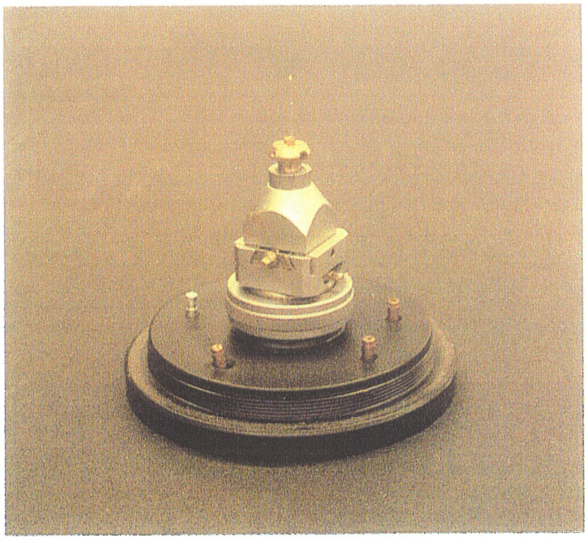

Figure 2-34 Capillary mounted crystal in a Huber goniometer head.

TABLE 2-16 Room temperature data collection parameters.
(a) a full data set could not be collected due to X-ray equipment failure

Crystal Form	Crystal Size	Exposure Time	$\Delta\phi$	ϕ range	Resolution Limit
original zinc	0.3 x 0.3 0.2 mm^3	-	-	-	3.4 Å
1:1 Zn:P pH 7 12mer	0.3 x 0.3 x 0.5 mm^3	10 minutes	1.5°	120°	3.4 Å
1:1 Zn:P pH 7 Br12mer	0.2 x 0.2 x 0.4 mm^3	10 minutes	1.5°	120°	3.4 Å
$Co(NO_2)_3$ low MPD	0.2 x 0.2 x 0.1 mm^3	15 minutes	1.0°	-	8.0 Å
spermine 12mer	0.1 x 0.1 x 0.2 mm^3	10 minutes	1.0°	102°	2.8 Å
$Co(NH_3)_6^{3+}$ 12mer	0.05 x 0.05 x 0.15 mm^3	10 minutes	1.0°	102°	2.8 Å
Drew & Dickerson 12mer	0.05 x 0.05 x 0.1 mm^3	10 minutes	1.0°	5°(a)	3.5 Å

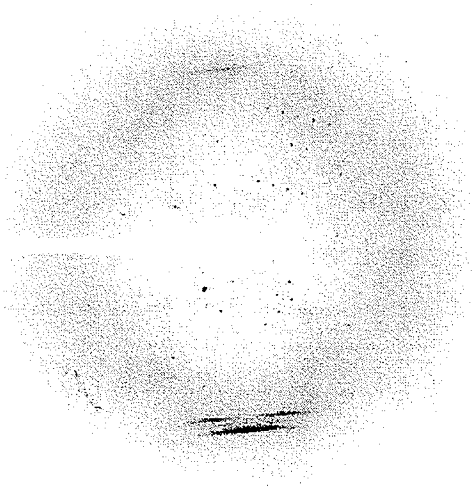

Figure 2-35 X-ray diffraction pattern for the 1:1 Zn:P 12mer crystal.

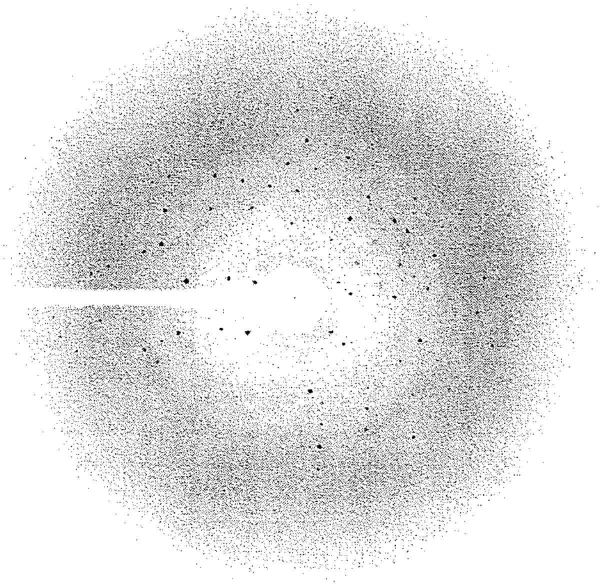

Figure 2-36 X-ray diffraction pattern for the 1:1 Zn:P Br12mer crystal.

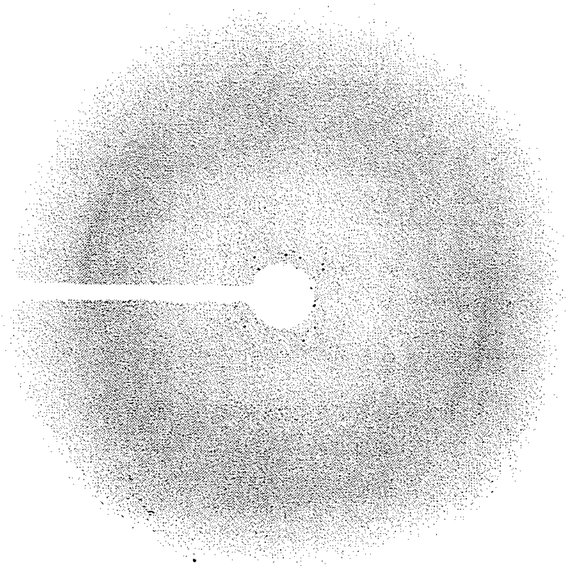

Figure 2-37 X-ray diffraction pattern for the low MPD $Co(NO_3)_2$ 12mer crystal.

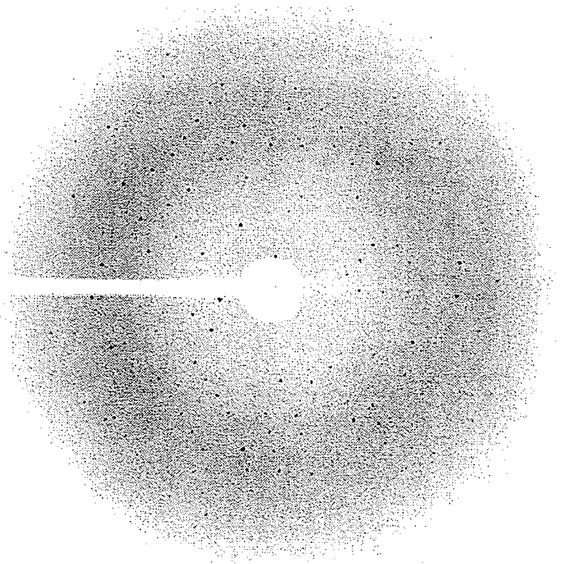

Figure 2-38 X-ray diffraction pattern for the spermine 12mer crystal.

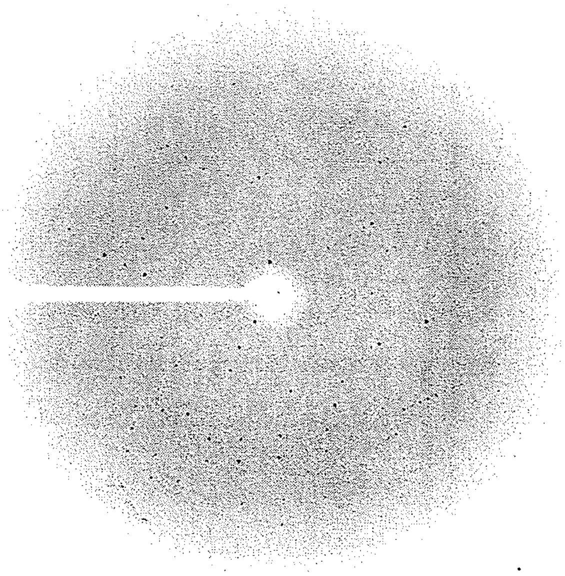

Figure 2-39 X-ray diffraction pattern for the cobalt hexammine 12mer crystal.

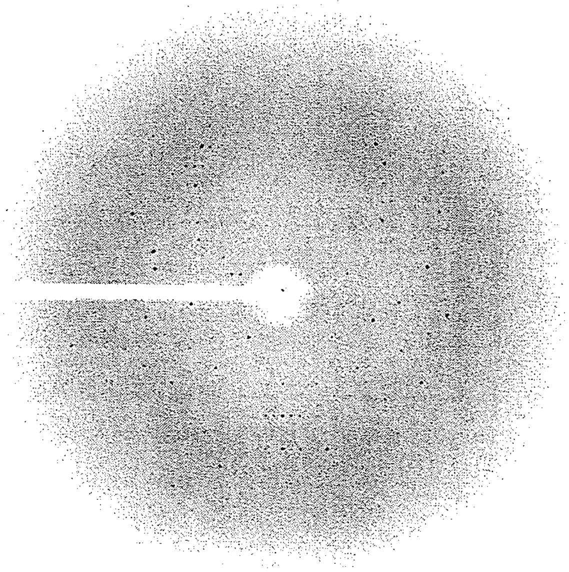

Figure 2-40 X-ray diffraction pattern for the Drew & Dickerson based crystal.

The data collection at cryogenic temperatures was carried out at the NSLS synchrotron (Long Island, NY) on beamlines X8C, X12B and X12C equipped with a CCD detector (either an ADSC Quantum 4 or Brandeis B1.3 detector, depending on the beamline), a cryostat (a low temperature device, Oxford Cryostream from Oxford Cryosystems) and a monochromator (recall synchrotron radiation is broad spectrum). The mounted, flash-frozen crystals were quickly removed from the cryovials (while submerged under liquid nitrogen) and then mounted on a Huber goniometer head (with a magnetic base support, Hampton Research) already in place on the diffractometer under the cold gas stream from the cryostat. The crystals were manually centered in the X-ray beam using a CCD video device (Figure 2-41). Like the room temperature data sets, these data sets were also collected using the oscillation method, the only difference being much shorter exposure times due to the high intensity of the synchrotron X-ray source. Data collection parameters for the various crystals forms are listed in Table 2-17.

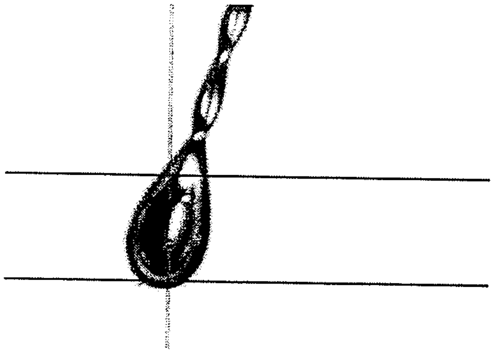

Figure 2-41 Centered flash-frozen loop-mounted crystal.

TABLE 2-17 **Cryogenic temperature data collection parameters.**

Crystal Form	**Exposure Time**	$\Delta\phi$	ϕ **range**	**Wavelength(s) & Beamline**	**Distance**	**Resolution Limits**
1:1 Zn:P pH 7.5 12mer	20 seconds	1°	120°	0.9789 Å X8C	200.0 mm	3.5 Å
1:1 Zn:P pH 7 Br12mer	10 seconds	1.5°	360°	0.9189 Å 0.9193 Å 0.9057 Å X8C	200.0 mm	3.5 Å
spermine 12mer	40 seconds	1°	168°	1.2847 Å X12B	120.0 mm	1.95 Å
spermine Br12mer	30 seconds 40 seconds 40 seconds	1°	333° 360° 339°	0.9236 Å 0.9240 Å 0.9116 Å X12C	99.6 mm	2.1 Å
$Co(NH_3)_6^{3+}$ 12mer	60 seconds	1°	180°	1.2827 Å X12B	100.0 mm	1.75 Å
Drew&Dickerson 12mer	60 seconds	1°	180°	1.2827 Å X12B	150.0 mm	2.7 Å
high MPD zinc 12mer	60 seconds	1°	-	1.2827 Å X12B	150.0 mm	~4 Å
high MPD $Co(NO_3)_2$	60 seconds	1°	-	1.2827 Å X12B	150.0 mm	~4 Å

2.4.2 Data Processing

The data collection results in a series of two-dimensional images which must be indexed, integrated, merged and scaled to provide a three-dimensional data set (the third dimension given by rotation of the ϕ angle) of structure factor magnitudes, $|\mathbf{F}(hkl)|$. In the oscillation method, a given diffraction image is recorded over an angular range. If the angular range is larger than the rocking curve (typically around 0.5°), a subset of the reflections is fully stimulated and a further subset is partially stimulated. The partially stimulated reflections, referred to as partials, will appear on contiguous images. The integration of the reflections from the raw data images requires knowledge of their index and position on the detector. While an approximate position for the strong reflections can be easily found by image-recognition software, the weak reflections can be found only by prediction based on the information from the strong reflections. The software suite HKL (Otwinowski & Minor, 1997) — used for processing all data sets collected in section 2.4.1— uses an autoindexing routine which finds a few dozen strong reflections to calculate the crystal cell parameters and the crystal orientation in the laboratory reference frame as well as generating indices for the reflections used. This information is then used to refine the crystal and detector parameters as well as to predict the diffraction pattern. Once the reflections can be accurately predicted for a diffraction image the pixels corresponding to a given reflection can be collected and the intensity of the reflection integrated from the images. The final step in the data processing involves scaling and merging the data from the individual images to form one complete data set. This step yields refined unit cell parameters, a mosaicity for the crystal as well as a number of statistics including the agreement between symmetry related reflection, R_{merge}. The data collection statistics for the various crystal forms are summarized in Table 2-18 & Table 2-19.

TABLE 2-18 **Room temperature data collection statistics.**

Crystal Form	Unit Cell	Space group	R_{merge}	Mosaicity	Completeness	Multiplicity
original zinc	a=b=46.67 c=101.67 α=β=90° γ=120°	R32	5.7%	0.7°	99.2%	3.1
1:1 Zn:P pH 7 12mer	a=b=47.0 c=101.43 α=β=90° γ=120°	R32	6.7%	0.68°	94%	2.9
1:1 Zn:P pH 7 Br12mer	a=b=46.99 c=102.1 α=β=90° γ=120°	R32	6.6%	0.8°	91.4%	4.7
spermine 12mer	a=b=41.85 c=78.58 α=β=90° γ=120°	$P6_122$	3.9%	0.46°	98.5%	9.3
$Co(NH_3)_6^{3+}$ 12mer	a=b=41.81 c=78.67 α=β=90° γ=120°	$P6_122$	7%	0.27°	96%	9.1
Drew & Dickerson 12mer	a=b=42.2 c=78.1 α=β=90° γ=120°	$P6_122$	-	-	-	-

TABLE 2-19 **Cryogenic temperature data collection statistics.**

Crystal Form	**Unit Cell**	**Space group**	**R_{merge}**	**Mosaicity**	**Completeness**	**Multiplicity**
1:1 Zn:P pH 7.5 12mer	a=b=46.5 c=99.88 α=β=90° γ=120°	R32	7.3%	0.69°	97.1%	5.6
1:1 Zn:P pH 7 Br12mer	a=b=47.0 c=101.99 α=β=90° γ=120°	R32	8%	0.8°	96%	6.5
spermine 12mer	a=b=41.789 c=78.112 α=β=90° γ=120°	$P6_122$	5.3%	0.74°	98%	16.5
spermine Br12mer	a=b=41.23 c=78.55 α=β=90° γ=120°	$P6_122$	11%	0.42°	99%	17.1
$Co(NH_3)_6^{3+}$ 12mer	a=b=41.479 c=78.186 α=β=90° γ=120°	$P6_122$	5.3%	0.61°	98%	19.2
Drew&Dick-erson 12mer	a=b=41. c=78.112 α=β=90° γ=120°	$P6_122$	9.4%	1.5°	98%	14.3

2.4.3 The MAD Experiments

In a MAD experiment it is necessary to know precisely the wavelength at which the data is collected. The first step in a MAD data collection is to calibrate the monochromator using X-ray absorption spectra. The monochromator is calibrated by centering on the absorption edge of a known sample (usually a foil or film of a pure element) near the expected wavelength of the absorption edge of the heavy atom in the crystal sample. For both MAD experiments the heavy atom in the crystal sample was bromine which has an absorption edge at approximately 0.92 Å. The calibration element used was selenium (as a film) which has a very sharp absorption edge at 0.979 Å. After calibration an absorption spectrum of the crystal must also be taken in order to choose the best wavelengths for data collection.

2.4.3.1 Br12mer Zinc Crystal Form

The data collection was carried out according to the parameters in Table 2-17 for the 1:1 Zn:P pH 7 Br12mer crystal form. The wavelengths for data collection were chosen based on the X-ray fluorescence scan (Figure 2-43). A full 360° of data were collected for each of the three wavelengths, 0.9189, 0.9193 and 0.9057 Å, corresponding to f''max, f''inflection (which corresponds with f'min) and a remote energy respectively. The data statistics are summarized in Table 2-19. The expected phasing signal is resolution dependant as shown by Figure 2-42. Since the crystal diffracted only to 3.5 Å resolution, the maximum phasing signal (*i.e.* the difference between Friedel pairs) is only about 5%. In order to be able to measure such a small difference the internal consistency of the data needs to be better than 5%. However, the R_{merge} for the data set was 8% and thus no useful phasing information could be gained from the experiment.

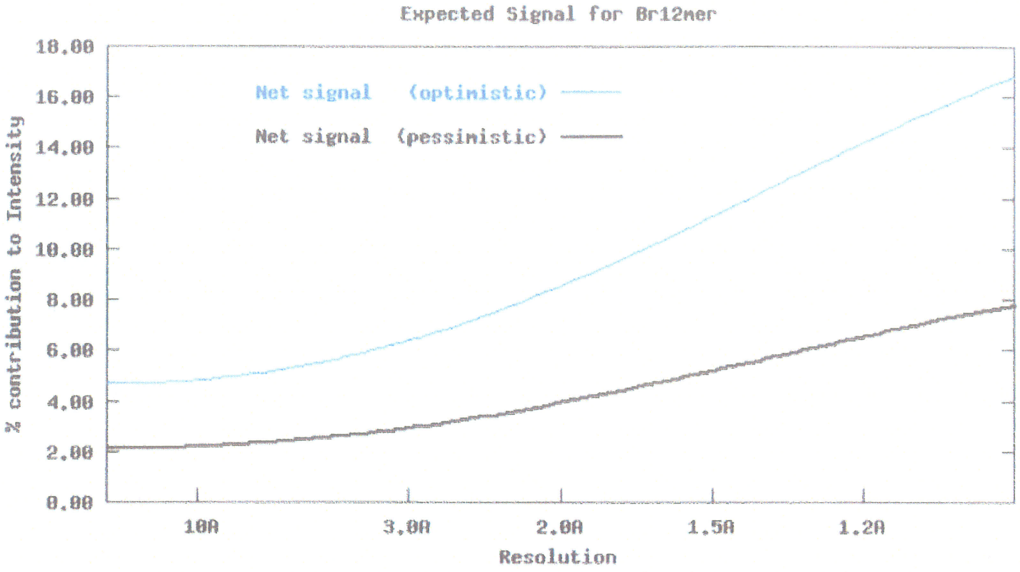

Figure 2-42 Estimate of signal from a bromine MAD experiment.
Signal estimate (see section section 1.2.7.2) based on 1 bromine atom per 12 nucleic acid residues and experimental scattering factor estimates of $f'_{min} = -8$, $f'_{max} = 0$ and $f''_{max} = 4$. In the pessimistic scenario 60% of the optimal f values are achieved.

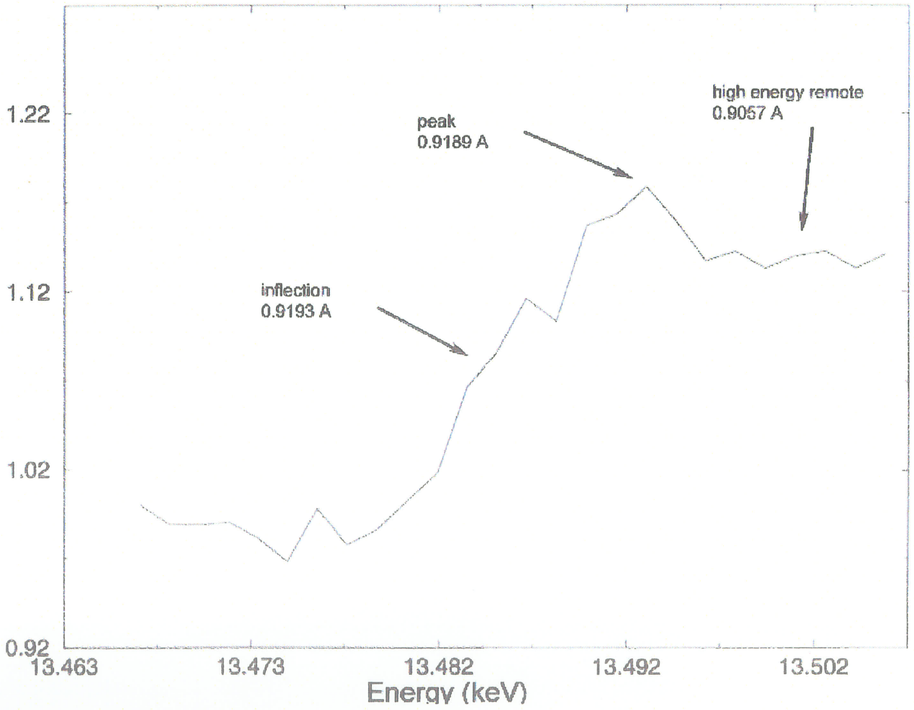

Figure 2-43 Fluorescence scan of Br12mer zinc crystal.

2.4.3.2 Br12mer Spermine Crystal Form

The data collection was carried out according to the parameters in Table 2-17 for the spermine Br12mer crystal form. From X-ray fluorescence scan (Figure 2-44a) and subsequent processing with CHOOCH (Figure 2-44b, Evans & Pettifer, 2001) — which calculates the wavelength dependent values of *f'* and *f"* — the wavelengths for data collection were chosen. Data sets were collected at wavelengths of 0.9198 Å (peak *f"*), 0.9202 Å (inflection of *f'*), and 0.9078 Å (high energy remote). The data collection statistics are summarized in Table 2-19.

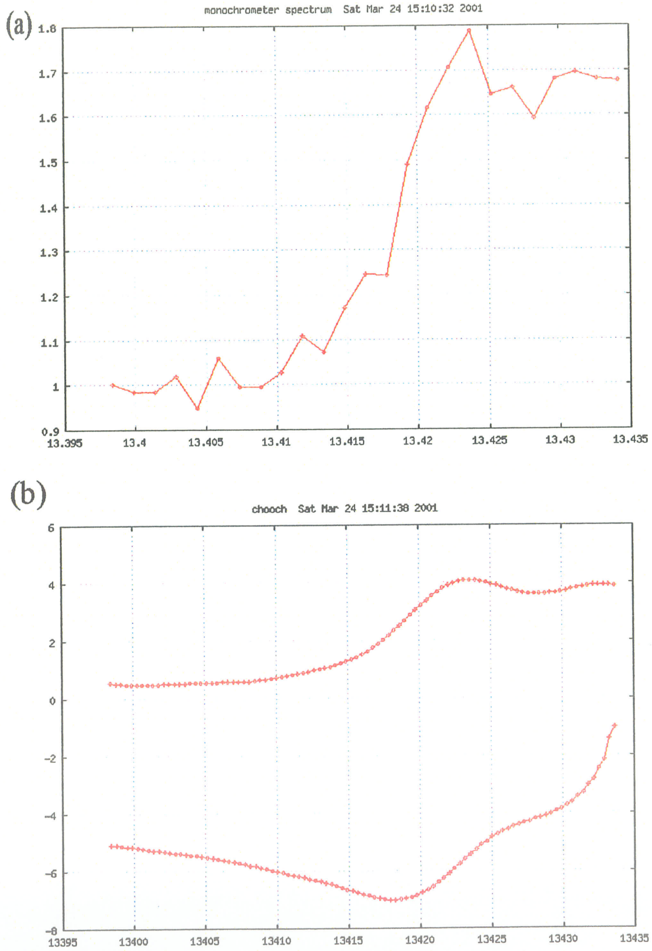

Figure 2-44 Spermine Br12mer MAD experiment setup.
(a) Fluorescence scan and (b) CHOOCH plot of the Br12mer spermine crystal form for the determination of the data collection wavelengths.

2.5 Structure Solution

2.5.1 Zinc Crystal Form

From the unit cell dimension and space group of the first zinc data set, the volume of the asymmetric unit was calculated to be 10859 Å^3. From the value of 1300 - 1600 Å^3 per base-pair, commonly observed in A- or B-DNA oligonucleotide crystals, it was apparent that the asymmetric unit contained a single dodecamer strand. The presence of diffuse 3.4 Å scattering as well as intense 3.4 Å reflections in the data set indicated that the DNA was likely in a B-type conformation. However, a Patterson map of the data (room temperature data set) did not show the expected "pancakes" of Patterson density separated by 3.4 Å which correspond to the base stacking in a B-DNA helix (Baikalov & Dickerson, 1998). Attempts were made to solve the structure using molecular replacement. A- and B-DNA models of the 12mer sequence were adapted from the fiber data of Arnott (Chanderasekaran & Arnott, 1989) to be used as the search models. The best result for the rotation solution was obtained with a B-DNA helix. However, the solution indicated that the B-DNA helix was parallel to the **c** axis of the unit cell. No reasonable translation solution, however, could be found. With the failure of molecular replacement an attempt was made to use heavy atom methods. The data collected from the 12mer and Br12mer zinc crystals were used to calculate a difference Patterson map. This map was expected to yield vectors corresponding to the interaction of the bromine atoms with the contents of the unit cell. The resulting Patterson map was, however, completely uninterpretable. Two possibilities existed: (*i*) the 12mer and Br12mer crystals were not completely isomorphous and (*ii*) the crystals were disordered. A MAD experiment was conducted to exclude the first possibility. Unfortunately, the zinc crystals do not freeze well and the resulting diffraction data was too poor to yield any phasing information.

It is believed that the improved crystallization conditions (higher MPD content crystals or the GG12mer crystals) will yield higher quality diffraction data that should enable structure solution. However, data has not yet been collected on these crystals due to time constraints.

2.5.2 Spermine Crystal Form

From the unit cell dimension and space group of the room temperature data set, the volume of the asymmetric unit was calculated to be 9845 Å^3. Based on the expected volume per base-pair, it was determined that the asymmetric unit contained a single dodecamer strand. The absence of diffuse 3.4 Å scattering and intense 3.4 Å reflections indicated that the DNA was likely in an A-type conformation. The Patterson map of the room temperature data (see Figure 2-45) closely resembled that of another A-DNA dodecamer crystal (Bingman *et al.*, 1992). A single stranded A-DNA 12mer model was adapted from the fiber data of Arnott (Chanderasekaran & Arnott, 1989) to be used as the search model. The rotation function yielded an orientation for the dodecamer, but attempts to find the correct position were unsuccessful.

