Structural Study of the DNA Fragment d(CGCTGGCCACCG) and a Preliminary Crystallographic Study of the Self-Complementary DNA Sequence d(CGGTGGCCACCG)

by

Mei Chen

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

MASTER OF SCIENCE

Department of Chemistry
University of Manitoba
Winnipeg, Manitoba

THE UNIVERSITY OF MANITOBA

FACULTY OF GRADUATE STUDIES

*****

COPYRIGHT PERMISSION PAGE

STRUCTURAL STUDY OF THE DNA FRAGMENT d(CGCTGGCCACCG) AND A PRELIMINARY CRYSTALLOGRAPHIC STUDY OF THE SELF-COMPLEMENTARY DNA SEQUENCE d(CGGTGGCCACCG)

BY

MEI CHEN

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 MASTER OF SCIENCE

Abstract

This thesis is aimed at elucidating the structures of two DNA fragments, d(CGCTGGCCACCG) (referred to as 'mismatched' DNA hereafter) and d(CGGTGGCCACCG). The fragment d(CGCTGGCCACCG), if duplexed, contains two C-C mismatches which are the least efficiently repaired mismatches and are considered a critical structural element in the development of *fragile X syndrome*. The fragment d(CGGTGGCCACCG) which is self-complementary contains the sequence GTG/CAC which frequently occurs in regulatory regions and is related to DNA-protein recognition.

Small crystals of the DNA fragment d(CGCTGGCCACCG) have been obtained. Circular dichroism, UV and gel electrophoresis techniques have been used with the goal of elucidating structural features of the anticipated C-C mismatched DNA in crystallizing solutions. The effect of individual crystallizing agents on the conformation of the 'mismatched' DNA fragment has been studied and discussed in detail. There is evidence that in the crystallizing solutions the 'mismatched' DNA forms a hairpin structure possibly with the first and the last four bases in the stem in B-DNA helical form and the central four bases in the loop which are partially stacked. Under the same crystallizing condition, the self-complementary DNA is duplexed and its helical type is B-form. Based on the crystallizing experiments, structural studies, and an extensive search of the literature, some suggestions regarding additional crystallizing conditions for the 'mismatched' DNA are made.

Two kinds of crystals of the self-complementary DNA sequence d(CGGTGGCCACCG) have been crystallized using $BaCl_2$ and $ZnCl_2$ as salt, respectively. The cell dimensions of crystals from $BaCl_2$ conditions are $a = b = 77$ Å, $c = 40$ Å, $\alpha = \beta = \gamma = 90^{\circ}$.

The unit cell is tetragonal. A complete x-ray diffraction intensity data set has been successfully collected to a resolution of 2.9 Å on a crystal crystallized with $ZnCl_2$. This appears to be the first oligonucleotide crystallized using $ZnCl_2$ as no report of a successful crystallization with Zn^{2+} has been published. The rhombohedral cell expressed in hexagonal form has dimensions of a = b = 47.313 (49) Å, c = 102.098 (63) Å, $\alpha = \beta = 90.0^{o}$, $\gamma = 120.0^{o}$, with a volume of 197,925 (418) $Å^3$. The intensity statistics indicate the Laue class is $R\bar{3}m$.

Acknowledgments

I would like to express my great gratitude to my supervisor, Dr. Anthony Secco, for his invaluable advice, guidance, support and patience throughout my study at the University of Manitoba. His remarkable insight and optimistic disposition have always been a source of inspiration and encouragement in difficult times in my work. I have enjoyed an excellent freedom in my research. This freedom combined with my arduous endeavors made the graduate school experience a very fruitful and unforgettable one.

I would like to express my sincere thanks to Dr. Les Tari for his continuous, selfless assistance and initiative suggestions throughout the whole DNA crystallographic data collection procedure and Dr. Frank Hruska for the fundamental principles of nucleic acids, dedicated discussions and constructive ideas. I extend special thanks to Gillian Henry and Lynda Donald for their many days of generous assistance in CD and gel electrophoresis studies, to Shaheen Shojania and Angela Toms for their many help and demonstrations in study and experiments, and to Dr. Ann McGregor for the always open lab for the UV experiments.

I am certainly greatly indebted to all my family members, especially my husband, Chuheng. Without their so many years of tremendous help and encouragement, I definitely would not have achieved so much in my professional pursuit. My son Jimmy also deserves a great deal of credit for never complaining about the extra long time Mum spent on her work.