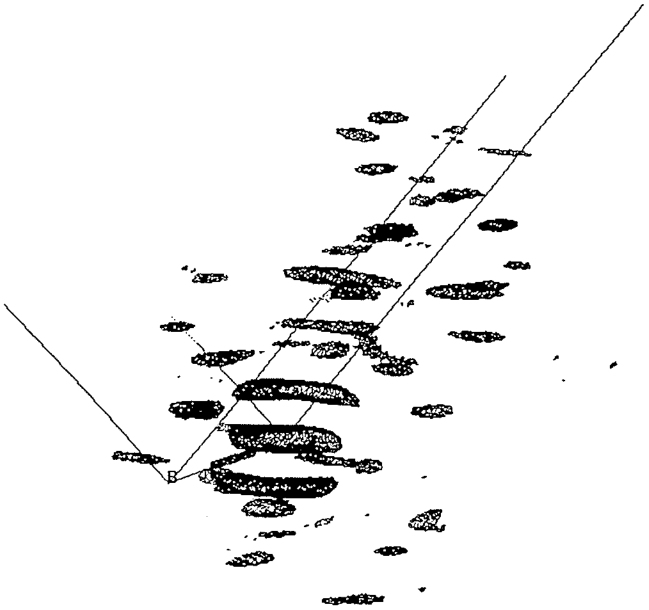

Figure 2-45 Native Patterson map of the spermine 12mer room temperature data.

When crystals of the Br12mer were obtained under the spermine conditions it was decided that a MAD experiment would be used to carry out the primary phasing. With the asymmetric unit consisting of a single dodecamer strand, only a single bromine atom position needed to be found. The program SOLVE/RESOLVE (Terwilliger & Berendzen, 1999) was used to find the position of the bromine atom and to calculate the initial phases for the data. The resulting electron density map, with no DNA structural input, clearly showed 11 of the 12 DNA bases (see Figure 2-46).

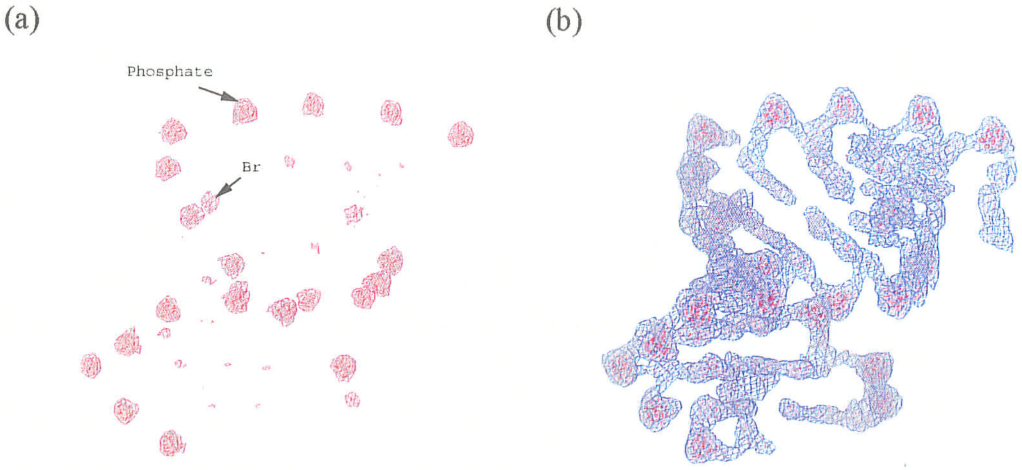

Figure 2-46 MAD phased electron density maps.
contoured at (a) 4σ and (b) 1.5σ sigma with the program xfit (McRee, 1999).

From the electron density map contoured at 4σ, the positions of the DNA backbone phosphates and the bromine atom on residue 7 were clearly visible (see Figure 2-46a). The electron density map contoured at 1.5σ revealed that the central 10 base-pairs had adopted a primary A-type conformation with the 3'- terminal guanine residue flipped away from the helix axis (see Figure 2-46b). It was unclear at this point where the remaining 5'-terminal cytosine was located.

The goal at this stage was to build a crude model of the 12mer into the existing electron density map. A single stranded 10 base model (comprising residues 2-11) of the sequence was manually built from an A-DNA decamer structure that had also crystallized in the space group $P6_122$ (NDB code adj065, Gao *et al.*, 1995). When the 10 base model was crudely positioned in the electron density map it became clear that residues 2 and 11 would need to be removed and manually rebuilt once the fit of the central 8 bases with the electron density had been improved (Figure 2-47). The position of the central 8 bases was improved using the rigid body refinement routine in the CNS program suite (Brunger *et* *al.*, 1998) with the experimentally phased data from the MAD experiment and the native data (collected at BNL) from the spermine crystal. The resolution of data included in the refinement was from 2.5 - 20 Å. The refinement target was a maximum likelihood function modified to make use of the experimental phase information in the calculation of the estimate of model error. The rigid body refinement improved the R-factor from 59% to 53%. The high R-factor at this point is due to the fact that 4 of the 12 bases in the structure have not yet been included in the model.

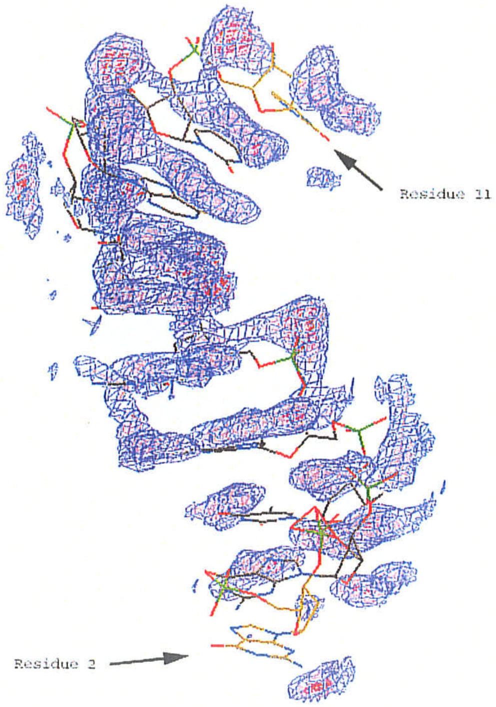

Figure 2-47 MAD electron density map and initial model placement. Crudely positioned 10-base A-DNA model (residues 2 to 11) in the experimentally phased electron density map.

With the fit of the central 8 bases improved, residues 2 and 11 were manually rebuilt in to the MAD electron density using the program xfit (McRee, 1999). The 10 base model was then subjected to further cycles of rigid body refinement resulting in an R-factor of 44.6%. At this point it was possible to build in the 3'-terminal flipped-out guanine residue. The R-factor after rigid body refinement for the 11 base model was 41.6%. With the crude model in position, atomic positional refinement of the structure (hereafter referred to as 12merSp) was initiated using only the native data.

2.5.2.1 Refinement of 12merSp

As noted in section section 1.2.8, crystallographic refinement involves the minimization of a global energy term consisting of an empirical energy term and an X-ray term. A nucleic-acid-specific parameter file was used for the empirical energy restraints which contains information about the average values for bond distances, bond angles and dihedral angles. The values resulted from a study of structures at atomic resolution which contained bases, sugars, or the phosphodiester linkage (Parkinson *et al.*, 1996). Additional restraints were added to restrain the bases to planarity. No restraints were required to maintain the Watson-Crick geometry of the central 10 base-pairs. The X-ray energy target used was based on a maximum likelihood function as implemented in the CNS program suite (Brunger *et al.,* 1998; Pannu & Read, 1996; Adams *et al.*, 1997). The maximum likelihood X-ray target requires setting aside a portion of the X-ray data for cross-validation and the calculation of R_{free} (section 1.2.8; Brunger, 1990) — 12% of the data was set aside for this purpose. The relative weight of the empirical energy term relative to the X-ray term was chosen at the various stages of refinement in order drive the refinement towards a lower R-factor without introducing unreasonable bond lengths and angles in the model.

The 11 base model was refined using the simulated annealing procedure incorporated in CNS with a torsion angle dynamics algorithm (Rice & Brunger, 1994). Torsion angle dynamics reduces the number of parameters being refined and hence reduces the degree of overfitting of the data. The simulated annealing involved heating the "system" from an initial value of 300 K to 5000K (recall this does not represent a physical temperature, but rather a parameter allowing the system to escape local minima), performing torsion angle dynamics, subsequently cooling the system in increments of 25 K according to the slow cooling annealing schedule (Brunger & Krukowski, 1990) and repeating the torsion angle dynamics at each temperature step.

The first round of simulated annealing was carried out using the 11 base model and all data greater than 5σ in the resolution range 20 - 2.5 Å. The errors in the model in the regions of manual building were corrected and the R-factor dropped to 29.7% with an R_{free} of 34.7%. Refinement continued with the inclusion of all data greater than 4σ up to 2.0 Å over several rounds of simulated annealing. This resulted in little improvement in the R-factor. Inspection of a 2Fo-Fc electron density map revealed that the sugar conformations of the terminal residues did not agree well with the observed electron density. In order to reduce the bias of the ring conformations towards the restraint values of the C3'-endo conformation, the nucleic acid parameter file was altered to remove all torsion angles involving the ring conformation. In order to maintain the geometry of the sugar pucker of residues 3 to 9, a weak dihedral energy restraint was added which allowed the ring torsion angles of these residues to deviate up to 10° from their ideal values. Several rounds of simulated annealing was carried out with the new parameter file and all data above 3σ in the resolution range 20 - 1.98 Å. The resulting structure had an R-factor of 27.8% (R_{free} = 33.1%) and the sugar moieties of the terminal residues had an improved fit with the electron density. Calculation of a Fourier difference electron density map for the structure at this point allowed the manual positioning of the missing C1 residue (Figure 2-48). Incorporation of the C1 residue into the model and subsequent refinement reduced the R-factor to 25.1%.

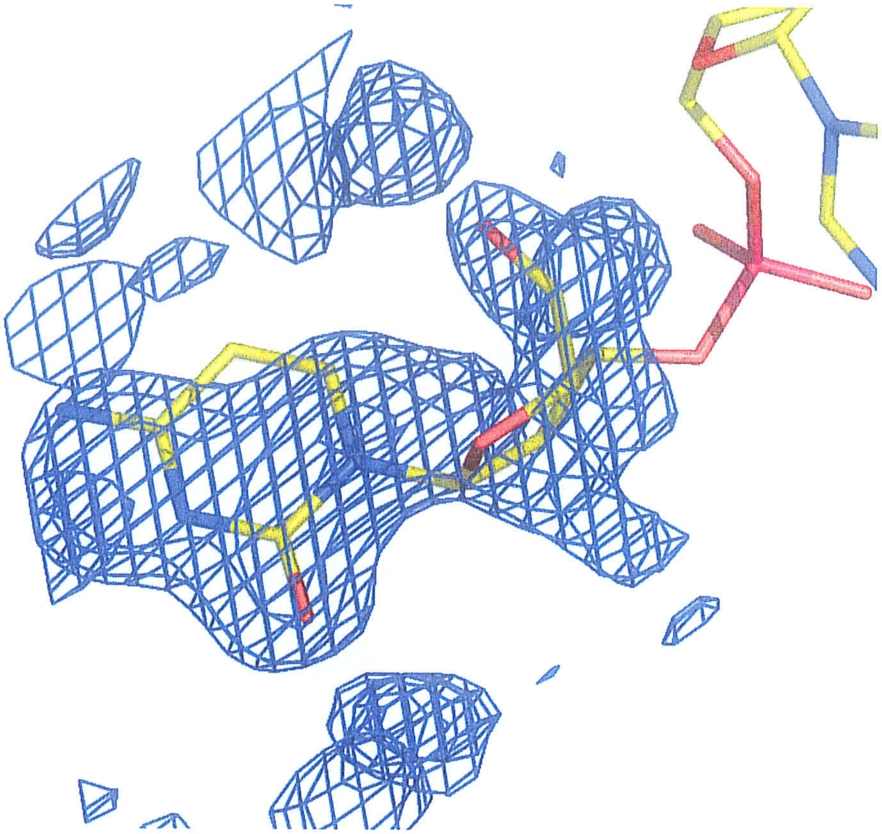

Figure 2-48 12merSp Fo - Fc electron density map at residue C1.
Contoured at 1.5 sigma

Inspection of Fourier sum (2Fo - Fc) maps revealed problems with the sugar phosphate backbone at residues A9 and C10 (Figure 2-49) which were attributed to conformational flexibility. The sugar phosphate backbone was modelled and refined in two alternate conformations each with an occupancy of 0.5 to account for the observed density (Figure 2-50). When all of the DNA atoms had been accurately positioned, atomic temperature factor refinement was carried out on all residues except for the partially disordered

C1 residue. The B-factors for residue C1 were fixed at a value approximately 25% greater than the average B-factor for the fully ordered residues to account for its apparent increased thermal motion. With a fixed residue B-factor of 45 Å^2, the occupancy of the C1 residue was refined to 0.8.

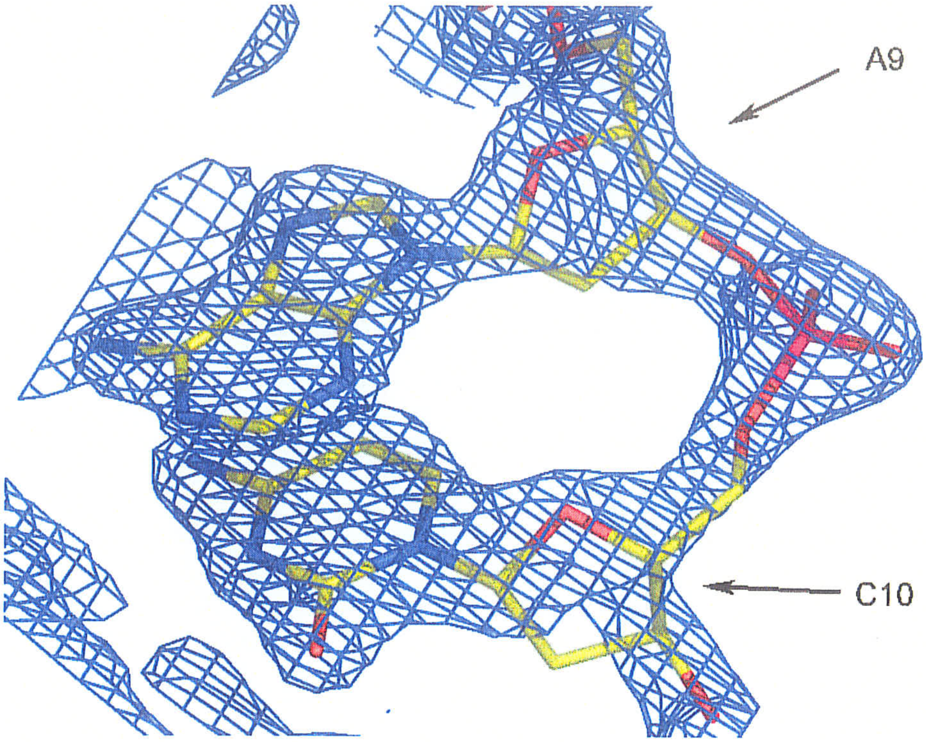

Figure 2-49 2Fo - Fc map of 12merSp at residues A9 and C10.
Contoured at 1.75 sigma

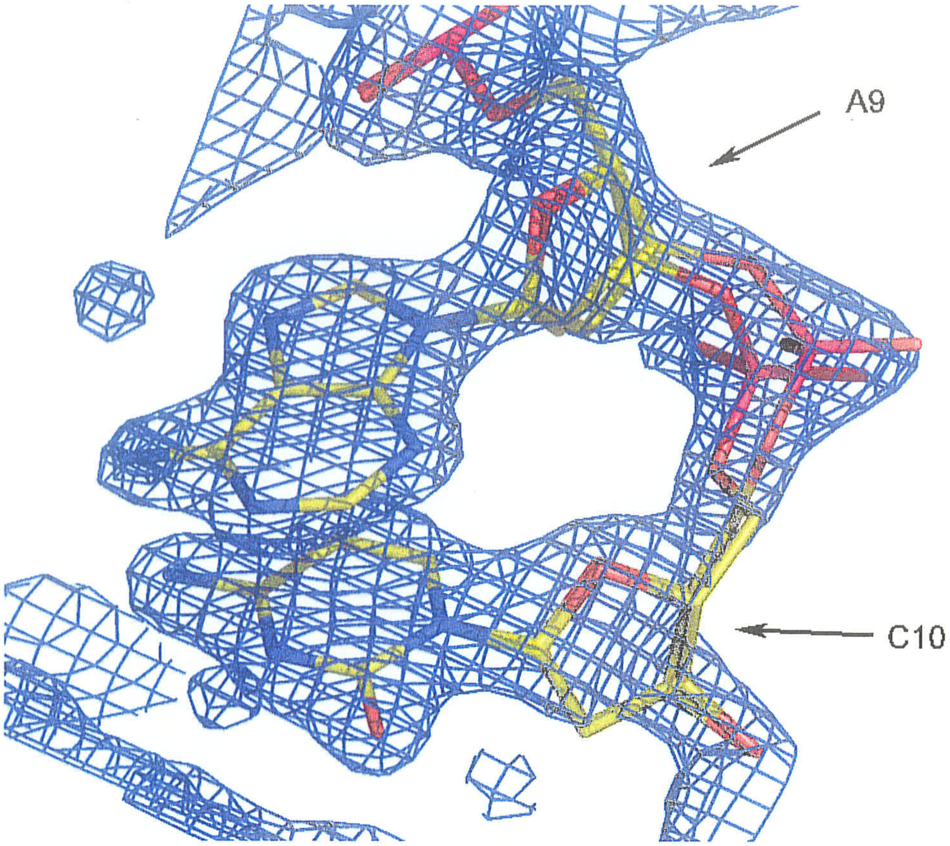

Figure 2-50 2Fo - Fc map of 12merSp with backbone conformational flexibility.
Contoured at 1.75 sigma

At this stage, well-ordered solvent water molecules located from difference Fourier maps were gradually included. After inclusion of 28 water molecules per 12mer single strand, continued refinement using 2733 reflections of greater than 2σ between 30 Å and 1.95 Å resolution gave an R-factor of 21.5% and an R_{free} of 25.2%.

2.5.3 Cobalt hexammine crystal form

The unit cell parameters of the 12mer crystals grown in the presence of cobalt hexammine are similar to those seen for the crystal structure of 12merSp (section 2.5.2). The crudely positioned 11 base model of 12merSp structure prior to atomic positional refinement was used as the starting point for the structure refinement of the 1.75 Å data set (collected at BNL) from the cobalt hexammine crystal form (hereafter referred to as the 12merCo structure). The structure refinement was carried out with the CNS program suite (Brunger *et al.,* 1998).

Rigid body refinement of the 11 base model was carried out using reflections in the 12merCo data set with I/σ greater than 5 in the resolution range 20 to 2.5 Å (10% of data set was aside for cross-validation). The R-factor after several cells of rigid body refinement was 45.6% with an R_{free} of 45.2%.

Atomic positional refinement was carried out with the same modified nucleic acid parameter file and simulated annealing schedule as the 12merSp structure. The first round of simulated annealing was carried out, using all data greater than 5σ in the resolution range 20 - 2.5 Å, on the 11 base model. The errors in the model in the regions of manual building were corrected and the R-factor dropped to 33.1% with an R_{free} of 36.1%. Difference Fourier maps revealed the location of one fully ordered and one partially ordered cobalt hexammine ion (Figure 2-51). Coordinates and empirical energy terms for the cobalt hexammine ion were obtained from HIC-Up (Kleywegt & Jones, 1998) and included in the refinement. Continued simulated annealing refinement with the inclusion of the cobalt hexammine ions and gradual inclusion of data to 1.9 Å and 3σ resulted in an R-factor of 28.5% (R_{free} = 27.3%).

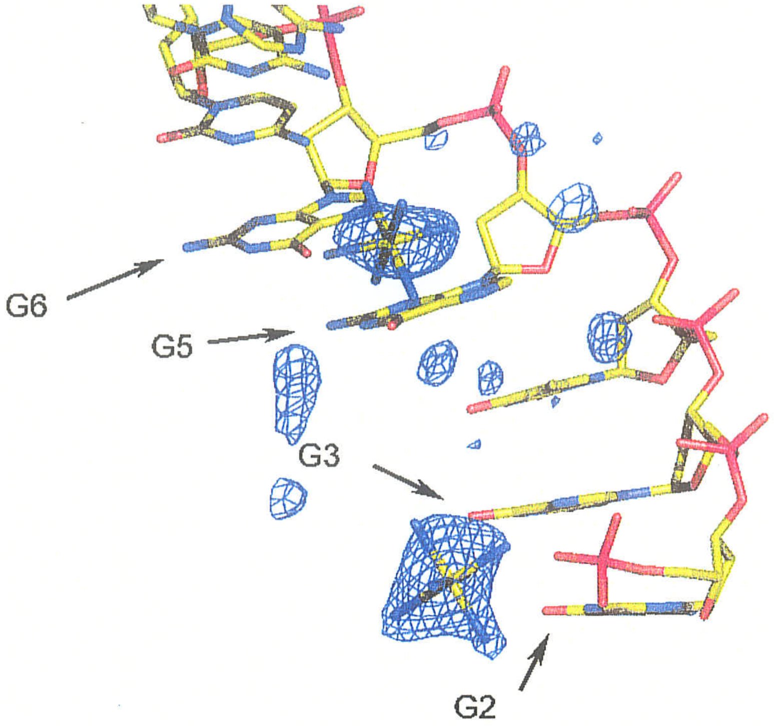

Figure 2-51 Difference Fourier map of 12merCo.
A portion of the difference (Fo - Fc) Fourier map of the 12merCo. Density near the center of the DNA suggests two partially occupied $[Co(NH_3)_6]^{3+}$ ions are located at the edges of the G5-G6 and the symmetry related G5*-G6* bases (only one of the ions is illustrated for clarity).

Calculation of a Fourier difference electron density map for the structure at this point allowed the manual positioning of the missing C1 residue (Figure 2-52). Incorporation of the C1 residue into the model and subsequent refinement reduced the R-factor to 28.5% (R_{free} = 25.8%). Refinement continued with the inclusion data to 1.8 Å over several rounds of simulated annealing.

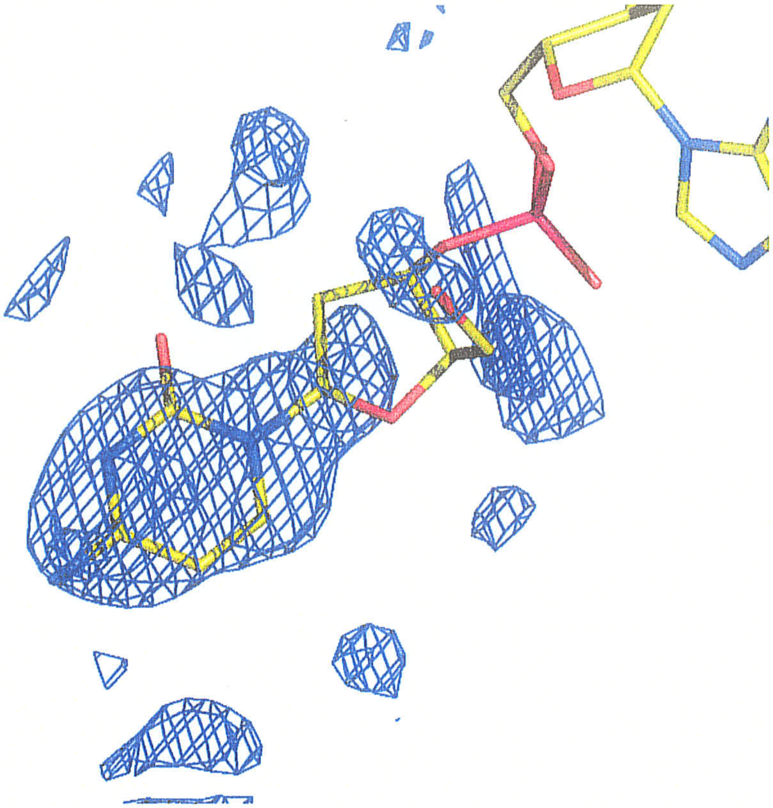

Figure 2-52 Difference Fourier map of 12merCo at residue C1.

Inspection of Fourier sum (2Fo - Fc) maps, as in the 12merSp structure, revealed problems with the sugar phosphate backbone at residues A9 and C10 (Figure 2-53) which were again attributed to conformational flexibility. The sugar phosphate backbone was modelled and refined in two alternate conformations each with an occupancy of 0.5 to account for the observed density (Figure 2-54). When all of the DNA atoms had been accurately positioned atomic temperature factor refinement was carried out on all residues except for the partially disordered C1 residue. The B-factors for residue C1 were fixed at value approximately 25% greater than the average B-factor for the fully ordered residues to account for its apparent increased thermal motion. With a fixed residue B-factor of 40 Å^2, the occupancy of the C1 residue was refined to 0.7.

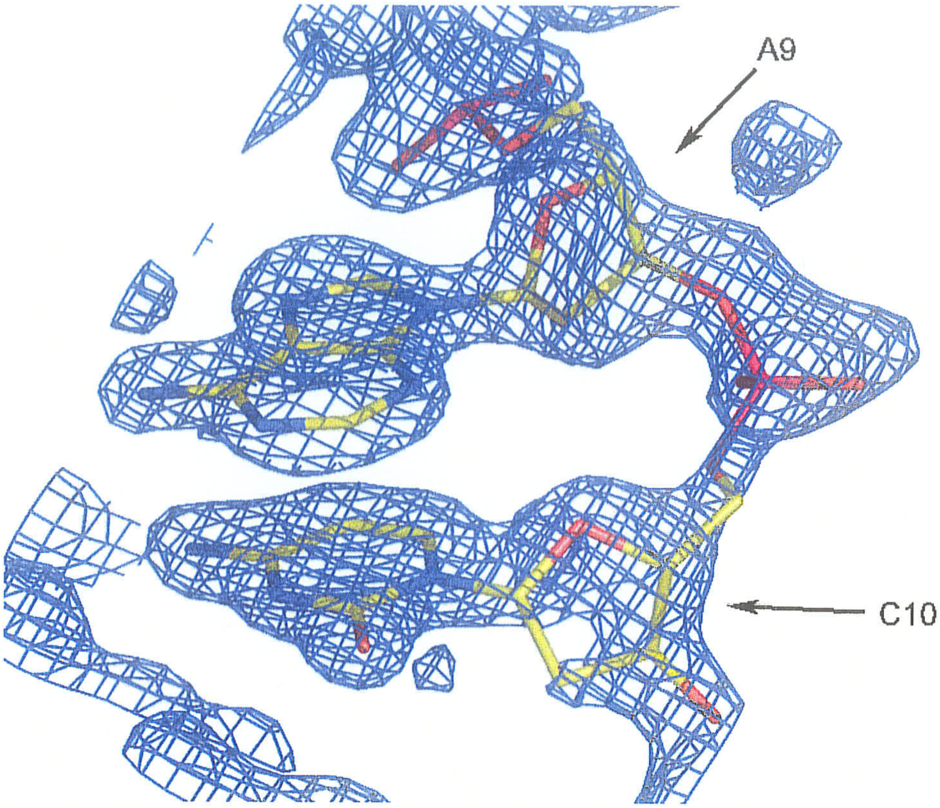

Figure 2-53 Fourier sum (2Fo - Fc) map of 12merCo at residues A9 and C10.

At this stage, well-ordered solvent water molecules located from difference Fourier maps were gradually included. After inclusion of 37 water molecules per 12mer single strand and 1.5 Cobalt hexammine ions, continued refinement using 3726 reflections of greater than 2σ between 30Å and 1.75 Å resolution gave an R-factor of 22.4% and an R_{free} of 25.2%

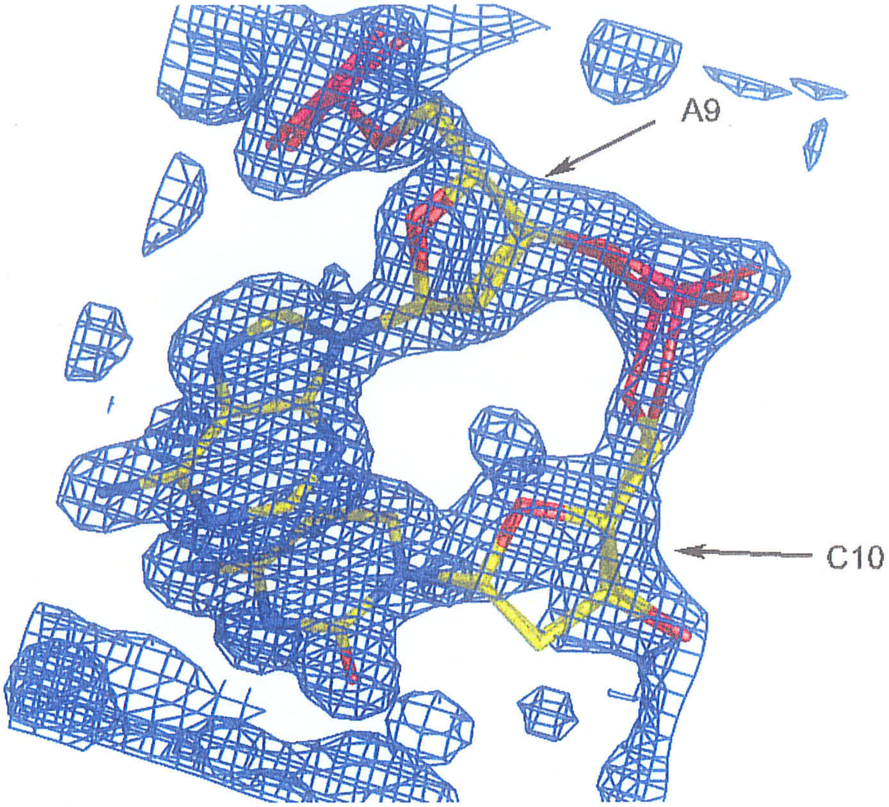

Figure 2-54 2Fo - Fc map of 12merCo with backbone conformational flexibility.

CHAPTER 3

Results and Discussion

3.1 C-C Mismatch Results

From the UV and CD thermal denaturation studies it was determined that the C-C mismatch-containing nonamer duplex was stable below 15°C and at low pH. This information was incorporated into the crystallization trials in order to decrease the number of crystallization parameters and to maximize the likelihood of obtaining crystals. The experiments indicated that the thermal stability of the C-C mismatch duplex was increased by the presence of a divalent cation and low concentrations of spermine. While no diffraction quality crystals were obtained, the solution studies provided valuable information regarding the global conformational and stability of the nonamer sequence. The solution studies indicated the conformational flexibility associated with the d(CGCTGGCCACCG) sequence (Chen, 1997), which exhibited biphasic melting, was not present in the nonamer d(CGGCCGCCG), thus making it more likely to crystallize. The initial crystallization screen provided promising results; further optimization of the conditions shown in Figure 2-24 might yield X-ray diffraction quality crystals. No one has yet published a crystal structure of an anti-parallel C-C mismatch, but our initial results are encouraging.

3.2 Crystallographic Results

A summary of the structural parameters and atomic numbering schemes used in describing the DNA structures is given in Appendix A.

3.2.1 The Structure of d(CGGTGGCCACCG)$_2$ - spermine form

Using the techniques described in the experimental section, the structure of d(CGGTGGCCACCG)$_2$ crystallized in the presence of spermine (hereafter referred to as 12merSp) was determined to 1.95 Å resolution. The central ten base pairs were found to adopt a primarily A-type conformation while the terminal 5'-C and G-3' residues are flipped out away from the helical axis. Within the crystal structure each dodecamer single strand forms base-pairing/stacking interactions with three other strands. Firstly, a major duplex interaction is formed by the central ten bases 5'-GGTGGCCACC-3' of the dodecamer (Figure 3-1a). These bases are involved in Watson-Crick base-pairing interactions with a two-fold related strand — the chemical dyad axis of the 12mer coincides with a crystallographic two-fold axis. The flipped-out terminal bases are stabilized by two types of interactions with other symmetry related strands. The G-3' residue is primarily stabilized by a base stacking interaction with a symmetry-related G-3' residue (Figure 3-1b). The two six-membered rings of the G-3' residues are stacked with a separation distance of approximately 3.4Å. The second interaction is formed through a water-mediated hydrogen-bond between the G-3' residue and a 5'-C1 base from yet another strand (Figure 3-1c). This second interaction likely plays only a minor role in the stabilization of this conformation as the 5'-C1 residue is partially disordered.

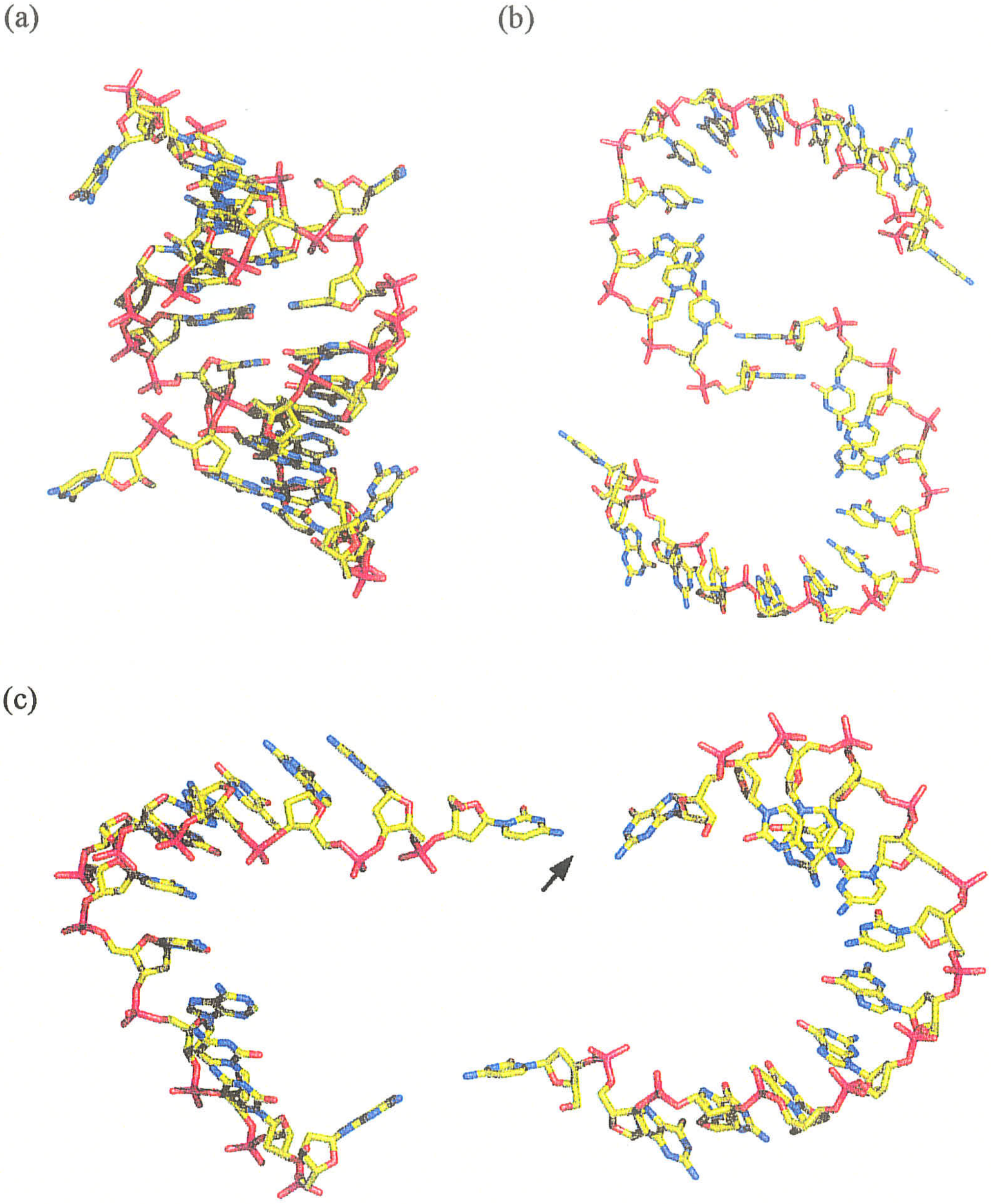

Figure 3-1 The base pairing/stacking interactions.
(a) Primary duplex interaction, (b) G-G base-stacking and (c) water-mediated H-bonding between C1 and G12 of symmetry related duplexes.

3.2.1.1 Sugar-Phosphate Backbone

The backbone torsion angles, as well as pseudorotation angles and sugar pucker types for the12merSp structure are given in Table 3-1 along with values from fiber diffraction studies for canonical A-DNA and B-DNA (Chandrasekaran & Arnott, 1989). The torsion angle wheel in Figure 3-2 graphically illustrates the backbone torsion angles for the 12merSp structure as well as the observed ranges for other A-DNA and B-DNA crystal structures (Schneider *et al.,* 1997).

TABLE 3-1 Backbone torsion angles and furanose ring conformations for 12merSp.

	α	β	γ	δ	ε	ζ	χ	Phase (degrees)	Pucker	P_i--P_{i+1} (Å)
C1	-	-	46.4	95.2	236.0	72.8	259.2	15.5	C3'-*endo*	-
G2	253.8	192.3	50.3	85.6	206.1	302.0	194.3	12.8	C3'-*endo*	5.5
G3	146.0	183.3	186.9	84.3	214.5	281.7	181.4	10.9	C3'-*endo*	6.7
T4	297.6	173.4	47.8	80.8	209.4	292.3	201.8	11.8	C3'-*endo*	5.5
G5	293.7	174.1	51.8	82.2	195.4	283.6	199.8	19.5	C3'-*endo*	6.0
G6	297.5	179.7	51.1	83.4	207.1	288.4	198.7	18.6	C3'-*endo*	5.8
C7	300.3	173.1	46.6	80.5	202.9	295.9	206.2	16.0	C3'-*endo*	5.7
C8	140.4	185.0	187.8	88.5	211.3	289.2	195.3	8.8	C3'-*endo*	6.7
A9	298.8	190.8	46.6	131.4	184.8	275.5	233.7	139.5	C1'-*exo*	6.3
C10	290.0	171.3	58.7	90.5	181.5	283.2	214.8	60.6	C4'-*exo*	6.5
C11	296.0	200.8	46.8	136.4	185.4	255.4	242.1	146.8	C2'-*endo*	6.4
G12	214.0	143.2	45.4	148.5	-	-	251.9	162.4	C2'-*endo*	-
A-DNA[a]	308	175	42	79	212	285	203	-10-40°	C3'-*endo*	5.5
B-DNA[a]	330	136	31	143	219	199	262	140-185°	C2'-*endo*	6.6

[a] Fiber diffraction data values (Chandrasekran & Arnott, 1989).

Focusing first on the central octamer structure, the backbone torsion angles for the most part fall within the observed range of other A-DNA structures. The α and γ torsion angles in nucleotides G3 and C8 adopt an extended *trans,trans* conformation instead of the more common *gauche*$^-$,*gauche*$^+$ conformation. This extended backbone conformation has been observed in other A-DNA structures (Haran *et al.*, 1987; Gao *et al.*, 1999) and is associated with a general opening of the helix and by an increase in the base separation distance (Olson, 1982). The glycosidic torsion angles, χ, are all in the usual anti-conformation.

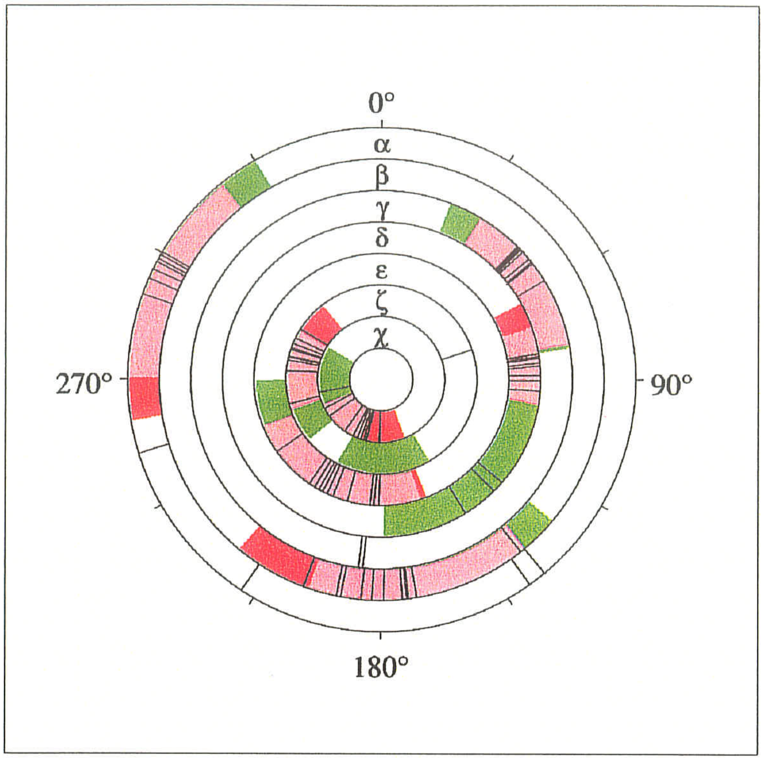

Figure 3-2 Backbone torsion angle wheel for 12merSp.
The black lines represent the values of the 12merSp structure, shown in red are the observed ranges for A-DNA crystal structures and shown in green are the observed ranges for B-DNA crystal structures.

The sugar-phosphate backbone exhibits local conformational flexibility around nucleotides A9 and C10 (Figure 3-3 and Table 3-2). The backbone disorder is associated with a concomitant change of the α and γ torsion angles of C10 from a *sc* to an *ac/ap* conformation, resulting in a 1.3 Å shift of the phosphate group. This flipping of the α, γ pair is the product of a crankshaft motion around the O5'-C5' bond (torsion angle β), creating the extended backbone variant of the A-DNA backbone. The local flexibility is not isolated to just these atoms and we see movement of greater than 0.4Å in atoms C2', C3' and O3' of A9 and atom C4' of C10. As a result we see changes in the A9 and C10 sugar puckers, with the pseudorotation phase angles deviating by 10 and 17° respectively between the alternate backbone conformations. Despite the conformational flexibility in the sugar-phosphate backbone, the orientation of the bases is invariant with minor changes in the glycosidic torsion angle of C10 (on the order of 10°).

TABLE 3-2 Local conformational flexibility for 12merSp.

	α	β	γ	δ	ε	ζ	χ	Phase (degrees)	Pucker	P_i--P_{i+1}[a] (Å)
A9	298.8	190.8	46.6	131.4	184.8	275.5	233.7	139.5	C1'-*exo*	6.3
A9	302.5	183.8	56.9	124.1	194.7	285.7	234.3	149.2	C2'-*endo*	6.7
C10	290.0	171.3	58.7	90.5	181.5	283.2	214.8	60.6	C4'-*exo*	6.5
C10	130.2	183.9	190.8	92.4	185.4	289.2	205.6	43.3	C4'-*exo*	6.9

[a] P-P intrastrand separation distance measured in Angstroms.

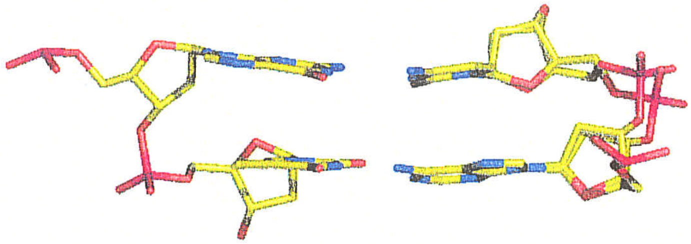

Figure 3-3 Local backbone conformational flexibility in 12merSp.
A view into the major groove of 12merSp at base-pairs G3-C10* and T4-A9* (the top base-pair is G3-C10*).

At both the 5' and 3' ends of the structure the backbone undergoes severe distortion. This distortion extends from the 5' end of the molecule to the α angle of base G2, and from the 3' end of the molecule (through the region of local conformational flexibility) to the δ of A9. The distortion occurs as the conformation of the DNA changes with movement of the 5'-C and 3'-G residues away from the major duplex interaction. The greater distortion in backbone torsion angles from canonical A-DNA values are observed towards the 3' end of the structure.

The sugar puckers for the nucleotides in the central region of the dodecamer are *C3'-endo*. As the backbone torsion angles undergo distortion there is a corresponding change in sugar ring conformation from that of pure *C3'-endo* type. For the sugar groups attached to bases A9, C11 and G12 we observe a *C2'-endo* conformation and for the sugar group attached to base C10 we observe a *C4'-exo* conformation. Sugar puckers of the *C2'-endo* type are uncommon in A-DNA structure although they have been observed (Thota *et al.,*1993). This change in sugar conformation to *C2'-endo* results in a change in the P-P intrastrand separation distances starting at the C8-A9 base step (Table 3-1). The P-P separation distances observed are longer than the 6Å distances normally observed in A-DNA structures (Lu *et al.,* 2000). As a result of the change in sugar pucker and concomitant increase in P-P separation distances, the DNA backbone is extended in this region which allows for the terminal G-3' residues to form a stable base-stacking interaction with a symmetry related strand. The precise determination of the sugar puckering mode of residue C1 was difficult due to its poor electron density. The poor electron density is likely due to the high thermal motion resulting from the poorly stabilized flipped-out conformation. Thus its puckering mode was restrained to that of a *C3'-endo* type during crystallographic refinement.

3.2.1.2 Helix Morphology and Base Stacking in the Central Decamer

The helix axis for the central decamer duplex is non-linear, with bending of the helix axis occurring near the ends in the direction of the DNA major groove. Figure 3-4 illustrates the origin of the bent helical axis. Bending of the helix axis towards the major groove is the result of an asymmetrical conformation of the sugar-phosphate backbone at the G2•C11*/G3•C10* and G5•C8*/G6•C7* base-steps where only one strand of the duplex exists in the extended α, γ conformation. Further asymmetrical distortions leading to bending of the helix axis result from the change in sugar pucker at residue A9 to the *C2'-endo* conformation. The total angle of curvature (CURVES, Lavery & Skelnar, 1989), is 32.8°. The curvature of the helix axis results in an overall shortening of the helix path-length by 6.43%. The helical axis for the central decamer and the pattern of axis curvature are shown in Figure 3-5.

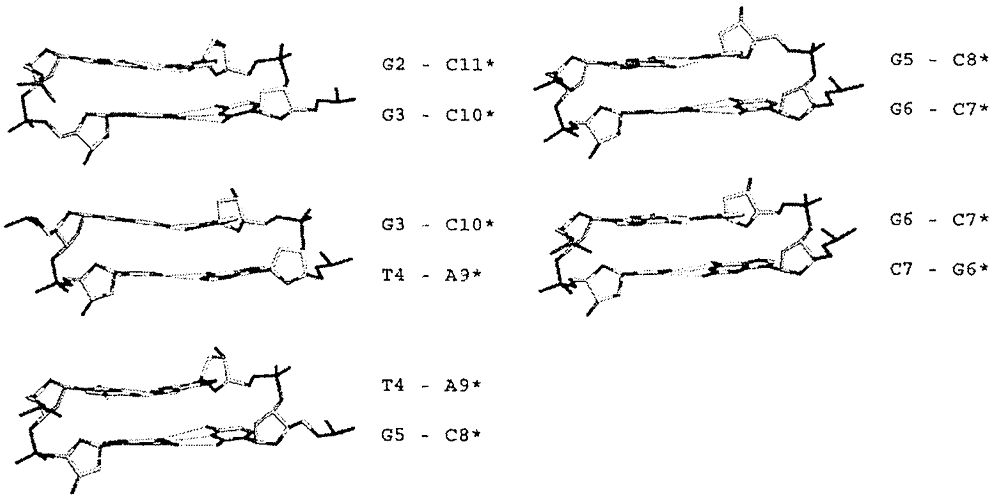

Figure 3-4 Base-pair steps of 12merSp viewed into the major groove.

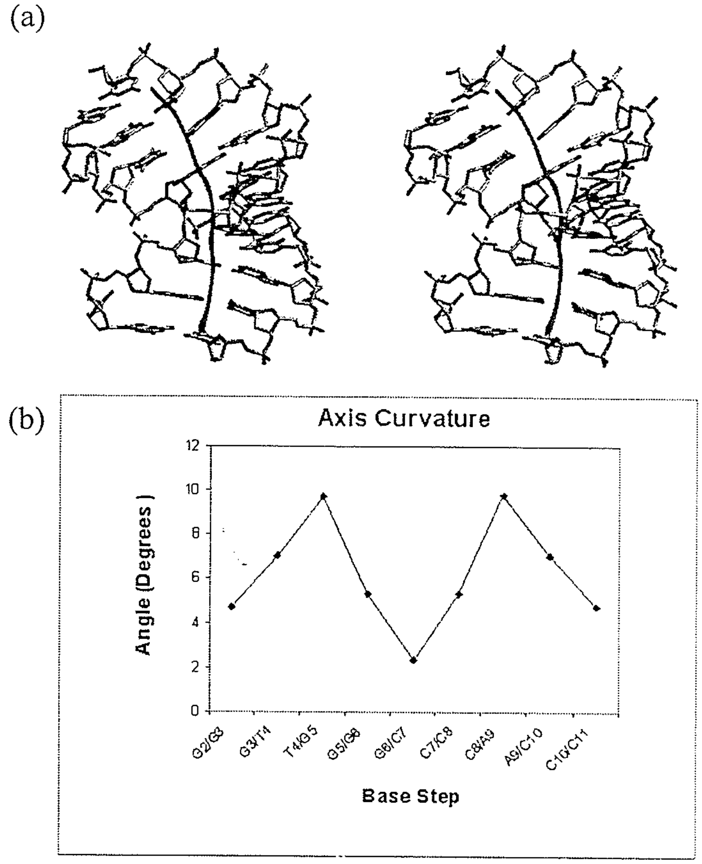

Figure 3-5 12merSp helical axis curvature.
(a) Stereo pairs of the central decamer demonstrating bending with the helix axis calculated by CURVES, (b) a graph of the angle of curvature per base step as calculated by CURVES.

Graphs of the helical parameters calculated for the central decamer duplex using a curved helical axis are shown in Figure 3-6. Based on the curved helix, the average rise/base pair is 2.68 Å, and the average helical twist/base pair is 30.5° (corresponding to 11.8 residues per turn). A noticeable feature in Figure 3-6a is the large base-pair rise for the G3•C10/T4•A9 step of 3.34 Å which is considerably larger than the expected value of 2.9 Å for A-DNA. One would also expect to find a higher twist at this step, however this is not the case here. The twist at this step is 31.8° which is close to the expected value for A-DNA of 32.7°. The increased axial rise here is accommodated by the bending of the helix in this region. A large negative roll is also associated with the G3•C10/T4•A9 base step (see Figure 3-6d) as well as an opening of the base pairs towards the major groove (see Figure 3-7f). This is a common mode of DNA bending and has been observed in other A-DNA structures such as d(GTGTACAC) (Thota *et al.*, 1993).

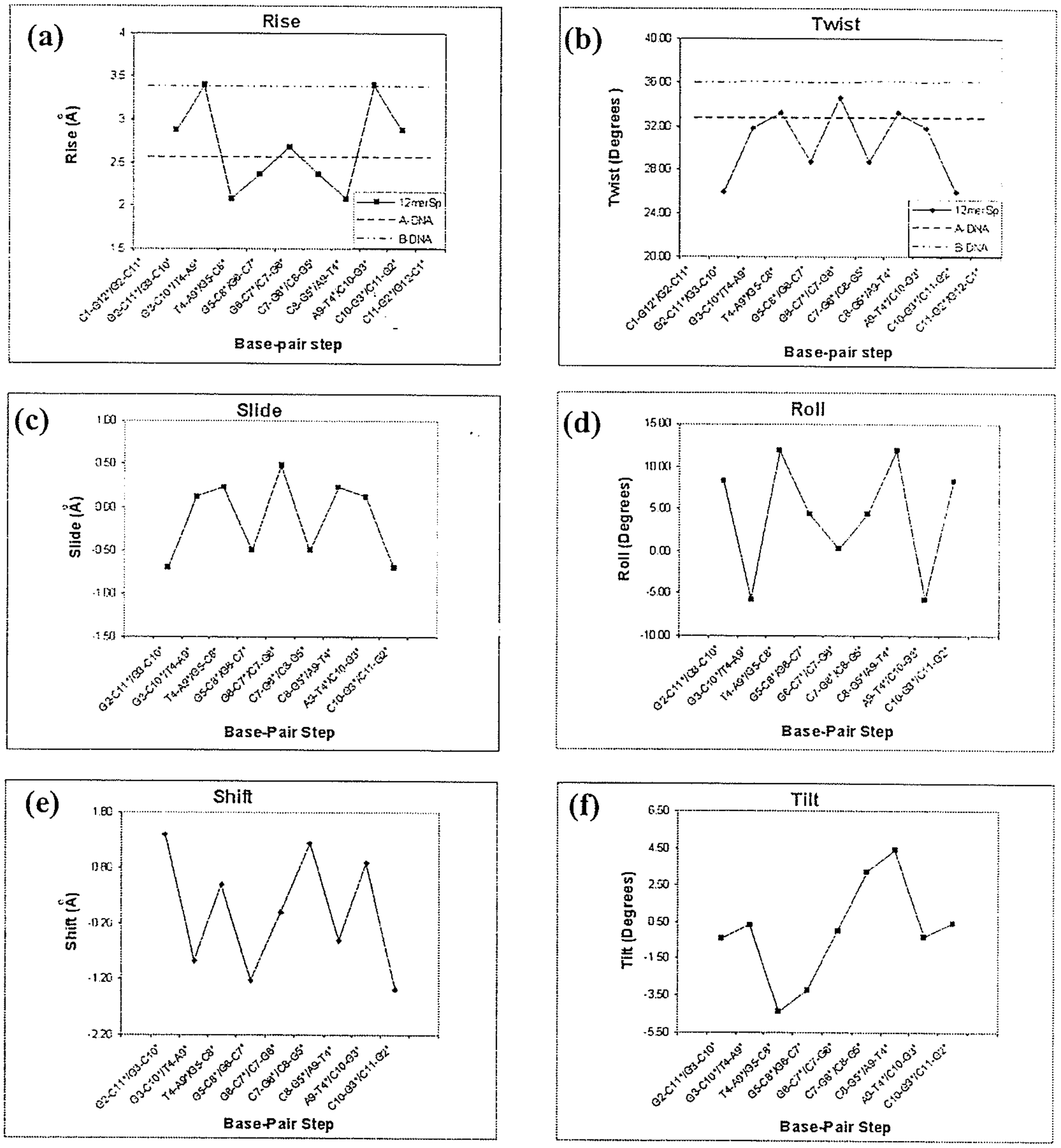

Figure 3-6 Plots of the global helical parameters for 12merSp.
Plots showing the variations in (a) rise, (b) twist, (c) slide, (d) roll, (e) shift, and (f) tilt across the helix.

Graphs of the base-pair parameters are shown in Figure 3-7 and illustrate the pattern of distortion of the helical structure. The central four base pairs are relatively undistorted. The structure deviates from an ideal A-type conformation as one approaches the flipped out terminal residues the base pairs become considerably more distorted. One notable feature in Figure 3-7b is the dramatic increase in base-pair buckle starting at base pair T4•A9*. This distortion may result from the asymmetrical conformation of the duplex backbone in this region. The α,γ extended backbone conformation may also be responsible for the opening of base pairs G3•C10* and G5•C8* towards the major groove. The opening of these base pairs results in a slight distortion of the inter-base hydrogen bonds, with the minor groove H-bonds being shortened and the major groove H-bonds being lengthened.

In A-DNA structures the purine pyrimidine base steps (RpY) have been found to display predominantly intrastrand stacking, while the YpR steps prefer interstrand stacking between the purine ring (Wahl & Sundaralingam, 1999 and references therein). The preferred stacking mode for GpG base-pair steps is for the five-membered ring of one guanine to lie directly over the six-membered ring of the other. The stacking of the dinucleotide steps for the 12merSp central decamer duplex are shown in Figure 3-4 and Figure 3-8. The effect of the bending of the helix axis on the base-pair overlaps is evident by comparing the two GpG base steps. In neither case are the guanine rings ideally stacked due to extended backbone conformations (Figure 3-4). In the case of the G5pG6 base-step, where the helix is less bent, the intrastrand base overlap is considerably higher. The only YpR base-pair step in the structure, T4pG5, exhibits a base-pair overlap typical of A-DNA despite the altered sugar pucker of residue A9.

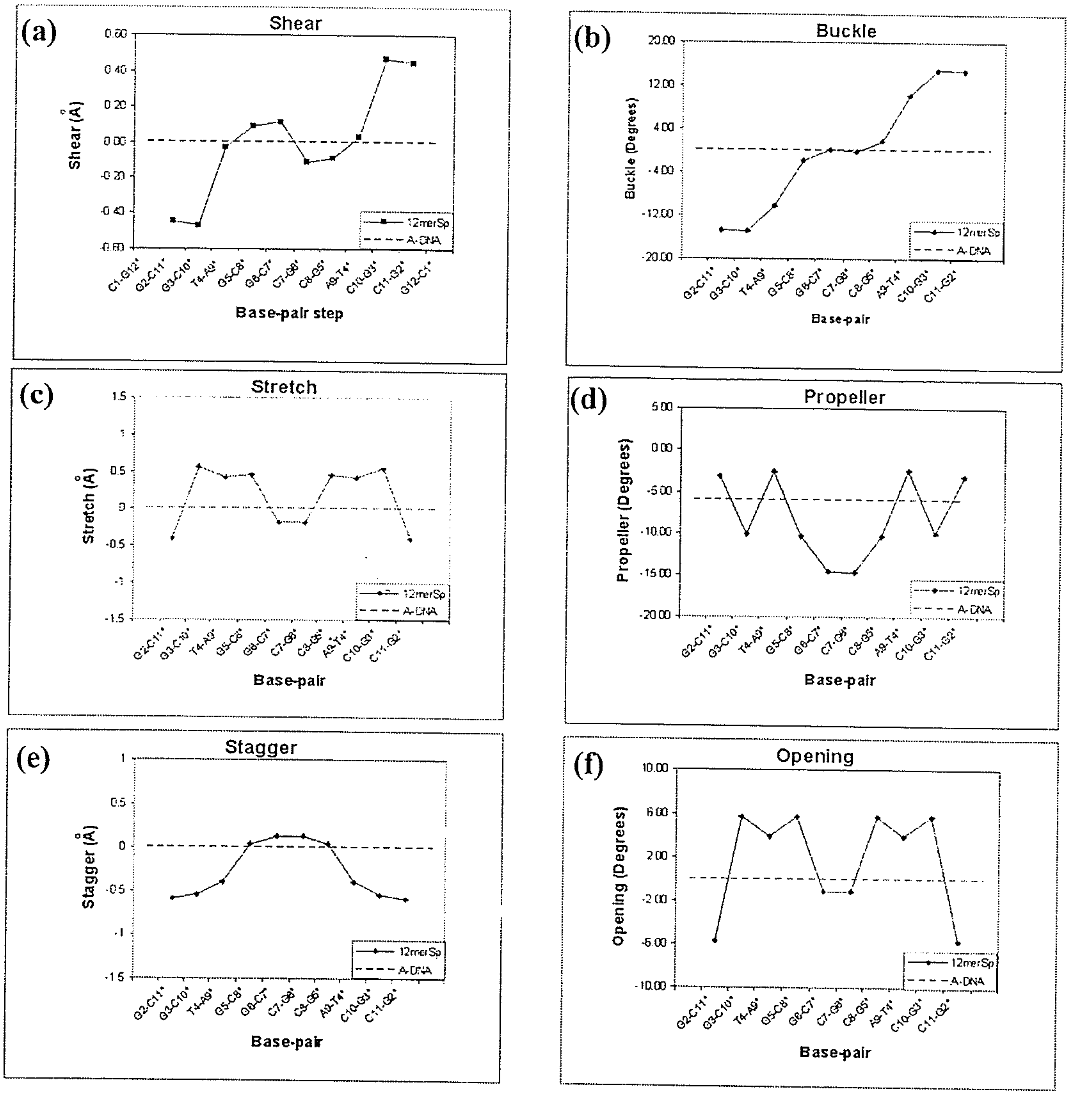

Figure 3-7 Plots of the global base-pair parameters for 12merSp.
Plots showing the variation in (a) shear, (b) buckle, (c) stretch, (d) propeller, (e) stagger, and (f) opening along the helix.

The minor groove geometry for the central decamer duplex is shown in Figure 3-9. The width and depth values observed for the central region of the structure show small variations and are close to canonical values of 11.0 and -0.5 Å. Moving away from the centre of the duplex, a reduction of the width of the minor groove and corresponding increase in the minor groove depth is observed. These changes in geometry accompany the distortions that occur in the sugar-phosphate backbone and resultant bending of the helix in this region.

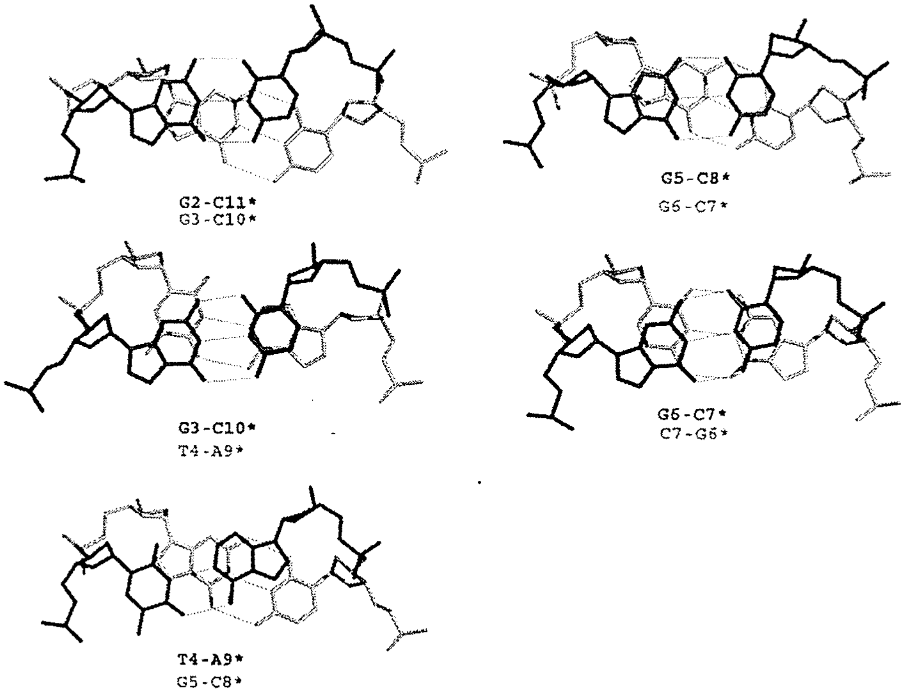

Figure 3-8 12merSp dinucleotide base-pair steps as viewed down helix axis.

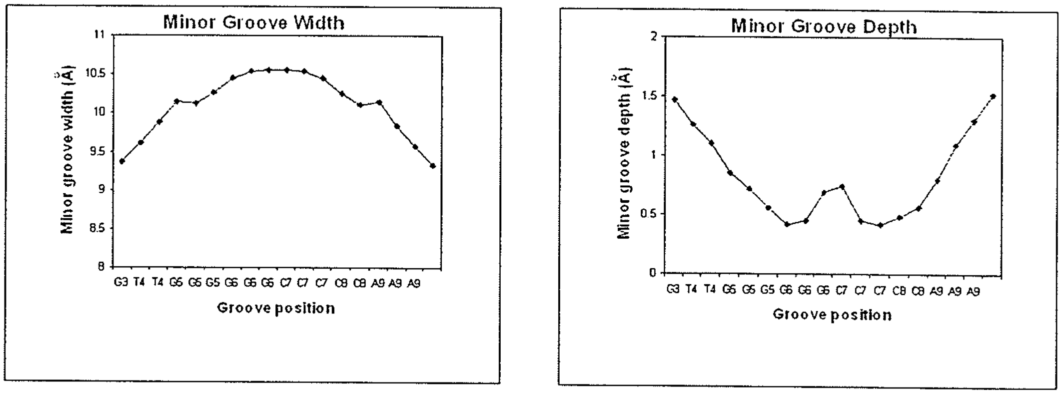

Figure 3-9 12merSp minor groove structure.

3.2.1.3 Crystal Packing

The unit cell packing is illustrated in Figure 3-10 and shows the DNA duplexes packed closely around the 3_1 axis in the unit cell. In this mode of packing each duplex interacts with 6 other duplexes and creates a solvent channel of approximately 26 Å in diameter centred around the *c* axis. This general mode of A-DNA packing is characteristic of other A-DNA structures crystallized in the $P6_122$ space group (Frederick *et al.*, 1989; Bingman *et al.*, 1992). The details of how the DNA molecules pack are unique to this structure. The first set of interactions involve the flipped out cytosine residues and penultimate base-pairs packing into the shallow minor groove of two neighbouring duplexes. The next set of interactions results from the flipped out guanine bases that are tucked into the minor groove of two other neighbouring duplexes. The final set of interactions involve two more neighbouring duplexes (interlocked through their flipped out guanine bases) which insert their flipped out cytosines and penultimate base-pairs into the minor groove of the central DNA duplex. The overall result of this packing scheme is that the majority of the apolar surface of the minor groove is buried by close interactions with symmetry-related molecules (Figure 3-11).

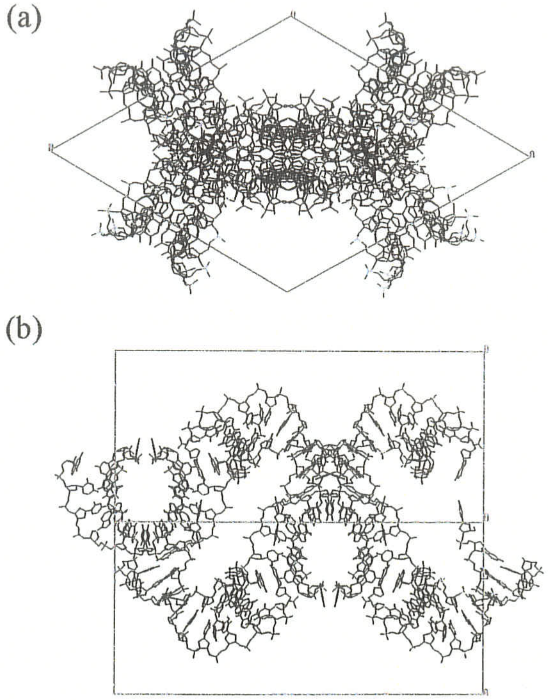

Figure 3-10 Unit cell packing in 12merSp.
Two views of the unit cell packing for 12merSp: (a) as viewed down the *c* axis and (b) parallel to the *c* axis.

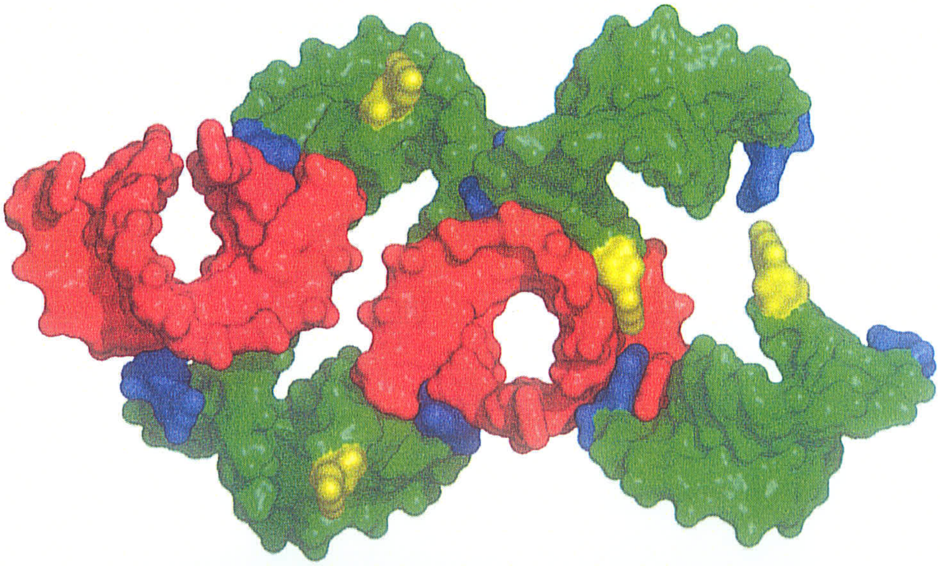

Figure 3-11 Surface packing diagram for the 12merSp crystal structure.
The occupation of the minor groove of the central DNA molecule (red) by the C1 (yellow) and G12 (blue) residues is illustrated.

3.2.1.4 Structure of the Terminal and Penultimate Base Pairs

The terminal guanine residue, G12, is oriented away from the helix axis such that it folds back towards the minor groove of the central decamer duplex. As a result of this orientation, the C8 atom of residue G12 forms a 3.4 Å contact with the phosphate backbone O3' of the preceding residue. A further consequence of this flipped out conformation is the creation of a binding pocket large enough to accommodate a symmetry-related terminal guanine residue. The symmetry-related guanine packs tightly into this pocket and forms base-stacking interaction with the flipped-out guanine residue. The two guanine bases are stacked such that their six-membered rings overlap with a separation distance of 3.4 Å (Figure 3-12). The interactions of the flipped-out guanine with the minor groove of the central duplex are primarily hydrophobic in nature, with the surface of the base packing against the sugar-phosphate backbone (Figure 3-13). There is a hydrogen bonding interaction of 3.01Å from the N2 amino group on the flipped out guanine to the O2 carbonyl of residue C11 in the duplex. Other potential hydrogen bonding interactions involve the 3'-terminal hydroxyl group of G12 with the amino group of G2 (Figure 3-13).

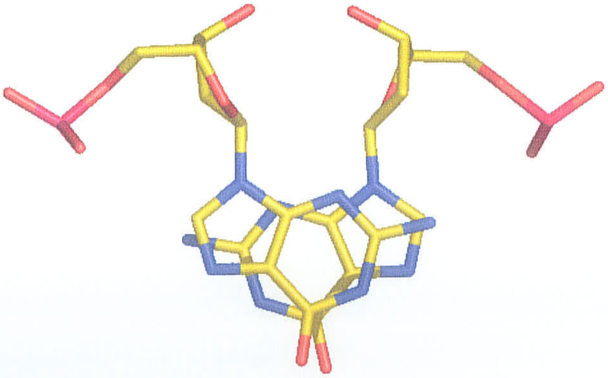

Figure 3-12 Stacking of the G-3' bases in the 12merSp crystal structure.

The 5'- terminal cytosine is flipped-out away from the helix axis such that the surface of the base forms contacts of a hydrophobic nature with the sugar-phosphate backbone of a neighbouring duplex. Within the minor groove the terminal cytosine makes a water-mediated hydrogen bonding interaction with a symmetry-related terminal guanine (Figure 3-14).

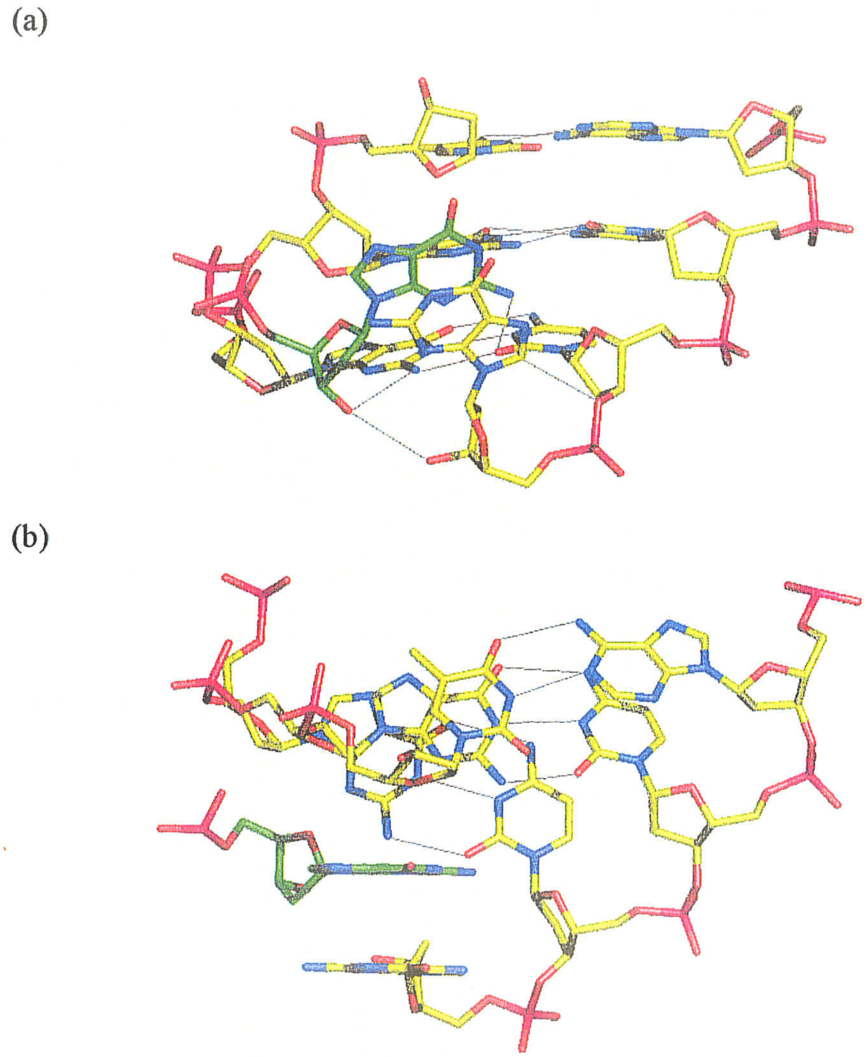

Figure 3-13 12merSp terminal guanine minor groove interactions.
Two views of the flipped-out guanine interactions with the minor groove: (a) as viewed into the minor groove and (b) perpendicular to the guanine bases. Base pairs T4-A9*, G3-C10*, G2-C11* and base G12 of the central DNA duplex are shown in yellow and the symmetry related terminal guanine in green.

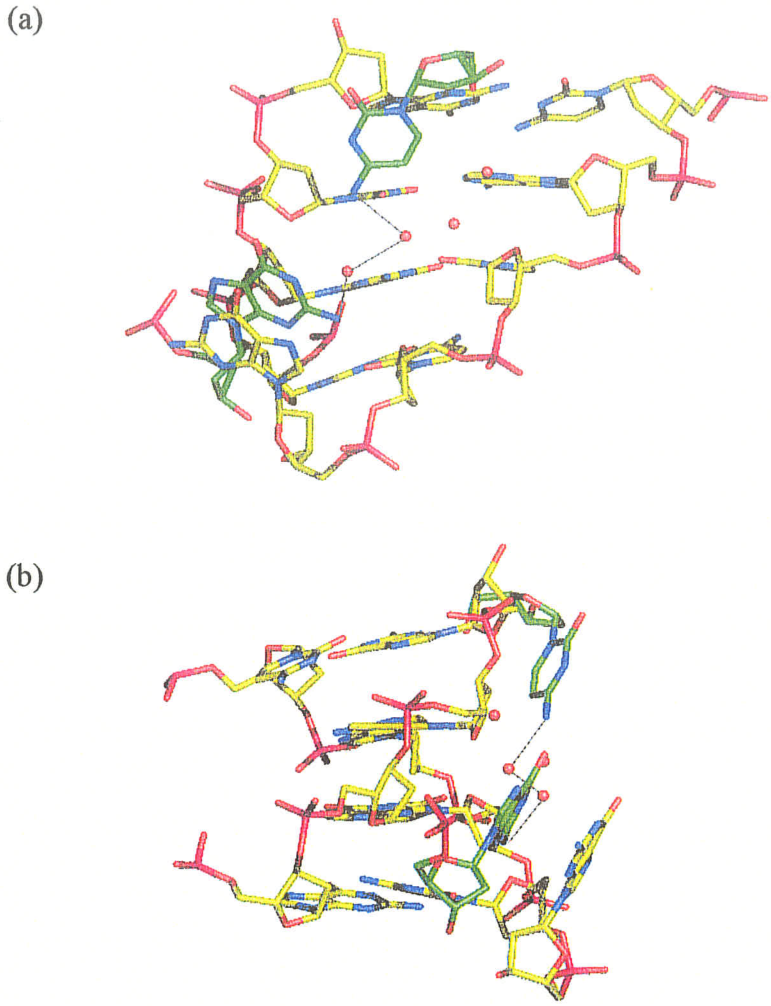

Figure 3-14 Water-mediated H-bonding in the minor groove of 12merSp. Two views of the water-mediated hydrogen bonding between the flipped-out terminal residues in the minor groove of a central duplex: (a) as viewed into the minor groove and (b) rotated 90° clockwise. The central duplex is shown in yellow, the symmetry related terminal bases in green and the water molecules as red spheres.

The penultimate base-pair, G2-C11*, is involved in hydrophobic interactions with the sugar-phosphate backbone of a neighbouring duplex (Figure 3-15). This type of interaction is generally observed for the terminal base pairs in A-DNA crystal packing schemes.

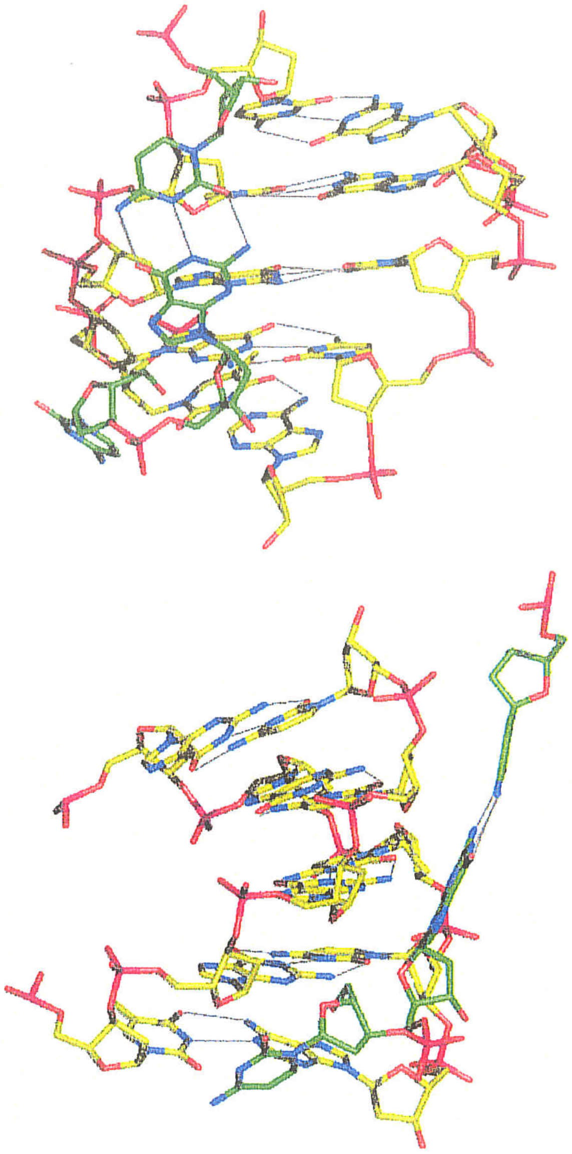

Figure 3-15 Interactions of the penultimate base-pair of 12merSp.
Two views of the minor groove interaction of the penultimate base-pair with a neighbouring duplex: (a) as viewed into the minor groove and (b) rotated 90° clockwise. The central duplex is shown in yellow and the symmetry related terminal bases in green.

3.2.1.5 Hydration

The asymmetric unit of the 12merSp crystal contains 28 water molecules. The major groove of the 12merSp structure is extensively hydrated (Figure 3-16a,b). Twelve water molecules (per DNA single strand) line the major groove and make contacts with polar DNA base atoms and/or other water molecules. The minor groove hydration shell is almost completely displaced by the hydrophobic surfaces of the penultimate and terminal base pairs of neighbouring duplexes. There are only three water molecules in the shallow minor groove (Figure 3-16c). Two of the water molecules make direct contact with polar base atoms of residues A9 and C10. The third water molecule is involved in water-water hydrogen-bonding linking the other two water molecules. Figure 3-17 illustrates the water contacts made with the polar base atoms in the major and minor grooves. The sugar-phosphate backbone is economically hydrated, with on average only a single water molecule per phosphate group.

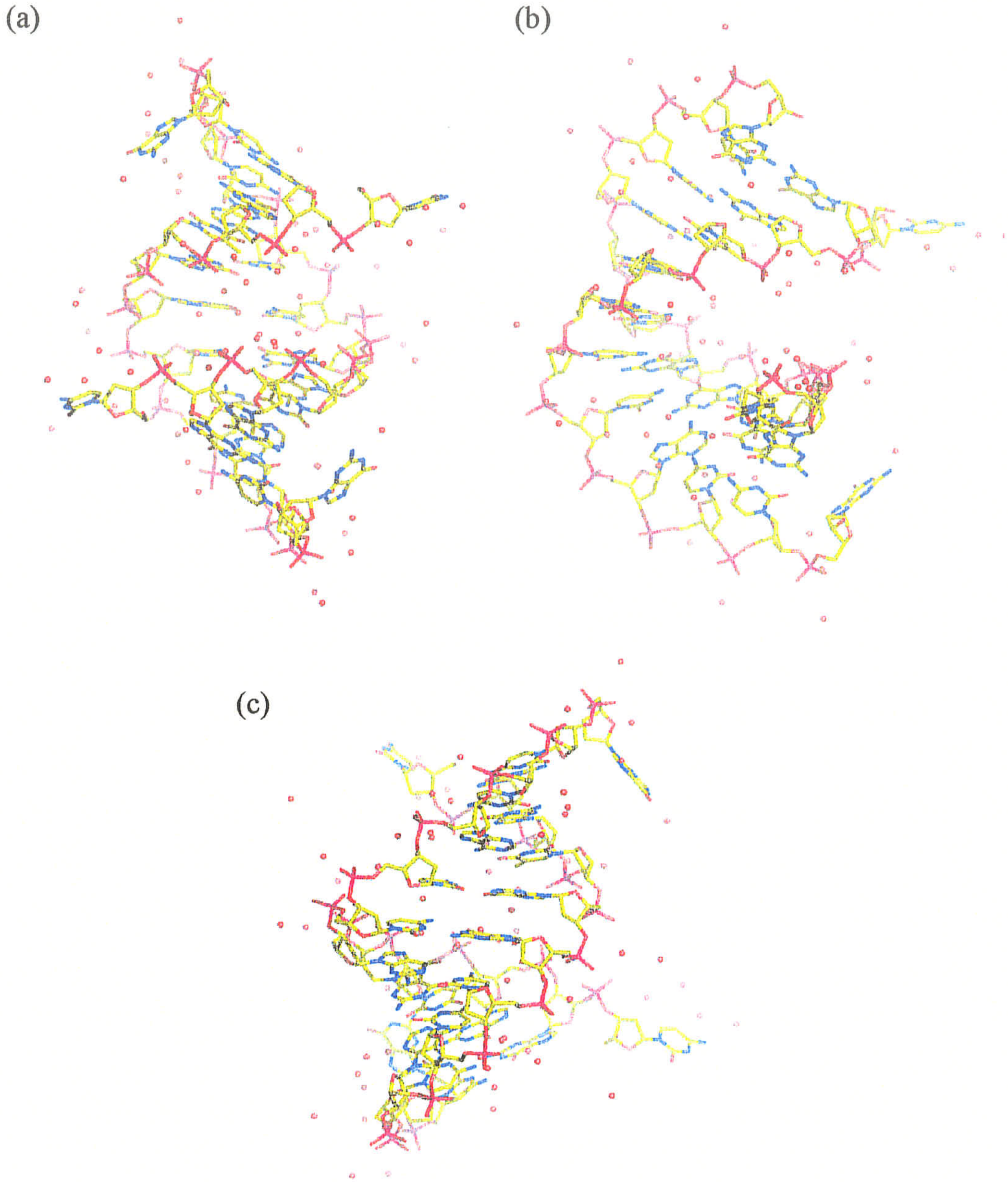

Figure 3-16 12merSp hydration.

Hydration of the 12merSp duplex structure as viewed (a) into the major groove, (b) rotated 90° and (c) into the minor groove.

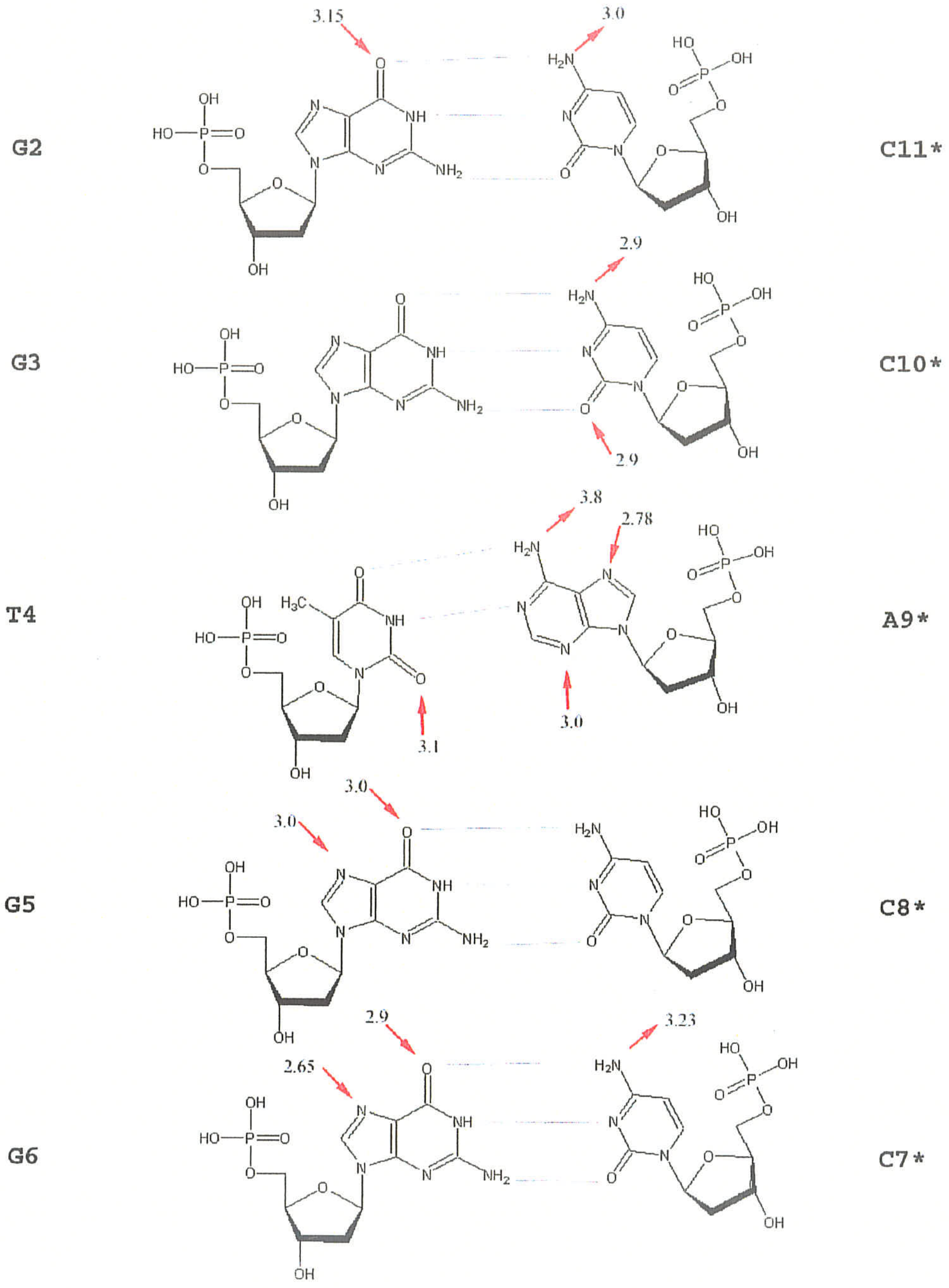

Figure 3-17 Schematic of 12merSp base hydration.
The H-bonds between the DNA and water molecules are shown in red and the distances are given in Å.

3.2.1.6 Thermal Parameters

In 12merSp average isotropic temperature factors range from 26 to 42 Å^2 for phosphate groups, from 25 to 45 Å^2 for deoxyribose units, and from 24 to 45 Å^2 for bases with overall mean values of 32.0, 30.4 and 29.2 Å^2, respectively. Overall the phosphate groups display the highest flexibility and the stacked bases are the least flexible among the DNA segments. The terminal nucleotide C1 displays the highest degree of thermal motion, with average base and sugar B values of 45 Å^2. The high thermal motion in C1 is likely the result of its poorly stabilized flipped-out conformation. Residue G5 also shows an increased degree of thermal motion than the rest of the helix with average base and sugar B values of 32.3 and 34.0 Å^2, respectively. The average thermal parameters for solvent is 41.6Å^2.

3.2.2 The Structure of d(CGGTGGCCACCG)$_2$ - cobalt hexammine form

The structure of d(CGGTGGCCACCG)$_2$ crystallized in the presence of cobalt hexammine (hereafter referred to as 12merCo) was determined to 1.75 Å resolution using the techniques described in the experimental section. The molecular and crystal structure of 12merCo is similar to that observed for 12merSp with an rms deviation between the two structures of 0.6Å, calculated for all DNA atoms. Within the crystal structure each dodecamer single strand forms base-pairing/stacking interactions with three other strands. A major duplex interaction is formed by the central ten bases 5'-GGTGGCCACC-3' of the dodecamer (Figure 3-18a). These bases are involved in Watson-Crick base-pairing interactions with a two-fold related strand — the chemical dyad axis of the 12mer coincides with a crystallographic two-fold axis. The overall conformation of the central decamer duplex is A-type. The terminal 5'-C and G-3' residues are swung away from the helix axis and assume a flipped-out orientation. The G-3' residue appears to be stabilized by forming a base stacking interaction with a symmetry related G-3' residue (Figure 3-18b). The two six-membered rings of the G-3' bases are stacked with a separation distance of approximately 3.4Å. The flipped-out terminal G-3' residue also interacts through a water mediated hydrogen-bond with a 5'-C1 base from yet another strand (Figure 3-18c). This second interaction likely plays only a minor role in the stabilization of this conformation as the 5'-C1 residue is partially disordered, as observed in 12merSp.

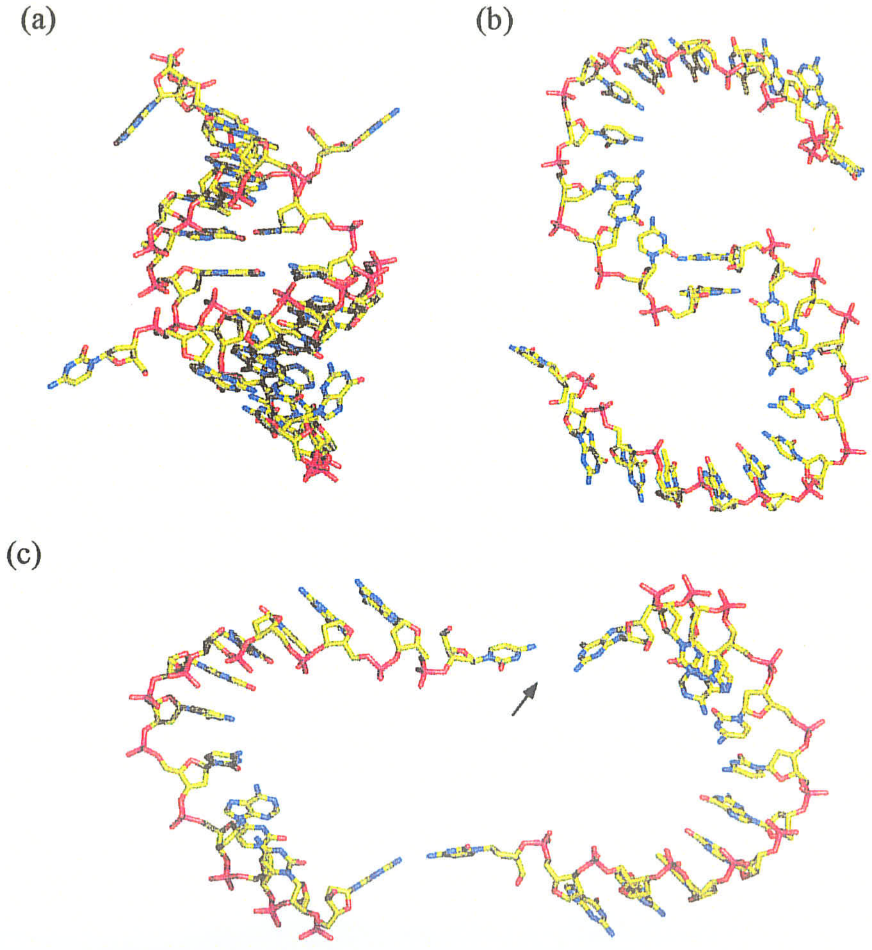

Figure 3-18 Base-pairing/stacking interactions in the 12merCo structure. (a) Primary duplex interaction, (b) G-G base-stacking and (c) water-mediated H-bonding between C1 and G12 of symmetry related duplexes.

3.2.2.1 Sugar-Phosphate Backbone

The backbone torsion angles, pseudorotation angles and sugar pucker types for the 12merCo structure are given in Table 3-3 along with values from fiber diffraction studies for canonical A- and B-DNA (Chandrasekaran & Arnott, 1989). Figure 3-19 illustrates the backbone torsion angles for the 12merCo structure superimposed on the observed ranges for other A- and B-DNA crystal structures (Schneider *et al.,* 1997).

TABLE 3-3 **Backbone torsion angles for 12merCo.**

	α	β	γ	δ	ε	ζ	χ	P (degrees)	Pucker	P_i--P_{i+1} (Å)
C1			26.7	126.1	292.2	321.2	156.2	133.8	C1'-*exo*	-
G2	293.9	199.2	63.8	86.9	205.4	293.0	192.9	17.8	C3'-*endo*	5.7
G3	145.3	186.9	188.4	86.7	212.9	281.2	182.9	13.1	C3'-*endo*	6.7
T4	292.7	176.8	50.4	80.2	212.7	289.4	198.6	16.9	C3'-*endo*	5.5
G5	294.9	167.8	51.8	79.5	195.4	288.0	196.2	12.1	C3'-*endo*	6.1
G6	295.1	179.5	54.8	82.4	211.7	285.1	196.2	21.2	C3'-*endo*	5.7
C7	296.9	169.7	48.5	81.0	204.1	296.4	201.3	24.8	C3'-*endo*	5.8
C8	148.3	183.7	183.7	84.3	203.0	289.2	196.4	11.1	C3'-*endo*	6.8
A9	296.9	198.1	51.0	145.9	183.4	274.3	247.3	165.2	C2'-*endo*	6.4
C10	276.4	170.2	65.4	84.5	182.4	279.5	211.8	60.8	C4'-*exo*	6.6
C11	293.4	196.1	51.3	136.8	192.4	251.8	238.3	147.7	C2'-*endo*	6.5
G12	220.6	135.1	47.2	146.8			249.8	160.9	C2'-*endo*	-
A-DNA^a	308	175	42	79	212	285	203	-10-40°	C3'-*endo*	5.5
B-DNA^a	330	136	31	143	219	199	262	140-185°	C2'-*endo*	6.6

^{a}Fiber diffraction data values (Chandrasekran & Arnott, 1989).

The backbone torsion angles for the central octamer structure for the most part fall within the observed range for other A-DNA structures. The nucleotides G3 and C8 also adopt the α(trans)/γ(trans) extended backbone conformation observed for 12merSp. The glycosidic torsion angles, χ, are all in the expected anti conformation.

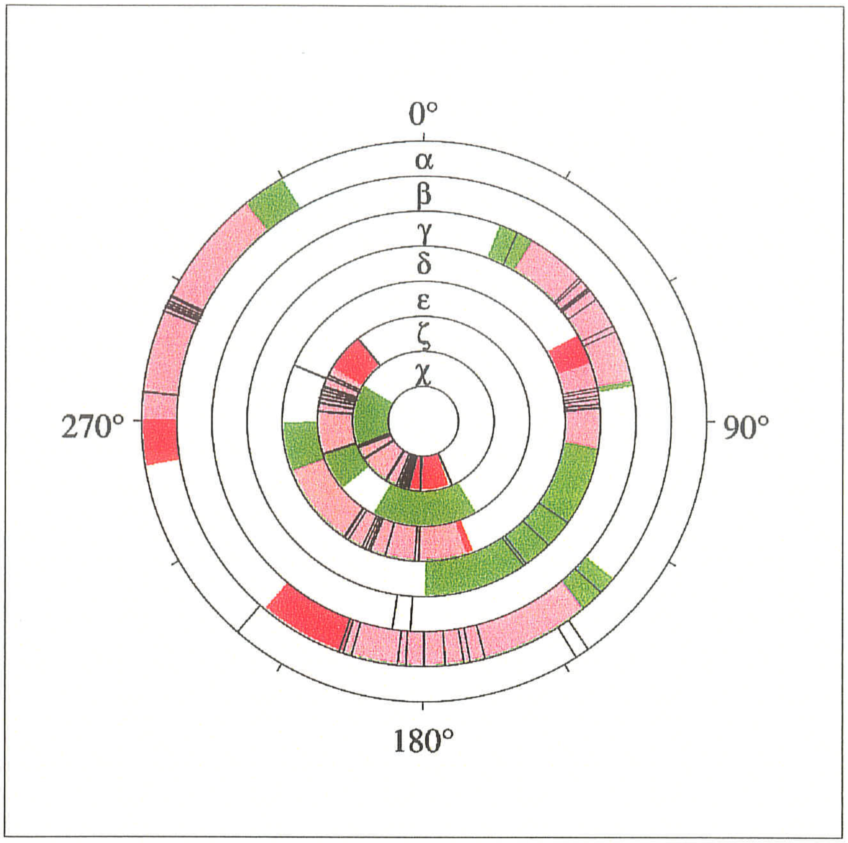

Figure 3-19 12merCo backbone torsion angle wheel.
The black lines represent the values of the 12merCo structure, shown in red are the observed ranges for A-DNA crystal structures and shown in green are the observed ranges for B-DNA crystal structures.

The 12merCo structure also exhibits a region of local sugar-phosphate backbone flexibility at the A9pC10 base-step (Figure 3-20 and Table 3-4). The backbone disorder is again associated with a concomitant change of the α and γ torsion angles of C10 from a *sc* to an *ac/ap* conformation, resulting in a 1.0 Å shift of the phosphate group. The local flexibility is not isolated to the phosphate group and movement of greater than 0.4 Å is observed in atoms O5', C5', and O3' of A9 and atom C4' of C10. Associated with the movement of these atoms are deviations in the β backbone torsion angle of A9 by 27° and pseudorotation phase angle of C10 by 23°. As in the 12merSp structure, the orientations of the bases are invariant with small changes (deviations on the order of 10°) in the glycosidic torsion angles of the two residues.

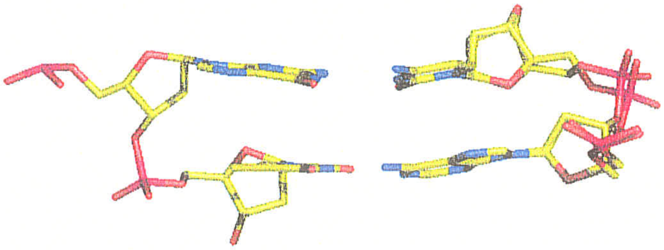

Figure 3-20 Backbone conformational flexibility in the 12merCo structure.
A view into the major groove of 12merCo at base-pairs G3-C10* and T4-A9* (the top base-pair is G3-C10*).

TABLE 3-4 Local backbone conformational flexibility in 12merCo.

	α	β	γ	δ	ε	ζ	χ	Phase (degrees)	Pucker	P_i--P_{i+1} (Å)
A9	296.9	198.1	51.0	145.9	183.4	274.3	247.3	165.2	C2'-*endo*	6.4
A9	309.9	171.5	61.5	138.6	191.3	268.0	236.9	165.5	C2'-*endo*	7.1
C10	276.4	170.2	65.4	84.5	182.4	279.5	211.8	60.8	C4'-*exo*	6.6
C10	111.9	196.2	195.8	85.8	196.5	288.1	204.1	37.3	C4'-*exo*	6.9

At the 3' end of the structure the backbone shows severe distortion, while the distortion at the 5' end of the structure is isolated to the terminal C1 residue. The distortion in the 3' end of the molecule extends from the G12 residue (through the region of local conformational flexibility) to the δ of A9. The distortion, as was observed in the 12merSp structure, is associated with the conformational changes in the DNA as the 5'-C and 3'-G residues move away from the major duplex interaction.

The sugar puckers for the nucleotides in the central region of the dodecamer are *C3'-endo*. As the backbone torsion angles undergo distortion there is a corresponding change in sugar ring conformation from that of pure *C3'-endo* type. For the sugar groups attached to bases A9, C11 and G12 we observe a *C2'-endo* conformation and for the sugar group attached to base C10 we observe a *C4'-exo* conformation. The change in sugar conformation to *C2'-endo* results in a change in the P-P intrastrand separation starting at the C8-A9 base step (see Table 3-3). The change in sugar pucker and concommitant increase in P-P separation create an extended DNA backbone conformation which allows for the terminal G-3' residue to form a stable packing interaction with another terminal G-3' from a symmetry related strand. A change in the sugar pucker type is also observed for the 5'-terminal base (residue C1 adopts a C1'-*exo* conformation).

3.2.2.2 Helix Morphology and Base Stacking in the Central Decamer

The helix axis for the 12merCo central decamer duplex is also non-linear, with bending of the helix axis occurring near the ends in the direction of the DNA major groove. Bending of the helix axis towards the major groove results from the asymmetrical conformation of the sugar-phosphate backbone at the G2•C11*/G3•C10* and G5•C8*/G6•C7* base-steps where only one strand of the duplex exists in the extended α, γ conformation. Further asymmetrical distortions leading to bending of the helix axis result from the change in sugar pucker at residue A9 to the *C2'-endo* conformation. The total angle of curvature (CURVES, Lavery & Skelnar, 1989) is 36.5°. The curvature of the helix axis results in an overall shortening of the helix pathlength by 7.7%. The overall pattern of axis curvature is similar to that observed for the 12merSp structure (Figure 3-5 and Figure 3-21).

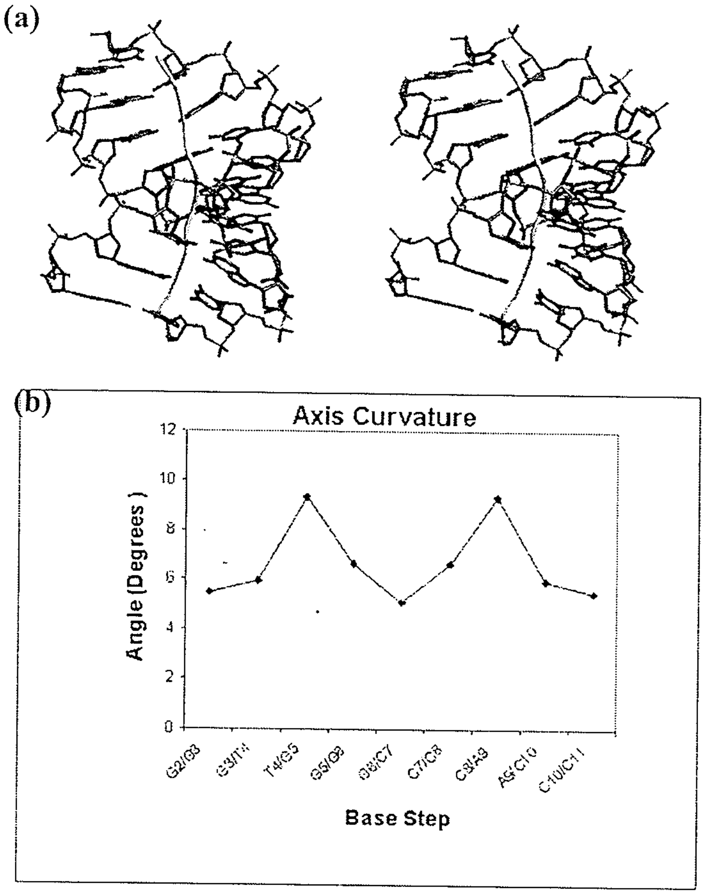

Figure 3-21 12merCo helix axis curvature.
(a) Stereo pairs of the central decamer demonstrating bending with the helix axis calculated by CURVES, (b) a graph of the angle of curvature per base step as calculated by CURVES.

Graphs of the helical parameters calculated for the central decamer duplex using a curved helical axis are shown in Figure 3-22. Based on the curved helix the average rise/base pair is 2.7 Å, and the average helical twist/base pair is 30.1° (corresponding to 12 residues per turn). As in the 12merSp structure, there is a base-pair rise for the G3•C10/ T4•A9 step of 3.35 Å (Figure 3-22a) without a substantial increase in twist. The increased axial rise is again accommodated by the bending of the helix in this region. A large negative roll is also associated with the G3•C10/T4•A9 base step (see Figure 3-22d) as well as an opening of the base pairs towards the major groove (see Figure 3-23f).

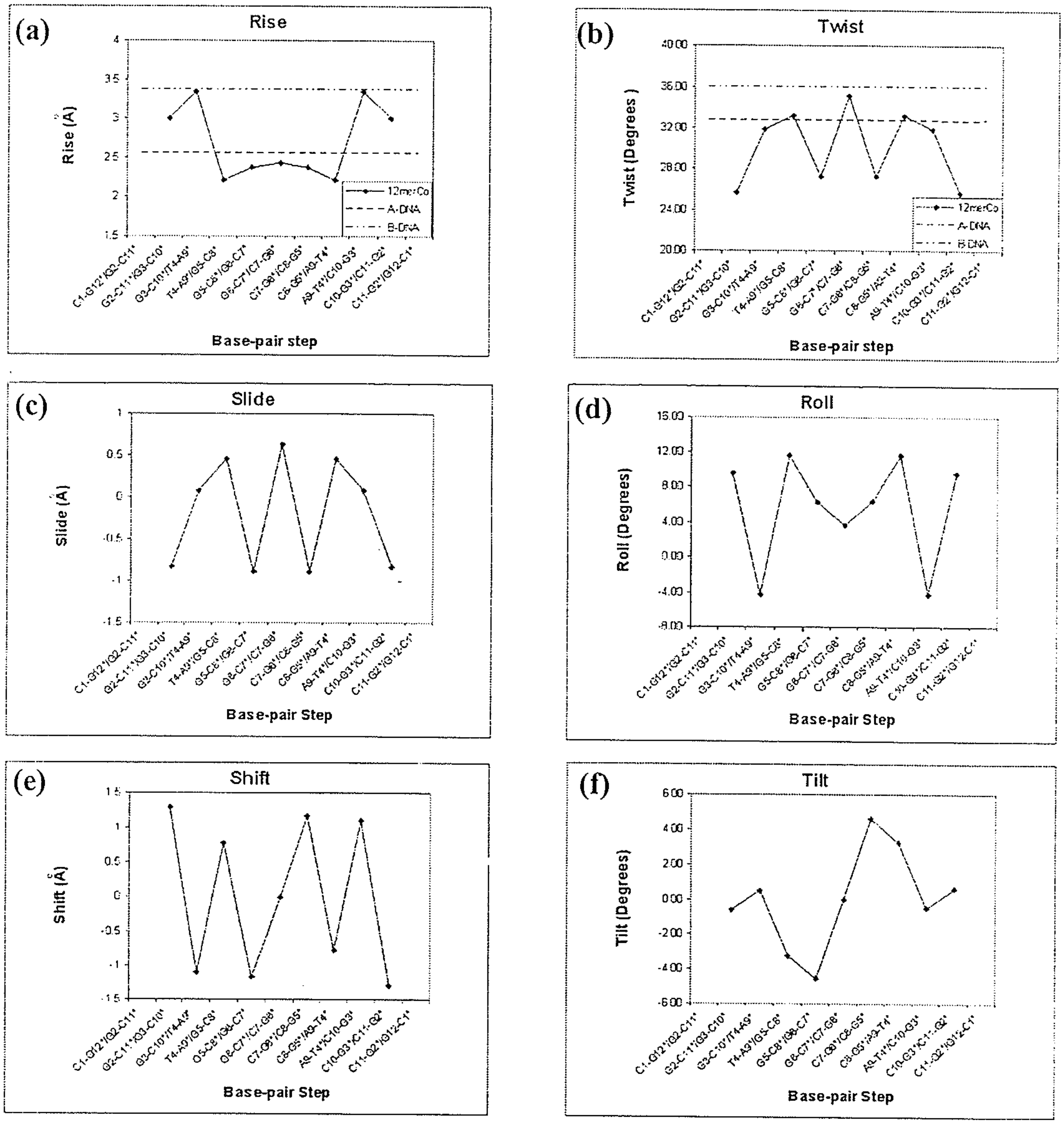

Figure 3-22 Plots of the global helical parameters for the 12merCo structure. Plots showing the variations in (a) rise, (b) twist, (c) slide, (d) roll, (e) shift, and (f) tilt across the helix.

Graphs of the base-pair parameters are shown in Figure 3-23. The central base-pairs have adopted a relatively undistorted conformation, as observed in the 12merSp structure, and the outer base pairs are considerably more distorted. A large increase in base-pair buckle is observed for the outer residues (Figure 3-23b) which corresponds with asymmetrical conformation of the duplex backbone. The α,γ–extended backbone conformation is likely also responsible for the opening of base pairs G3•C10* and G5•C8* towards the major groove. The opening of these base pairs results in a slight distortion of the inter-base hydrogen bonds, with the minor groove H-bonds being shortened and the major groove H-bonds being lengthened.

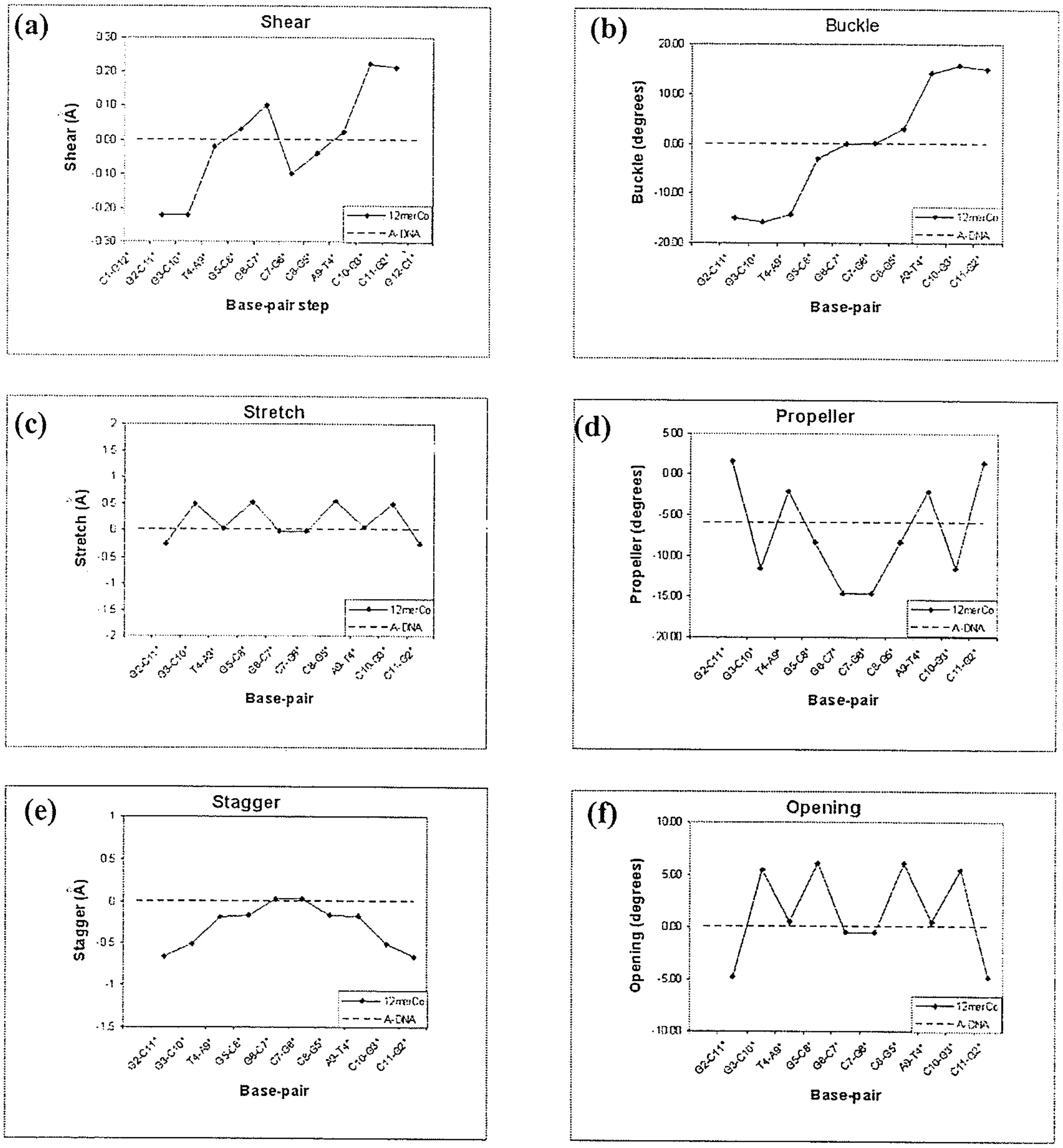

Figure 3-23 Plots of the global base-pair parameters for the 12merCo structure. Plots showing the variation in (a) shear, (b) buckle, (c) stretch, (d) propeller, (e) stagger, and (f) opening along the helix.

3.2.2.3 Crystal Packing

The unit cell packing is illustrated in Figure 3-24 and Figure 3-25. While the overall packing mode is very similar to the 12merSp crystal structure, the volume per base-pair in the 12merCo structure is reduced by 20Å^3 which leads to only minor changes in the packing contacts as well as in the orientation of the 5'-C residue. Three types of packing interactions are observed in the 12merCo crystal structure. The first set of interactions involve the flipped-out cytosine residues and penultimate base-pairs occupying the shallow minor groove of two neighbouring duplexes. The next set of interactions results from the flipped-out guanine bases which are tucked into the minor groove of two other neighbouring duplexes. The final set of interactions involve two more neighbouring duplexes (interlocked through their flipped out guanine bases) which insert their flipped out cytosines and penultimate base-pairs into the minor groove of the central DNA duplex.

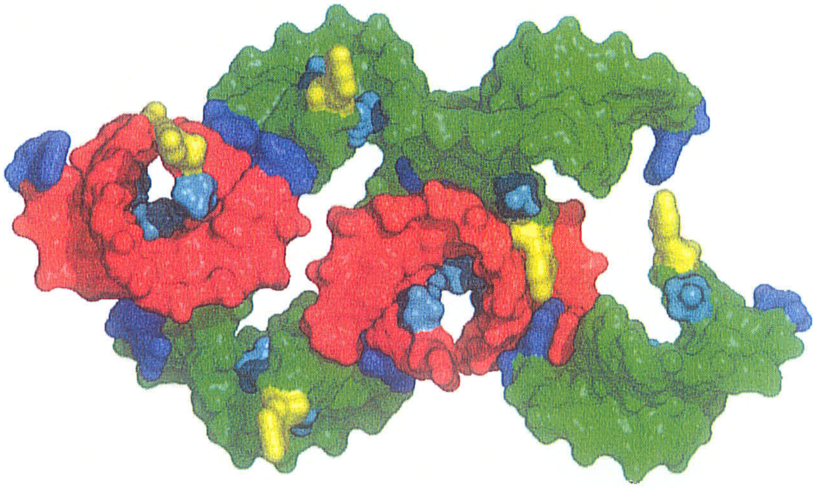

Figure 3-24 Surface packing diagram for the 12merCo crystal structure. The occupation of the minor groove of the central DNA molecule (red) by the C1 (yellow) and G12 (blue) residues is illustrated as well as the location of the cobalt hexammine ions (teal).

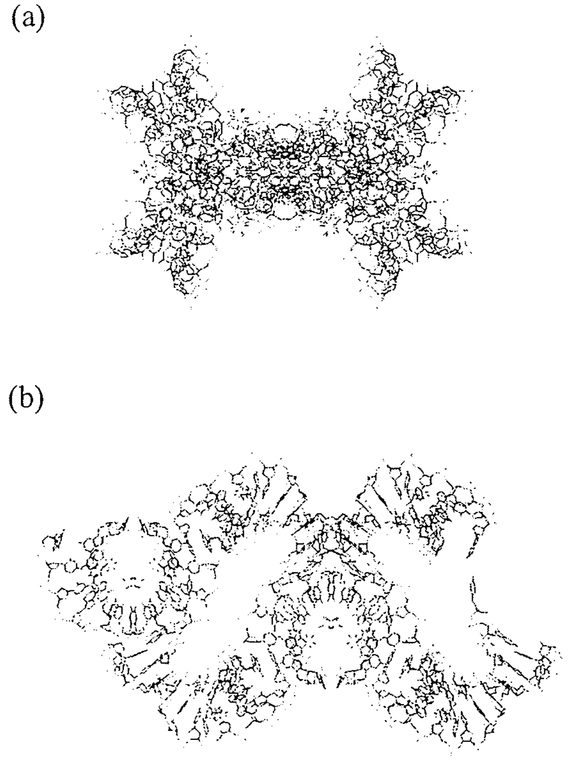

Figure 3-25 Unit cell packing in 12merCo.
Two views of the unit cell packing for 12merCo: (a) as viewed down the *c* axis and (b) parallel the *c* axis.

3.2.2.4 Structure of the Terminal Base Pairs

The terminal guanine, as observed in the 12merSp structure, is rotated away from the helix axis towards the minor groove to form a small binding pocket. In this conformation the C8 atom of residue G12 makes a 3.3 Å contact with O3' of the phosphate backbone at residue C11. The small minor groove binding pocket is again occupied by a symmetry-related 3'-terminal guanine. The terminal guanines are stacked such that their six-membered rings overlap with an interbase separation distance of 3.4 Å (Figure 3-26). The hydrophobic surface of the symmetry-related terminal guanine is in close contact with the sugar-phosphate backbone while the N2 amino group forms an H-bonding interaction the O2 carbonyl of residue C11 (Figure 3-27).

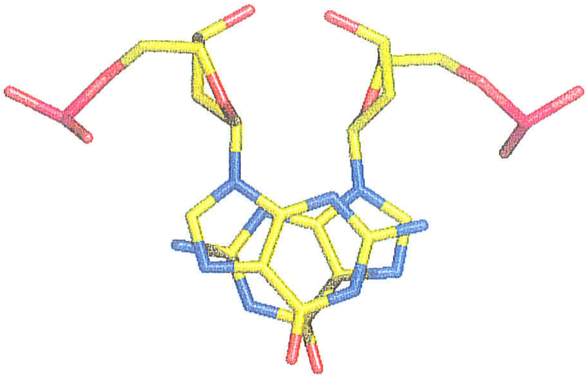

Figure 3-26 12merCo 3'-terminal guanine base-stacking.

The 5'-terminal cytosine is oriented such that it lies in the minor groove of a neighbouring duplex. The surface of the cytosine lies against the sugar-phosphate backbone while the polar base atoms O2 and N4 are involved in water-mediated hydrogen-bonding to a nearby 3'terminal guanine (Figure 3-28).

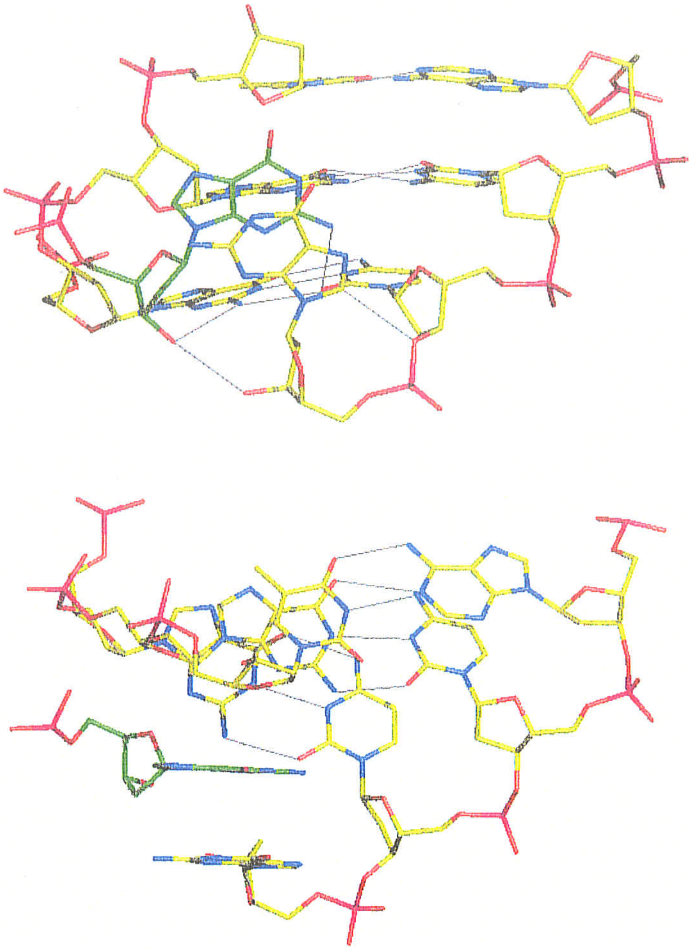

Figure 3-27 Minor groove interactions of the 12merCo terminal guanine.
Two views of the flipped-out guanine interactions with the minor groove: (a) as viewed into the minor groove and (b) perpendicular to the guanine bases. Base pairs T4-A9*, G3-C10*, G2-C11* and base G12 of the central DNA duplex are shown in yellow and the symmetry related terminal guanine in green.

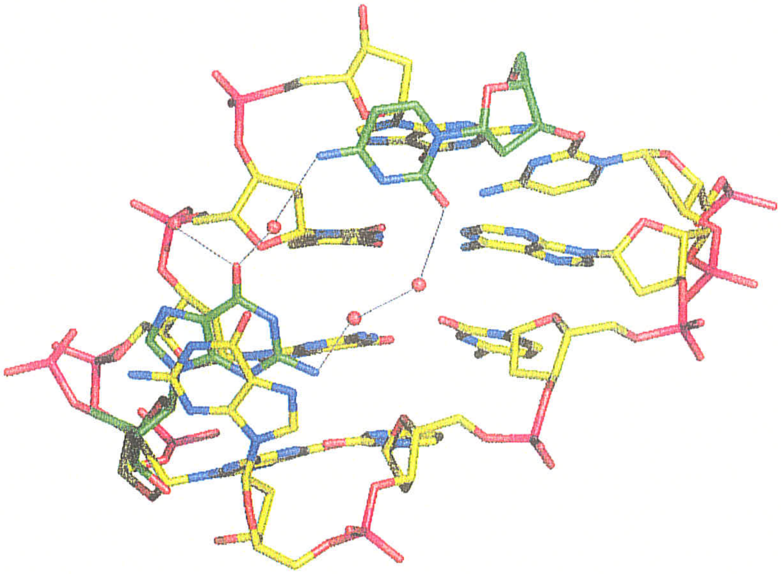

Figure 3-28 Water-mediated H-bonding in the minor groove of 12merCo.
Water-mediated hydrogen-bonding between flipped-out terminal residues C1 and G12 as viewed into the minor groove of the central DNA duplex. The central duplex is shown in yellow, the symmetry related terminal bases in green and the water molecules as red spheres

3.2.2.5 Cobalt Hexammine Binding

One and a half cobalt hexammine ions are observed in the structure per single strand of the DNA dodecamer. The hydrogen-bonding interactions between the ammine ligands and the DNA are shown in Figure 3-29. The first $[Co(NH_3)_6]^{3+}$ ion lies within the major groove and coordinates to residues G2 and G3 of the DNA. The ammine ligands of the ion hydrogen bond directly to the N7 and O6 atoms of both residues. The second $[Co(NH_3)_6]^{3+}$ ion exhibits a disordered binding mode. The disordered $[Co(NH_3)_6]^{3+}$ ion is located near the crystallographic twofold axis at the centre of the deep major groove where it binds to the G5-G6 bases. However, with the chemical dyad axis of the DNA lying on a crystallographic twofold axis, a twofold-related $[Co(NH_3)_6]^{3+}$ ion binds to the

G5*-G6* bases within the same major groove. The two symmetry-related sites are too close together for both $[Co(NH_3)_6]^{3+}$ ion binding sites to be occupied at the same time. Thus only a single $[Co(NH_3)_6]^{3+}$ ion can bind to a GGCC:CCGG sequence. A similarly disordered cobalt hexammine binding mode was observed in the structure of d(ACCG-GCCGGT) (Gao *et al.*, 1995).

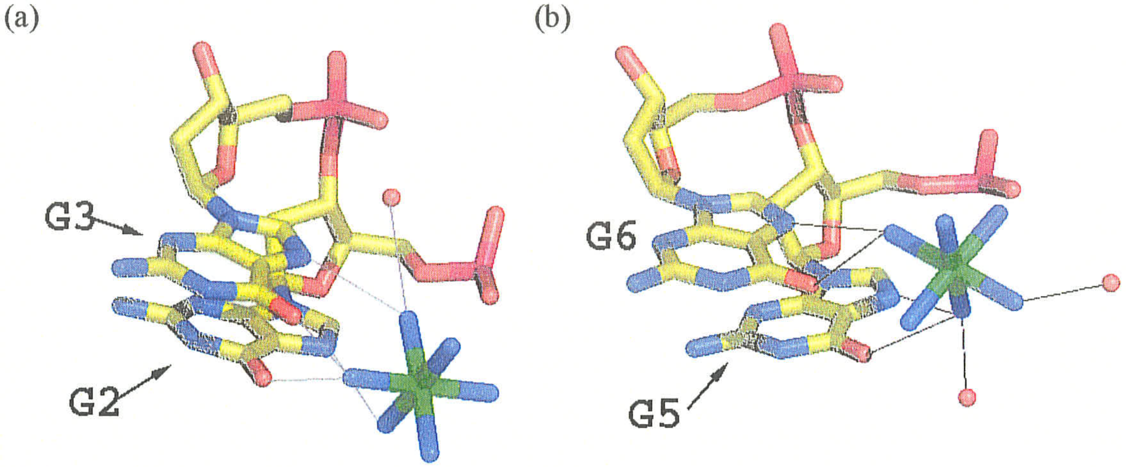

Figure 3-29 Cobalt hexammine binding.
The hydrogen-bonding interactions from the cobalt hexammine ion to the hydrogen-bond donor groups of the dodecamer base-steps (a) G2/G3 and (b) G5/G6 as well as to water molecules in the major groove.

3.2.2.6 Hydration

The asymmetric unit of the 12merCo crystal contains 37 water molecules. The major groove of the 12merCo structure is extensively hydrated. 14 water molecules (per DNA single strand) line the major groove and make contacts with polar DNA base atoms and/or other water molecules (Figure 3-30a,b). The minor groove hydration shell is almost completely displaced by the hydrophobic surfaces of the penultimate and terminal base pairs. The 3 water molecules in the shallow minor groove are involved primarily in the water-mediated H-bonding between the flipped-out terminal bases(Figure 3-30c). Figure 3-31 illustrates the water contacts made with the polar base atoms in the major and minor grooves. The sugar-phosphate backbone is economically hydrated, with on average only 1.5 water molecules per phosphate group.

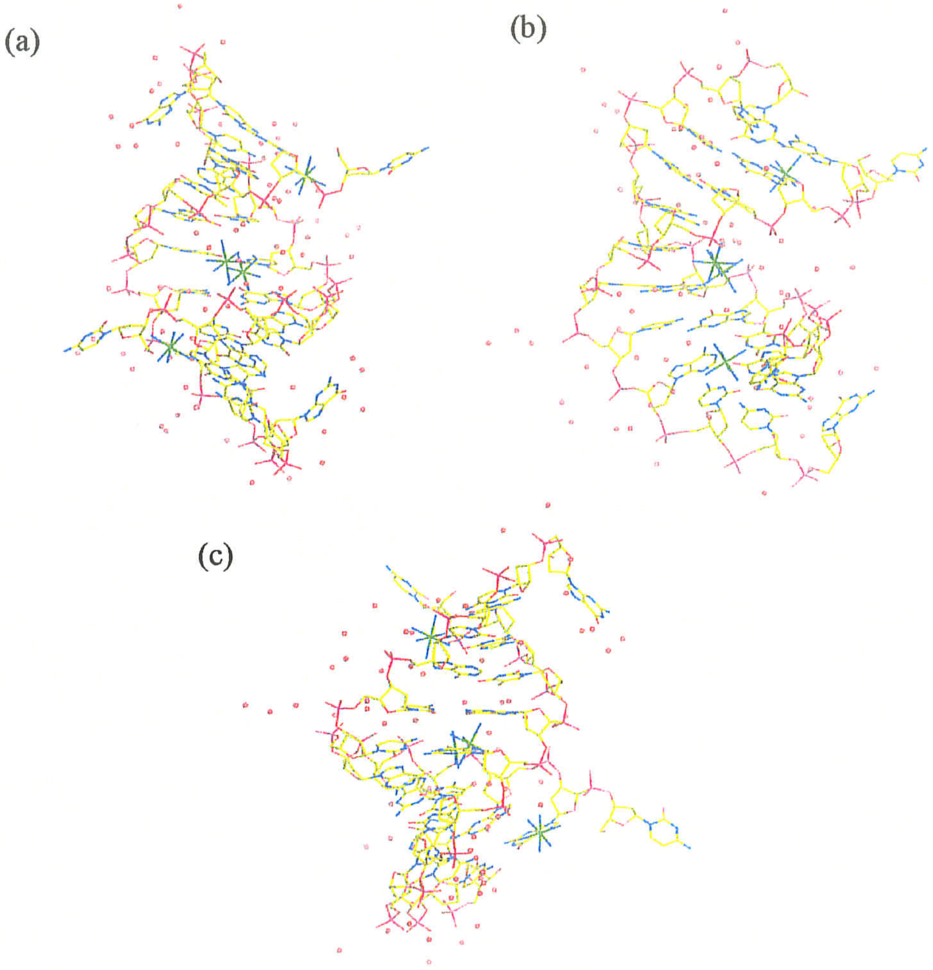

Figure 3-30 12merCo hydration.
Hydration of the 12merCo duplex structure as viewed (a) into the major groove, (b) rotated 90° and (c) into the minor groove.

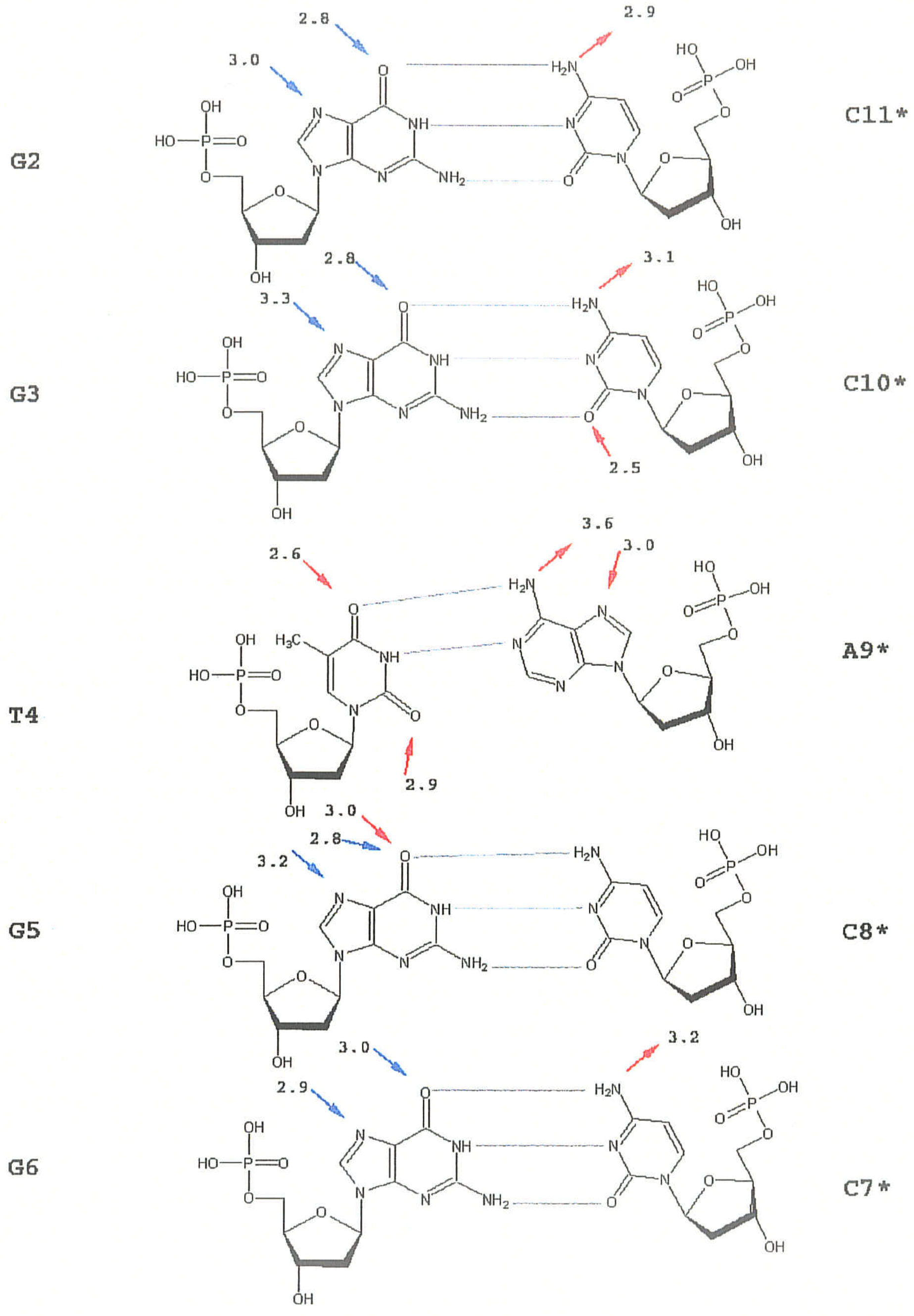

Figure 3-31 Groove hydration structure in 12merCo.
The water H-bonds are shown in red and the cobalt hexammine H-bonds in blue.

3.2.2.7 Thermal Parameters

In 12merCo average isotropic temperature factors range from 21.5 to 40 Å^2 for phosphate groups, from 19 to 40 Å^2 for deoxyribose units, and from 19 to 40 Å^2 for bases with overall mean values of 27.0, 25.1 and 24.0 Å^2, respectively. Overall the phosphate groups display the highest flexibility and the stacked bases are the least flexible among the DNA atoms. The terminal nucleotide C1 displays the highest degree of thermal motion, with average base and sugar B values of 40 Å^2. The high thermal motion in C1 is likely the result of its poorly stabilized flipped-out conformation. The average thermal parameters for solvent is 35.2 Å^2.

3.3 Structures and Biological Implications

3.3.1 Structural Comparison

Despite the differences in crystallization conditions, the two structures, 12merSp and 12merCo, are remarkably similar with a rms deviation for all DNA atoms of 0.6 Å (Figure 3-32). The two structures exhibit slightly altered patterns of helix bending as illustrated in Figure 3-33. The 12merCo structure shows increased bending in the centre of the helix where the disordered cobalt hexammine binds. Significant deviations between the two structures exist in the sugar-phosphate backbone of residues A9 and C10, a region where both structures exhibit local conformational flexibility (Figure 3-34). As a result of the movement of the sugar-phosphate backbone in this region, the phosphate group of the partially disordered cytosine must alter its position between the two structures in order to alleviate close packing contacts between symmetry-related duplexes. The orientation of the flipped-out cytosine residue also differs significantly. Thus the cytosine residue has adapted its conformation in response to the changes in its local environment.

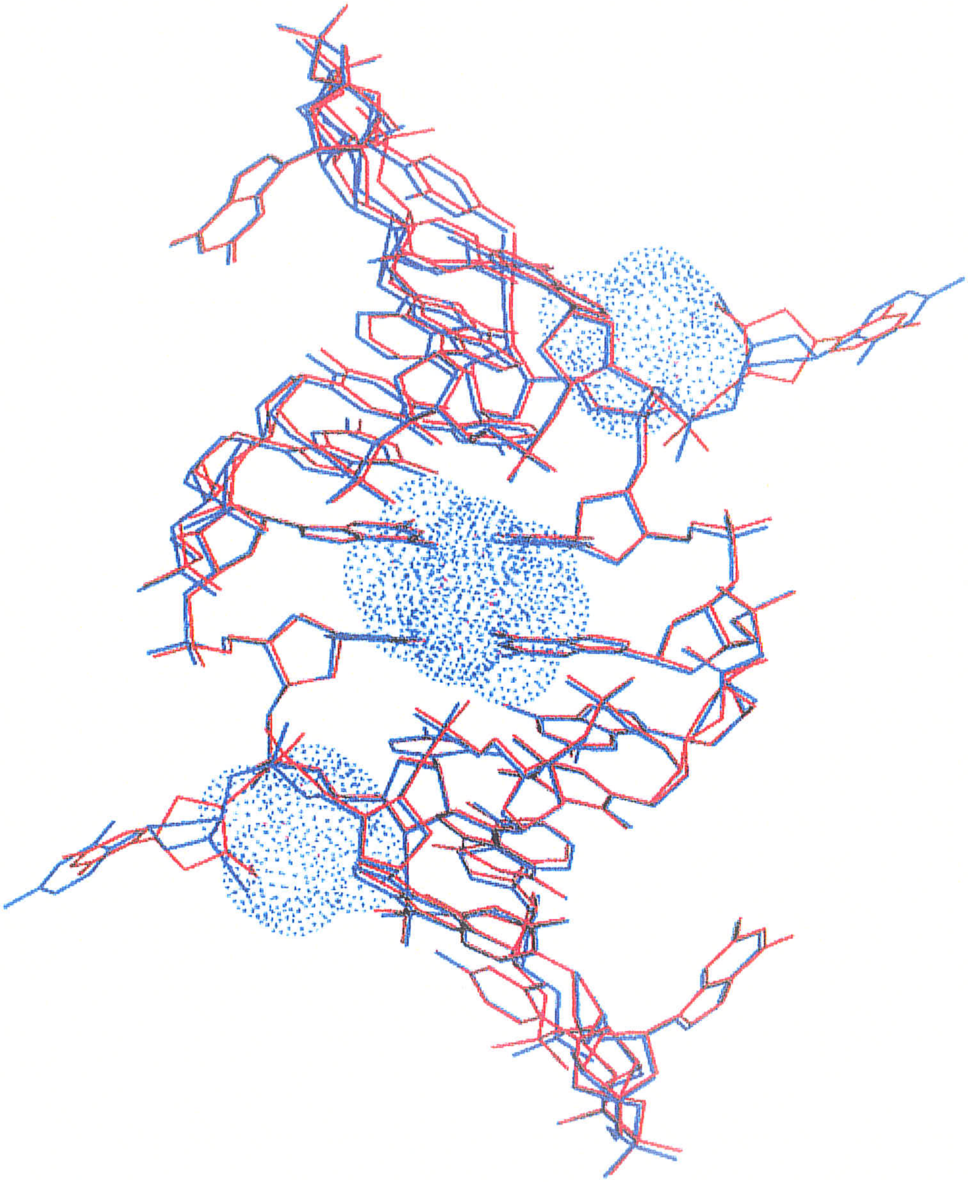

Figure 3-32 Least-squares fit of 12merSp and 12merCo structures.
Overlay of the structures as viewed into the minor groove showing the van der Waals volume around the cobalt hexammine ions. 12merSp shown in red and 12merCo in blue.

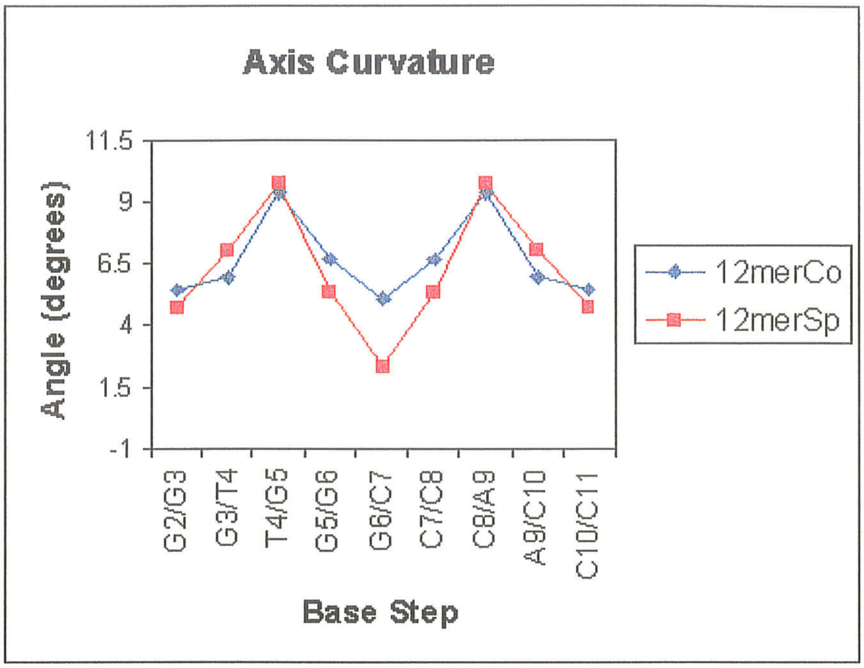

Figure 3-33 Comparison of the axis of curvature of 12merSp and 12merCo.
As calculated by CURVES.

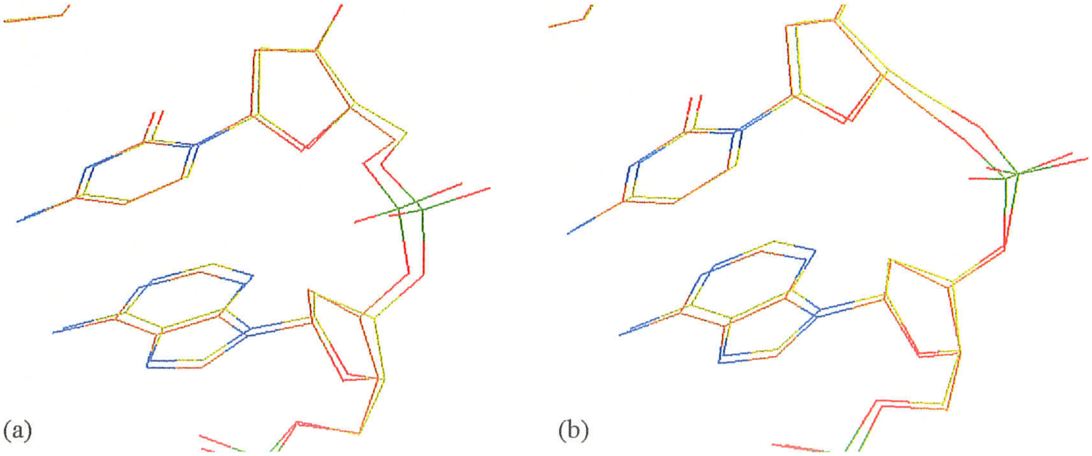

Figure 3-34 Overlay of 12merSp and 12merCo at the A9/C10 base step.
An overlay of the alternate backbone conformations at residue C10:(a) g+g- backbone conformation and (b) all trans conformation.

A comparison of selected local base-pair and helical parameters (Figure 3-35 & Figure 3-36) reveals that the differences between the two structures are primarily associated with the bases involved in cobalt hexammine binding. It appears that the DNA backbone is able to make compensations for the variable local conformations to achieve a similar global conformation.

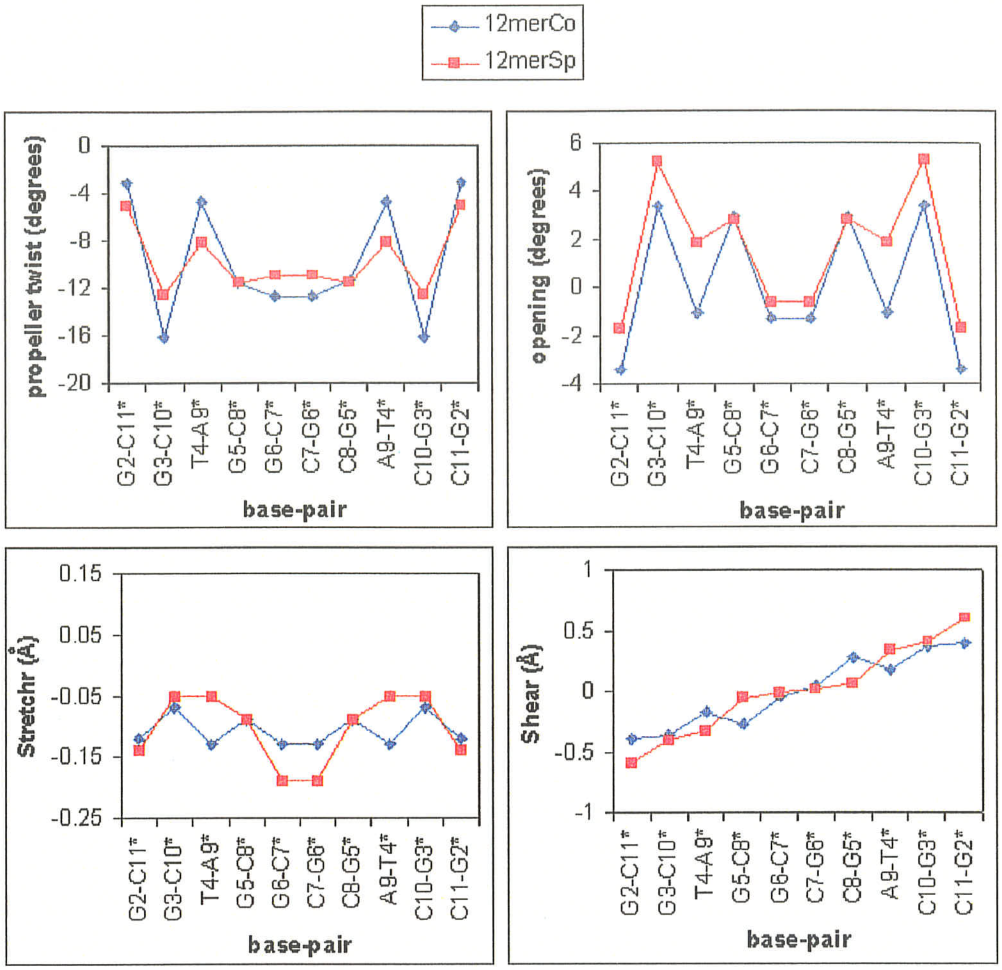

Figure 3-35 Selected local base-pair parameters.
Calculations carried out using the program 3DNA (Lu *et al.*, 2000).

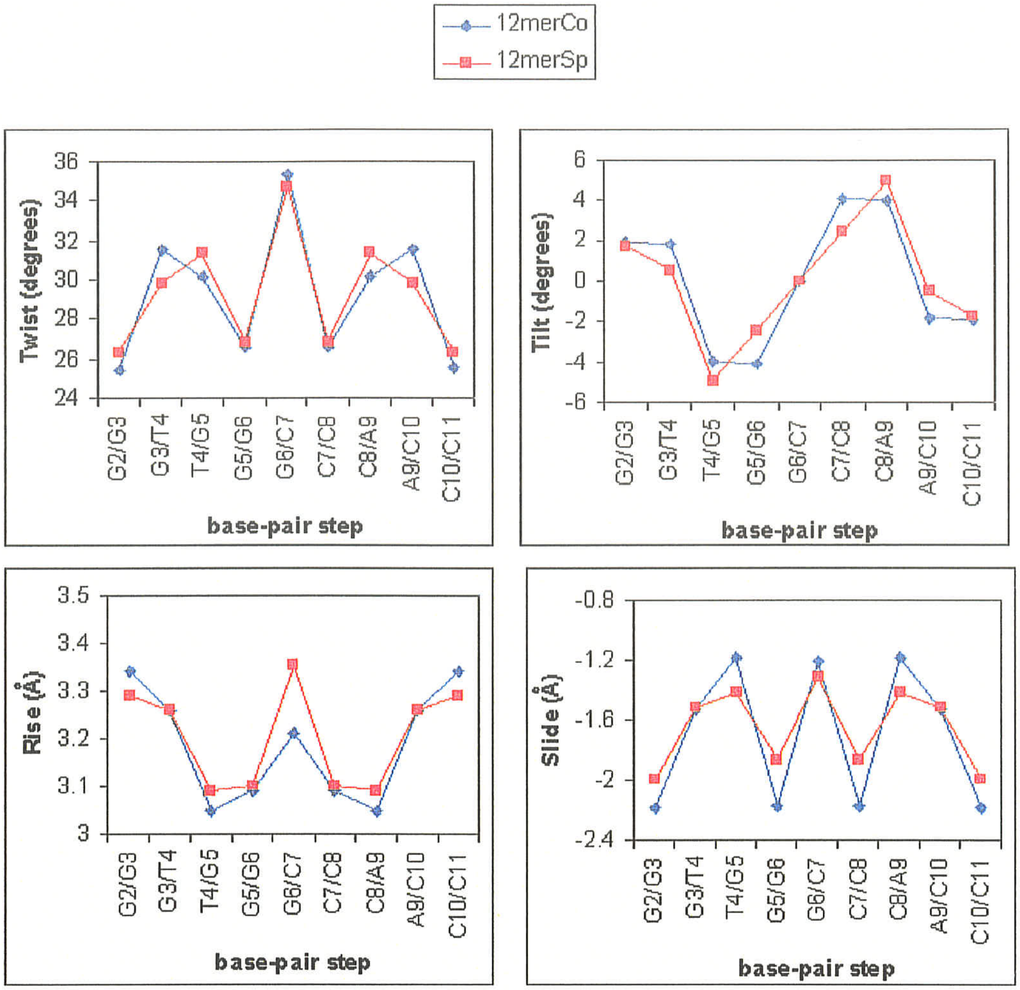

Figure 3-36 Selected local base-step parameters.
Measurements as determined by the program 3DNA (Lu *et al.*, 2000).

There are 12 solvent sites common to the two structures which are located primarily in the major groove (Figure 3-17 & Figure 3-31). In the case of the G5/G6 base step, one of the cobalt hexammine ions in the 12merCo crystal structure displacing the water molecules found in the 12merSp crystal structure. The hydration of the sugar-phosphate backbone is somewhat more variable due to the dependence of the hydration patterns on the intra-phosphate separation distances (Feig & Pettitt, 1998).

3.3.2 The Influence of Crystal Packing

The origin of the high quality of the X-ray diffraction data from the 12merSp and 12merCo crystals becomes clear when examining how the DNA molecules are packed within the unit cell. Two DNA duplexes interlocked through their flipped-out guanine residues are able to act as a molecular handle and latch on to the minor groove of a third duplex. Through this motif, propagation of the growth of highly ordered crystals is achieved. Thus the cylindrical pseudo-symmetry normally encountered in the crystallization of DNA oligomers has been broken.

The local conformational flexibility in both crystal structures is likely the result of the crystal packing arrangement. Figure 3-37 illustrates the atoms in both structures which exhibit movement of greater than 0.4 Å. There is a close-packing interaction on the minor groove side of the phosphate moiety of residue A9 which results in the movement of the sugar-phosphate backbone between residues A9 and C10 which resides at the surface of a small solvent channel.

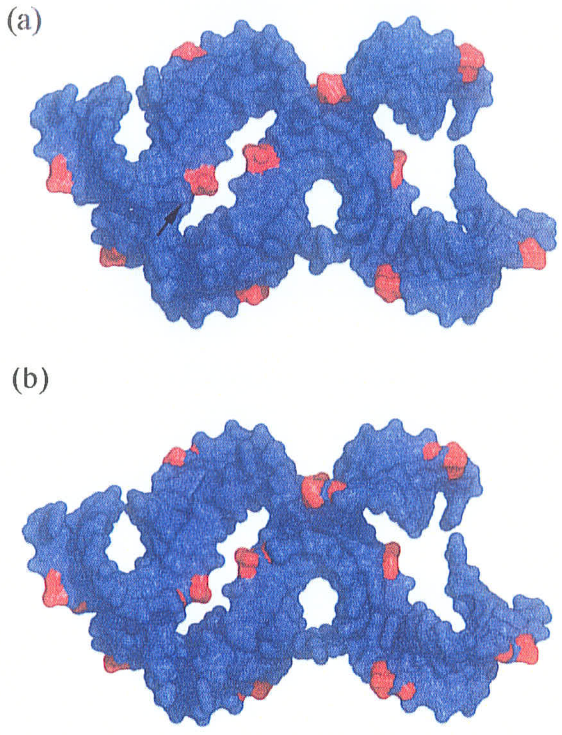

Figure 3-37 Local Conformational flexibility surface packing.
Surface packing diagrams for the (a)12merSp (b)12merCo crystal structures illustrating in red the atoms which exhibit local conformational flexibility of greater than 0.4 8Å, the arrow points to the phosphate group of residue A9.

3.3.3 Biological Implications

3.3.3.1 The Significance of the A-form

The biological significance of crystal structures of DNA oligomers in the A-form have traditionally been questioned. This is due to the belief that the B-form is the biologically relevant conformation for DNA, while the A conformation is the biologically relevant form for RNA molecules. In more recent years, crystal structures of protein-DNA complexes have revealed the DNA duplex to be surprisingly malleable (Dickerson, 1998). The TATA-binding protein, for instance, induces an A-type conformation in the transcriptional promoter of eukaryotes (J. L. Kim *et al.*, 1993, Y. Kim *et al.*, 1993). The deformability of DNA is an important determinant of its ability to interact with proteins (Dickerson and Chin, 1998). There is also exists a growing number of DNA crystal structures that exhibit properties of both A- and B-DNA (Wang *et al.*, 1982; Malinina *et al.*, 1999; Ng *et al.*, 2000; Vargason *et al.*, 2001). Crystallographic studies have shown that DNase I can bind to both an A-DNA and a B-DNA sequence, however, it can only cleave the B-DNA sequence (Suck *et al.*, 1988; Lahrn & Suck, 1991; Weston *et al.*, 1992). The A↔B conformational conversion in DNA is a phenomenon that raises important biological and chemical questions and deserves further study.

The structures of 12merSp and 12merCo are the second instance of a DNA oligomer with a 5'-terminal CG starting sequence crystallizing in an A-type conformation. The CG starting sequence is common to oligonucleotides crystallized in the B-form. The first DNA oligomer with a CG starting sequence to crystallize in the A-form was d(CGCCCGCGGGCG), which exhibits a highly kinked A-DNA structure with a B-type base-step at the terminal CG base-step (Malinina *et al.*, 1999). Malinina *et al.* speculated that the unusual structure of the terminal base-pairs may be an intrinsic feature of the CG base-step when placed at the ends of oligonucleotides in the A-form, or alternatively the distortion of the entire molecule may result in the stretching of the terminal base-step conformation. In the 12merSp and 12merCo structures the terminal base-pairs also do not exist in an A-type conformation, but instead are flipped-out away from the helix axis. The sugar moiety of the 3'-terminal residue has adopted a C2'-*endo* conformation, commonly observed in B-DNA. Crystallographic studies of other 5'-CG sequences are required to further investigate this anomaly.

3.3.3.2 Base-Flipping

The structure of d(CGGTGGCCACCG) in the 12merSp and 12merCo crystal structures has a central ten base tract 5'-GGTGGCCACC-3' which forms Watson-Crick basepairs with a 2-fold related strand resulting in a primarily A-type decamer duplex. Distortion of the DNA occurs in both structures close to the 5' and 3' ends of the oligomer to give two terminal base pairs which are oriented away from the helix axis. The flipped out 5'-C and 3'-G bases are stabilized by base stacking and water-mediated H-bonding interactions with the terminal residues from symmetry-related strands. The structure of the DNA decamers d(AGGCATGCCT) (Nunn & Neidle, 1996) and d(AGGGGCCCCT) (Gao *et al.*, 1999) show a similar orientation for the terminal 5'-A and 3'-T bases which are swung out of the duplex to form Watson-Crick base-pairing interactions with flipped out bases from symmetry-related strands. While the existence of the flipped out bases in an A-type structure is not unique, the interlocking of the duplexes through the flipped out purine bases has not been observed before.

Base-flipping is the phenomenon whereby a base in DNA is swung completely out of the helix into an extra-helical position where it can interact with the concave catalytic binding pocket of the protein. Base-flipping by DNA enzymes was first observed in the structure of a DNA cytosine 5-methyltransferase bound to substrate DNA (Klimasauskas *et al.*, 1994). Since that time, base-flipping has been found in some of the enzyme-DNA complexes involved in the base excision repair pathway and predicted in others where an enzyme requires access to an individual DNA base (Roberts & Cheng, 1998). There is little hard data about the mechanism of base flipping as no intermediates have yet been characterized. There are currently two main theories. The first theory suggests that the flipping is an active process in which the base is pushed out of the helix from the minor groove side by an appropriate residue in the enzyme. Once flipped, the base is pulled into the active site of the enzyme and trapped. The second theory is that during the normal breathing of DNA, the bases naturally spend time in a completely flipped-out position, and it is the transient conformation in the DNA that is recognized and caught by the protein.

The existence of both native DNA and protein-DNA complexes containing flipped-out bases demonstrates that DNA molecules are very flexible and will readily undergo this type of deformation at any required position along a DNA duplex.

3.3.3.3 A-T Base-pair Opening and Future Work

In an effort to verify the A•T opening observed in the crystal structure of d(CGGTGG)/d(CCACCG), the crystal structure of d(CGGTGGCCACCG) was determined. In the 2.5 Å crystal structure of the non-self-complementary hexamer the DNA helix adopted a B-type conformation with significant structural anomalies at the GTG site (Tari & Secco, 1995). The A•T opening observed yielded structural features which could serve as a signal for DNA-protein recognition. In the crystal structures of 12merSp and 12merCo the DNA helix adopted an A-type conformation. While no structural anomalies were found in the region of the GTG trimer, this does not dismiss the possibility of A•T base-pair opening in the B-DNA form of this sequence. While the backbone of A-form helices has been shown to be somewhat malleable, the sequence dependent conformational flexibility of the bases tends to be less prominent in A-DNA than in B-DNA (Gao *et al.*, 1995).

It has been shown that certain DNA sequences may favour the A-type conformation (Frederick *et al.,* 1987). In particular, the GpG base step has been shown to favour the A-form, with poly(G)·poly(C) existing almost exclusively in the A-from (Arnott & Selsing, 1974). The source of this stabilization of the A-form has been linked to the base-pair stacking. In the A-DNA geometry the guanines can optimally stack with the five-membered ring over six-membered ring. In order to obtain a B-form GTG containing sequence it may be necessary to change our DNA construct to reduce the number of GpG base steps. The self-complementary decamer sequence d(CGGTGCACCG), while still containing the GTG trimer of interest, contains two fewer GpG base steps. The reduced number of GpG steps are expected to yield a sequence which would be more likely to crystallize in the B-form and provide further evidence for the A•T base-pair opening.

The influence of divalent cations on the DNA conformation may provide access to the B-form of d(CGGTGGCCACCG). Metal ion coordination to nucleic acids is not only required for charge neutralization, it is also essential for the biological function of nucleic acids. The structural impact of the coordination of different metal ions on DNA helices has yet to be determined fully. Among divalent cations the effects of Cu^{2+} and Cd^{2+} on DNA structure and stability are relatively clear cut. They are known to mediate the unstacking of the bases and produce a random coil form by binding to the nucleotide bases at high enough concentrations (Saenger, 1984). The effects of the alkaline earth cations, such as Mg^{2+} and Ca^{2+}, are equally clear cut, with their stabilization effects resulting from binding primarily to the phosphate backbone. The effects of the transition metal ions such as Zn^{2+}, Ni^{2+}, Co^{2+}, and Mn^{2+} are less clear cut. They possess fairly strong affinities for both the nucleotide bases as well as the phosphate backbone. As a result they are known to selectively stabilize some ordered structures while destabilizing others depending on the ligand strength of the nucleotide base as well as the preferred conformation of the DNA oligomer. The preliminary diffraction data on the zinc and cobalt nitrate crystal forms seemed to indicate the DNA had adopted a B-type conformation. These cations may to bind at the N7 positions of successive GpG base steps (Froystein *et al.*, 1993) and may interfere with the formation of the A-DNA geometry. Further investigation of the 12mer crystals grown in the presence of Zn^{2+} and Co^{2+} ions are expected to yield the B-DNA 12mer structure as well as to provide more information on the interaction of these transition metal ions with DNA oligomers.

While the main goal of investigation of the deformability of the A•T base-pair in a B-DNA context was not achieved, a greater understanding of the deformability of the A-DNA backbone was gained. The interlocking of the 3'-terminal guanine residues in the minor groove is a novel packing motif and has implications for minor groove recognition. Design of DNA-binding drugs, based on the interactions of cobalt hexammine with the dodecamer, which bind preferentially to GpG base steps and lock the conformation of the DNA into an A-type structure may have applications in turning off gene expression in diseasese such as cancer.

APPENDIX A

DNA Structure and Nomenclature

DNA is a linear polymer that is made up of monomers called nucleotides. The nucleotide is constructed from three components: a nitrogenous base, a 2'-deoxyribose sugar, and a phosphate group. There are four types of bases found in DNA: the purines, adenine (A) and guanine (G), and the pyrimidines, thymine (T) and cytidine (C). The bases and their numbering scheme are shown in Figure A-1. Each base is connected to a sugar via a β glycosyl linkage between the N(9) atom in purines or the N(1) atom in pyrimidines to the C1' atom of the sugar moiety. The nucleotide units are connected via phosphodiester linkages between the O3'and O5' atoms and the phosphate group. The connectivity and numbering scheme for the nucleotide unit is shown in Figure A-2.

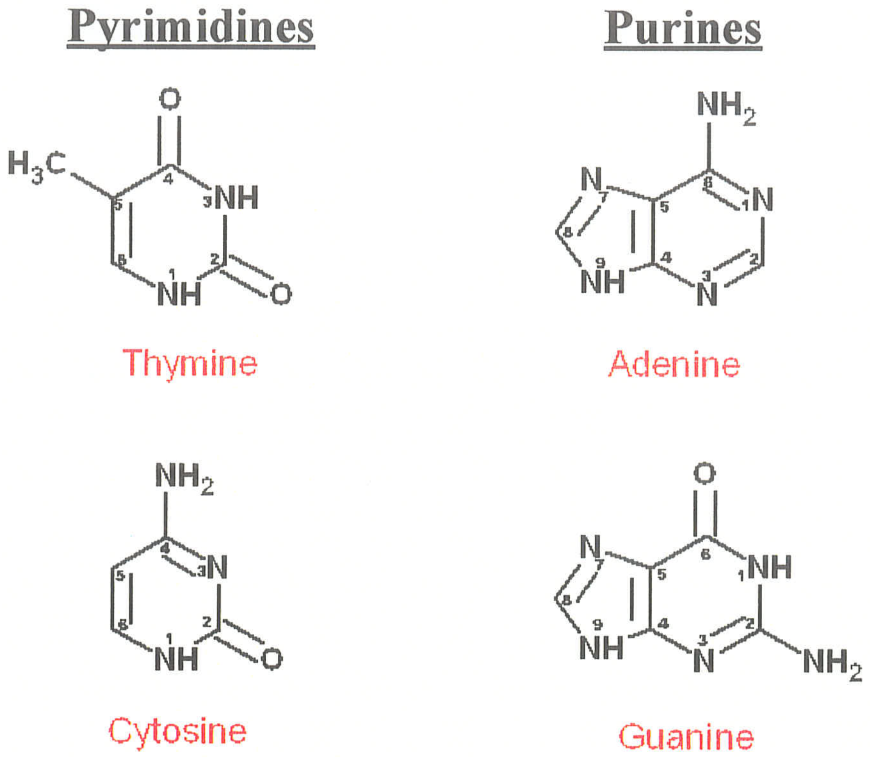

Figure A-1 Structure of the pyrimidine and purine bases.

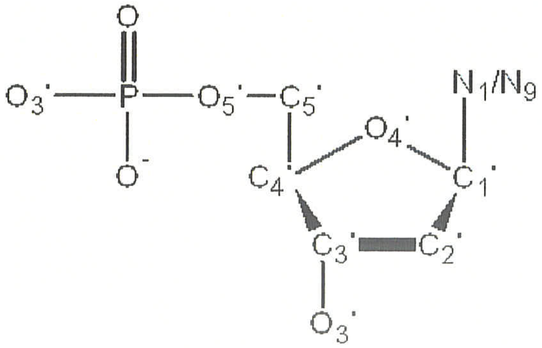

Figure A-2 Structure of a 2'deoxyribose nucleotide.

The overall geometry of the DNA double-helix is a consequence of the conformational preferences of the individual nucleotide units. The geometry of the nucleotides is described by the phosphate backbone torsion angles α, β, γ, δ, ε, and ζ, the sugar conformation, and the glycosidic torsion angle χ (Figure A-3). The deoxyribose sugar adopts a conformatation in which one of the five ring atoms lie out of the plane defined by the other four. This is termed an 'envelope' or E conformation. The convention is to name the conformation after the out-of-plane atom and then to specify whether it is out of plane on the same side as the C5' atom (*endo*) or the opposite side (*exo*). Ten such conformations exist: five *endo* and five *exo*, which can be generalized to a continuous distribution of intermediate conformations, characterized by a pseudorotation angle, P (Altona & Sundaralingam, 1972).

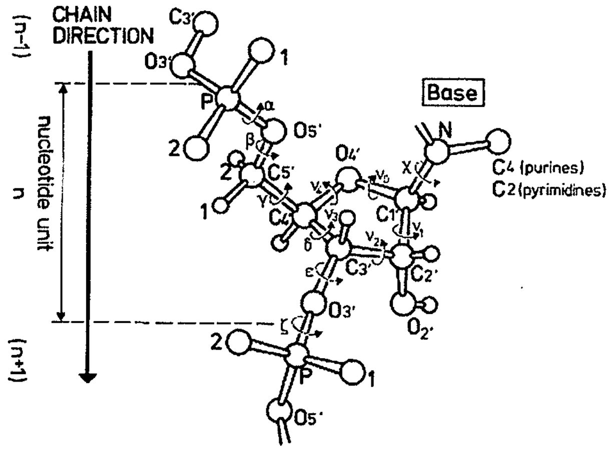

Figure A-3 IUPAC torsion angle notation for polynucleotide chains.
(reproduced from Saenger, 1984)

The results of fiber and single crystal x-ray diffraction studies have shown that DNA is capable of a high degree of structural polymorphism, and that the global DNA structure is intrinsically related to the extent of hydration of the double-helix. DNA can adopt two major right-handed conformations, the A- and B-forms (Figure A-4). All double helical DNA possess characteristic grooves, whose dimensions depend on the DNA conformation. The minor groove is on the

O2 (pyrimidine) or N3 (purine) side of the base-pair while the major groove is on the opposite side. A-DNA is a stable right-handed double-helix that is favoured by certain sequences (*e.g.* $(dG)_n(dC)_n$) or under low humidity (Benevides *et al.*, 1986; Mazur, 1989). The A-form has a wide and compact helix, with a deep, narrow major groove and a wide, shallow minor groove. The base-pairs in an A-DNA helix are displaced off-axis. In aqueous solution and *in vivo*, DNA is found almost exclusively in the B-form. The B-form is slimmer and more elongated than the A-form with base pairs essentially perpendicular to the helix axis. The major groove in B-DNA is wide and the minor groove is narrow. The base-pairs sit directly on the helix axis so that the major and minor grooves are of comparable depth.

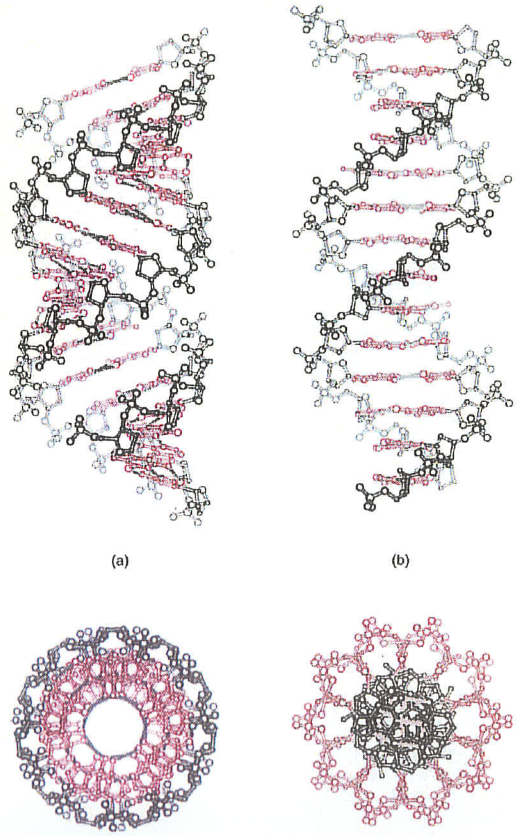

Figure A-4 A- and B-DNA structures.
(a) Ideal A-DNA and (b) ideal B-DNA as viewed into the major groove (top) and down the helix axis (bottom) (reproduced from Saenger, 1984).

A common point of reference is required to describe the three-dimensional arrangements of bases and base-pairs in nucleic acid structures. The geometry of a dinucleotide step is completely defined by: (*i*) the base-pair parameters that describe the relative position and orientation of one-base with respect to the other in a standard Watson-Crick base-pair (Figure A-7), and (*ii*) the step parameters that describe the relative position and orientation of the two base-pairs (Figure A-5). The base-pair step parameters can be calculated either with respect to local helical axes (Figure A-6) between the two base-pairs, the so-called local base-step parameters, or with respect to an overall helix axis, the global base-step parameters.

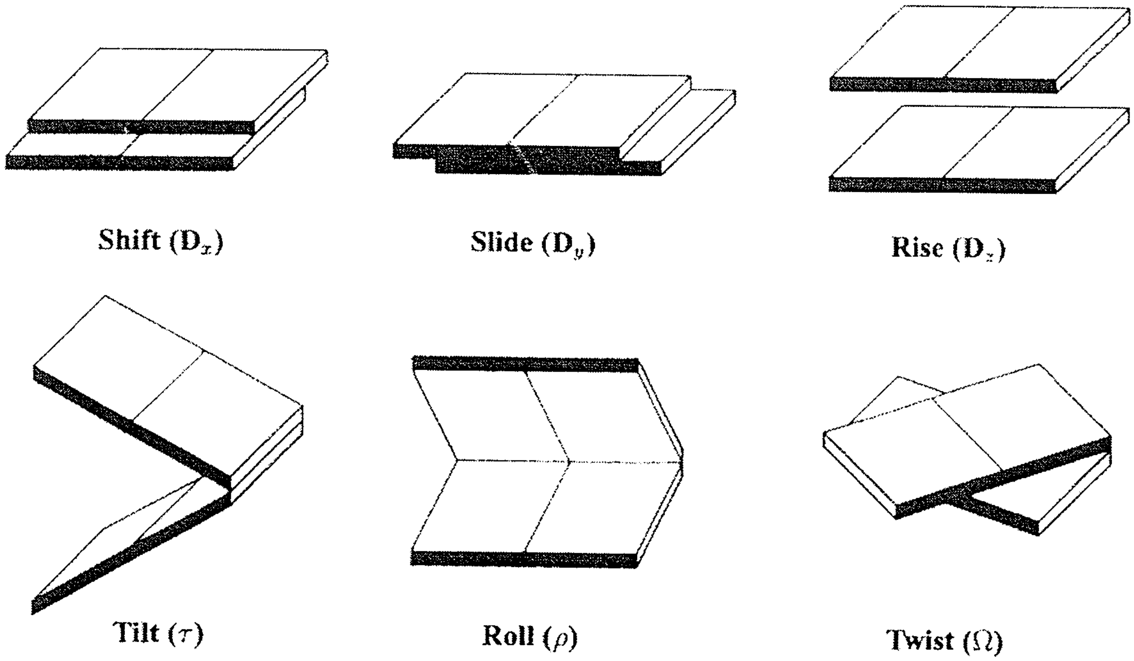

Figure A-5 Base-step parameters.
The top row describes the various translations involving two successive base-pairs and the bottom row describes the various rotations between successive base-steps.A positive tilt opens the angle between base-pairs towards strand I (see Figure A-6). A positive roll opens the angle between base-pairs towards the minor groove. (reproduced from Dickerson, 1989).

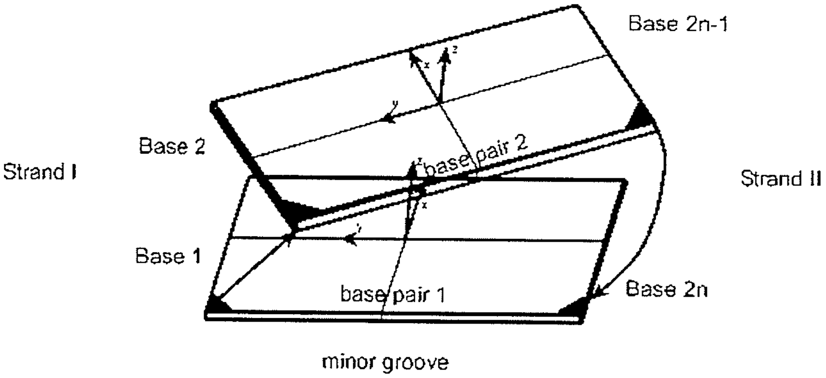

Figure A-6 Definition of the local reference axes (x,y,z).
Definition of (x,y,z) shown for the first two base-pairs of an n bae-pair double helix. The shaded corners indicate the bonds to sugar C1' atoms. Curved arrows indicate 5'-to-3' direction of each backbone strand.

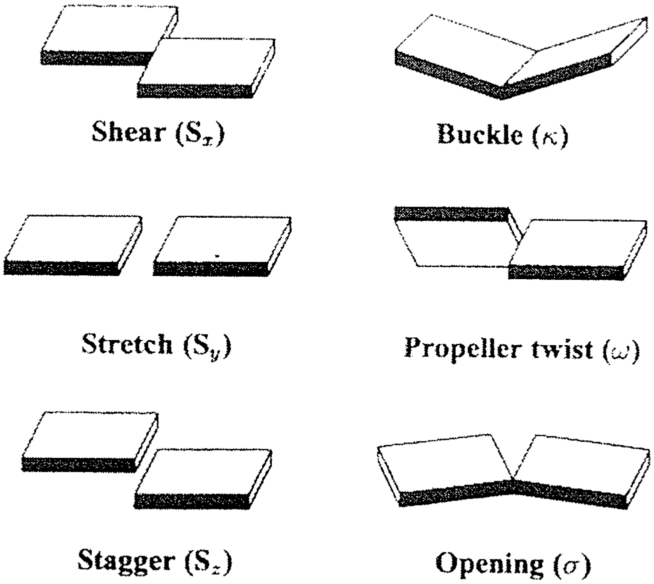

Figure A-7 Base-pair parameters.
The left column describes various translations involving two bases and the right column describles the rotations (reproduced from Dickerson, 1989).

References:

Aboul-ela, F., Koh, D., Jr., I. T. and Martin, F. H. (1985) *Nucleic Acids Research,* **13,** 4811-4824.

Adams, P. D., Pannu, N. S., Read, R. J. and Brunger, A. T. (1997) *Proc. Natl. Acad. Sci. USA,* **94,** 5018-5023.

Adams, P. D., Pannu, N. S., Read, R. J. and Brunger, A. T. (1999) *Acta Cryst.,* **D55,** 181-190.

Adhya, S. and Miller, W. (1979) *Nature,* **279,** 492-494.

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

Arnott, S. and Selsing, E. (1974) *J. Mol. Biol.,* **88**, 551-552.

Avizonis, D. Z. and Kearns, D. R. (1995) *Biopolymers,* **35,** 187-200.

Baikalov, I. and Dickerson, R. E. (1998) *Acta Cryst.,* **D54,** 324-333.

Benevides, J. M., Wang, A. H.-J., Rich, A., Kyogoku, Y., van der Marel, G. A., van Boom, J. H., and Thomas, G. J. Jr. (1986) *Biochem.*, **25**, 41-50.

Berger, I., Kang, C., Sinha, N., Wolters, M. and Rich, A. (1996) *Acta Cryst,* **D52,** 465-468.

Bhattacharyya, A. and Lilley, D. M. J. (1989) *J. Mol. Biol.,* **209,** 583-597.

Bingman, C., Jain, S., Zon, G. and Sundaralingam, M. (1992a) *Nucleic Acids Research,* **20,** 6637-6647.

Bingman, C. A., Zon, G. and Sundaralingam, M. (1992b) *J. Mol. Biol.,* **227,** 738-756.

Boulard, Y., Cognet, J. A. H. and Fazakerley, G. V. (1997) *J. Mol. Biol.,* **268,** 331-347.

Brennan, S. and Cowan, P. L. (1992) *Rev. Sci. Instrum.*, **63**, 850-853.

Brown, T., Leonard, G. A., Booth, E. D. and Kneale, G. (1990) *J. Mol. Biol.,* **212,** 437-440.

Brunger, A. T. (1991) *Annu. Rev. Phys. Chem.,*197-223.

Brunger, A. T. (1992) *Nature,* **355,** 472-475.

Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-

Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, M. M., Simonson, T. and Warren, G. L. (1998) *Acta Cryst.,* **D54,** 905-921.

Brunger, A. T. and Krukowski, A. (1990) *Acta Cryst.,* **A46,** 585-593.

Brunger, A. T., Kuriyan, J. and Karplus, M. (1987) *Science,* **235,** 458-460.

Bustamante, C., Guthold, M., Zhu, X. and Yang, G. (1999) *J. Biol. Chem.,* **274,** 16665-16668.

Bush, C. Allen (1974) in *Basic Principles in Nucleic Acid Chemistry* (Ed. Paul Tso), Academic Press Inc., New York, **Vol II**, pp. 92-165.

Chadrasekaran, R. and Arnott, S. (1996) *J. Biomol. Struct. Dynam.,* **13,** 1015-1027.

Chandrasekaran, R. and Arnott, S. (1989) *Landolt-Bornstein Numerical Data and Functional Relationships in Science and Technology (Group VII, Biophysics),* **Subvolume VII 1b,** 31.

Chen, M. (1997) *Structural Study of the DNA fragment d(CGCTGGCCACCG) and a priliminary crystallographic study of the self-complementary DNA sequence d(CGGTGGCCACCG);* M. Sc. Thesis, Department of Chemistry, University of Manitoba, Winnipeg.

Chen, X., Mariappan, S. V. S., Catasti, P., Ratliff, R., Moyzis, R. K., Laayoun, A., Smith, S. S., Bradbury, E. M. and Gupta, G. (1995) *Proc. Natl. Acad. Sci. USA,* **92,** 5199-5203.

Choy, H. E. and Adhya, S. (1992) *Proc. Natl. Acad. Sci. USA,* **89,** 11264-11268.

Choy, H. E., Park, S.-W., Parrack, P. and Adhya, S. (1995) *Proc. Natl. Acad. Sci. USA,* **92,** 7327-7331.

Crick, F. (1970) *Nature,* **227,** 561-563.

Davis, T. M., McFail-Isom, L., Keane, E. and Williams, L. D. (1998) *Biochem.,* **37,** 6975-6978.

Dickerson, R. E. (1989) *J. Biomol. Strct. Dynam.*, **6**, 627-634.

Dickerson, R. E. (1998) *Nucleic Acids Research,* **26,** 1906-1926.

Dohet, C., Wagner, R. and Radman, M. (1985) *Proc. Natl. Acad. Sci. USA,* **82,** 503-505.

Drenth, J. (1994) *Principles of Protein X-ray Crystallography,* Springer-Verlag, New York.

Egli, M., Tereshko, V., Teplova, M., Minasov, G., Joachimiak, A., Sanishvili, R., Weeks, C. M., Maier, M. A., An, H., Cook, P. D. and Manoharan, M. (1998) *Biopolymers,* **48,** 234-252.

Eisenstein, M. and Shakked, Z. (1995) *J. Mol. Biol.,* **248,** 662-678.

Evans, G. and Pettifer, R. F. (2001) *J. Appl. Cryst.,* **34,** 82-86.

Fasman, G. O. (1976) *Handbook of Biochemistry and Molecular Biology: Nucleic Acids,* CRC Press, Cleveland.

Feig, M. and Pettitt, B. M. (1998) *Biopolymers,* **48,** 199-209.

Frederick, C., Quigley, G. J., Teng, M.-K., Coll, M., van der Marel, G. A.,van Boom, J. H. van., Rich, A. and Wang, A. H.-J. (1989) *Eur. J. Biochem.,* **181,** 295-307.

Frederick, C. A., Saal, D., van der Marel, G. A., van Boom, J. H., Wang, A. H. and Rich, A. (1987) *Biopolymers,* **26 Suppl,** S145-S160.

Froystein, N. A., Davis, J. T., Reid, B. R. and Sletten, E. (1993) *Acta Chemica Scandinavica,* **47,** 649-657.

Gao, Y.-G., Robinson, H., Boom, J. H. v. and Wang, A. (1995) *Biophysical Journal,* **69,** 559-568.

Gao, Y.-G., Robinson, H. and Wang, A. H.-J. (1999) *Eur. J. Biochem.,* **261,** 413-420.

Goodrich, J. A. and McClure, W. R. (1992) *J. Mol. Biol.,* **224,** 15-29.

Harran, T. E., Shakked, Z., Wang, A. H.-J. and Rich, A. (1987) *Journal of Biomolecular Structure and Dynamics,* **5,** 199-217.

Hendrickson, W. A. (1985) *Methods in Enzymology,* **115,** 252-270.

Hunter, W. N., Brown, T., Anand, N. N. and Kennard, O. (1986) *Nature,* **320,** 552-555.

Irani, M., Musso, R. and Adhya, S. (1989) *J. Bacteriology,* **171,** 1623-1630.

Ivanov, V. I., Minchenkova, L. E., Schyolkina, A. K. and Poletayev, A. I. (1973) *Biopolymers,* **12,** 89-110.

Jack, B. A. and Levitt, M. (1978) *Acta Cryst,* **A34,** 931-935.

Jain, S., Zon, G. and Sundaralingam, M. (1987) *J. Mol. Biol.,* **197,** 141-145.

James, R. W. (1982) *The Optical Principles of the Diffraction of X-rays,* Ox Bow Press, Woodbridge Connecticut.

Kejnovsky, E. and Kypr, J. (1998) *Nucleic Acids Res.,* **26,** 5295-5299.

Kim, J. L., Nikolov, D. B. and Burley, S. K. (1993) *Nature,* **365,** 520-527.

Kim, Y., Geiger, J. H., Hahn, S. and Sigler, P. (1993) *Nature,* **365,** 512-520.

Kleywegt, G. J. and Brunger, A. T. (1996) *Structure,* **4,** 897-904.

Kleywegt, G. J. and Jones, T. A. (1998) *Acta Cryst.,* **D54,** 1119-1131.

Klimasauakas, S., Kumar, S., Roberts, R. J. and Cheng, X. (1994) *Cell,* **76,** 357-369.

Konnert, J. H. and Hendrickson, W. A. (1980) *Acta Cryst.,*, **A36,** 344-350.

Kouchakdjian, M., Li, B. F. L., Swann, P. F. and Patel, D. J. (1988) *J. Mol. Biol.,* **202,** 139-155.

Kramer, B., Kramer, W. and Fritz, H.-J. (1984) *Cell,* **38,** 879-887.

Kremer, E. J., Pritchard, M., Lynch, M., Yu, S., Holman, K., Baker, E., Warren, S. T., Schlessinger, D., Sutherland, G. R. and Richards, R. I. (1991) *Science,* **252,** 1711-1714.

Lahm, A. and Suck, D. (1991) *J. Mol. Biol.*, **221**, 645-647.

Lavery, R. and Sklenar, H. (1989) *J. Biomol. Struct. Dynam.,* **4,** 655-667.

Lee, J. S. and Latimer, L. J. P. (1993) *Biochem. Cell Biol.,* **71,** 162-168.

Lu, P., Cheung, S. and Arndt, K. (1983) *Journal of Biomolecular Structure and Dynamics,* **1,** 509-521.
Lu, X.-J., Shakked, Z. and Olson, W. K. (2000) *J. Mol. Biol.,* **300,** 819-840.

Malinina, L., Fernandez, L. G., Huynh-Dihn, T., and Subirana, J. A. (1999) *J. Mol. Biol.*, **285,** 1679-1690.

Mazur, J. (1989) *Biopolymers,* **28,** 1223-1233.

McRee, D. E. (1999) *Journal of Structural Biology,* **125,** 156-165.

Murshudov, G. N., Vagin, A. A. and Dodson, E. J. (1997) *Acta Cryst.,* **D53,** 240-255.

Ng, H.-L., Kopka, M. L. and Dickerson, R. E. (2000) *Proc. Natl. Acad. Sci. USA,* **97,** 2035-2039.

Nunn, C. M. (1996) *Journal of Biomolecular Structure and Dynamics,* **14,** 49-55.

Nunn, C. M. and Neidle, S. (1996) *J. Mol. Biol.,* **256,** 340-351.

Olson, W. K. (1982) *Nucleic Acids Research,* **10,** 777-787.

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., Sklenar, H., Suzuki, M., Tung, C.-S., Westhof, E., Wolberger, C. and Berman, H. M. (2001) *J. Mol. Biol.,* **313,** 229-237.

Otwinowski, Z. and Minor, W. (1997) *Methods in Enzymology,* **276,** 307-326.

Pannu, N. S., Murshudov, G. N., Dodson, E. J. and Read, R. J. (1998) *Acta Cryst.,* **D54,** 1285-1294.

Pannu, N. S. and Read, R. J. (1996) *Acta Cryst.,* **A52,** 659-668.

Parkinson, G., Vojtechovsky, J., Clowney, L., Brunger, A. T. and Berman, H. M. (1996) *Acta Cryst.,* **D52,** 57-64.

Peyret, N., Seneviratne, P. A., Allawi, H. T. and SantaLucia, J. Jr. (1999) *Biochem.,* **38,** 3468-3477.

Ramakrishnan, B. and Sundaralingam, M. (1993) *J. Mol. Biol.,* **231,** 431-444.

Rice, L. M. and Brunger, A. T. (1994) *Proteins: Structure, Function, and Genetics,* **19,** 277-290.

Roberts, R. J. and Cheng, X. (1998) *Annu. Rev. Biochem.,* **67,** 181-198.

Robinson, H., Gao, Y.-G., Sanishvili, R., Joachimiak, A. and Wang, A. H.-J. (2000) *Nucleic Acids Research,* **28,** 1760-1766.

Robinson, H. and Wang, A. H.-J. (1996) *Nucleic Acids Research,* **24,** 676-682.

Rossman, M. G. and Blow, D. M. (1962) *Acta Cryst.,* **15,** 24-31.

Rouzina, I. and Bloomfield, V. A. (1998) *Biophysical Journal,* **74,** 3152-3164.

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

Schneider, B., Neidle, S. and Berman, H. M. (1997) *Biopolymers,* **42,** 113-124.
Sines, C. C., McFail-Isom, L., Howerton, S. B., VanDerveer, D. and Williams, L. D. (2000) *J. Am. Chem. Soc.,* **122,** 11048-11056.

Suck, D., Lahrn, A. and Oefner (1988) *Nature*, **332**, 464-468.

Sutherland, G. R. and Richards, R. I. (1994) *American Scientist,* **82,** 157-163.

Tari, L. W. (1995) *Protein-DNA recognition: The importance of DNA and its components*; Ph. D. Thesis, Department of Chemistry, University of Manitoba, Winnipeg.

Tari, L. W. and Secco, A. S. (1995) *Nucleic Acids Res.,* **23,** 2065-2073.

Terwilliger, T. C. and Berendzen, J. (1999) *Acta Cryst.,* **D55,** 849-861.

Thota, N., Li, X. H., Bingman, C. and Sundaralingam, M. (1993) *Acta Cryst.,* **D49,** 282-291.

Tippin, D. B. and Sundaralingam, M. (1997) *J. Mol. Biol.,* **267,** 1171-1185.

Tronrud, D. E., Eyck, L. F. T. and Matthews, B. W. (1987) *Acta Cryst.,* **A43,** 489-501.

Vargason, J. M., Eichman, B. F. and Ho, P. S. (2000) *Nature Structural Biology,* **7,** 758-761.

Vargason, J. M., Henderson, K. and Ho, P. S. (2001) *Proc. Natl. Acad. Sci. USA,* **98,** 7265-7270.

Verdaguer, N., Aymami, J., Fernandez-Forner, D., Fita, I., Coll, M., Huynh-Dinh, T., Igolen, J. and Subirana, J. A. (1991) *J. Mol. Biol.,* **221,** 623-635.

von Hippel, P. H. and Berg, O. G. (1986) *Proc. Natl. Acad. Sci. USA,* **83,** 1608-1612.

von Hippel, P. H. and Berg, O. G. (1989) *J. Biol. Chem.,* **264,** 675-678.

Wahl, M. C. and Sundaralingam, M. (1997) *Biopolymers,* **44,** 45-63.

Wahl, M. C. and Sundaralingam, M. (1999) In *Oxford Handbook of Nucleic Acid Structure* (Ed, Neidle, S.) Oxford University Press, New York, pp. 117-144.

Wang, A. H.-J., Fujii, S., van Boom, J. H., and Rich, A. (1982) *Proc. Natl. Acad. Sci. USA,* **79,** 3968-3972.

Weickert, M. J. and Adhya, S. (1993) *Molecular Microbiology,* **10,** 245-251.

Weston, S. A., Lahrn, A. and Suck, D. (1992) *J. Mol. Biol.*, **226**, 1237-1256.

Wing, R., Drew, H., Takano, T., Broka, C., Tanaka, S., Itakura, K. and Dickerson, R. E. (1980) *Nature,* **287,** 755-758.