I am truly grateful to my supervisor for granting me this opportunity and to the University of Manitoba for the financial assistance of the University of Manitoba Graduate Fellowship.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

General Introduction

"Mistakes during DNA replication can have disastrous effects on an organism, such as genetic diseases or even death. It is only by understanding the nature of these mistakes that we will be able to minimize them."[1]

1.1 DNA Structure: background

Deoxyribonucleic acid (DNA) is the 'molecule of life'. DNA stores and transmits genetic information in living systems. DNA usually has a double stranded structure, or duplex. This duplex consists of two complementary strands interwound in an antiparallel fashion to form a helix. The two-strand helix is held together by hydrogen bonds between complementary heterocyclic bases (Figure 1).

Almost all cells contain a complete copy of the organism's genetic information. For this information to be utilized, it must be copied or transcribed into messenger ribonucleic acid (mRNA). mRNA carries information from DNA to ribosomes where proteins are synthesized. Each time an mRNA molecule completes its journey through a ribosome, a new protein molecule is produced. The central dogma of molecular biology that 'DNA makes RNA makes proteins' was established in 1950's and has greatly accelerated the development of modern molecular biology.

DNA is constructed by four kinds of bases, adenine (A), guanine (G), cytosine (C), and thymine (T). The genetic information within DNA is coded in the form of three letter words called codons. Each codon specifies an amino acid. When a cell divides, its double-stranded DNA splits into two single strands, each of which acts as a template to direct the synthesis of its new complementary strand of DNA. As a result, each new cell has its own complement of duplex DNA. This process is known as replication and it is not perfect. The human genome has 3×10^9 base-pairs and inevitably, a few mistakes are made during copying. A single mistake alters one codon and may lead to an incorrect amino acid. Each time a protein is synthesized under the direction of DNA with a codon error, a mutant rather than a normal protein may be produced. As a consequence the protein may mal-

function resulting in a catastrophic event such as a carcinogenic lesion, a genetic disease, or death of the organism[1, 2].

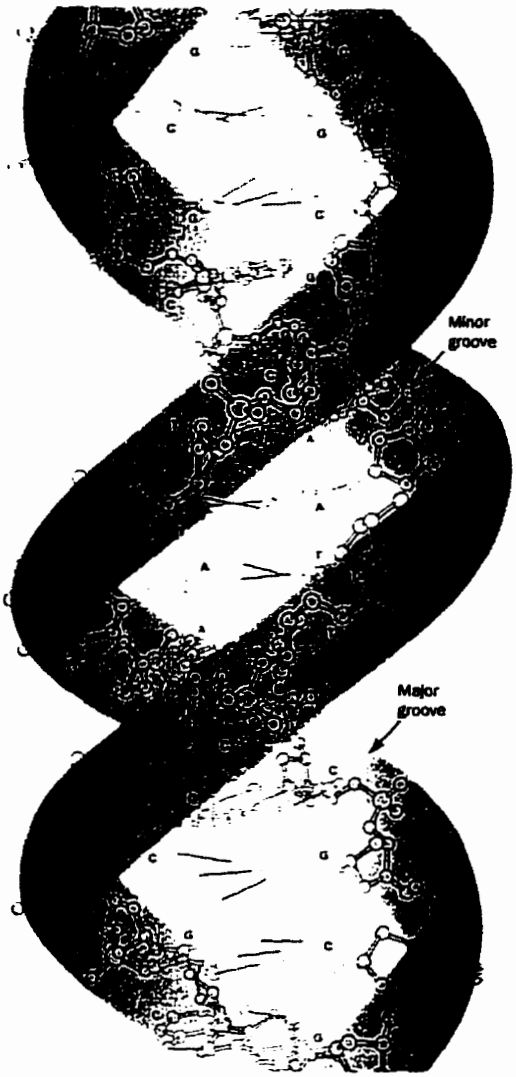


Figure 1. The structure of B-DNA. From Ref. 2.

Because mistakes during replication are usually detrimental, they must be minimized. The formation of so-called Watson-Crick G-C and A-T base-pairs during replication ensures the fidelity of DNA replication, *i.e.* that the information contained in the

parent DNA strand can be copied accurately to the daughter strand, thus preserving the integrity of genetic message from generation to generation. Nevertheless, mistakes are made and mispairing of bases results. There are eight possible base mispairs: A-A, G-G, A-G, C-C, T-T, C-T, A-C, and G-T. Such mispairs arise as a consequence of genetic recombination or as a result of DNA biosynthetic errors or deamination. One of the tasks of DNA repair enzymes is to correct these mistakes. Mismatches are not corrected with equal frequency. These differences may be due to some mismatches being readily recognized by the repair machinery, while others are not. C-C mismatches are the least efficiently repaired and these mismatches are considered critical structural elements in the development of *fragile X syndrome*[3].

Fragile X syndrome is the most common genetic cause of heritable mental retardation in humans. It is so named because of an aberration in a fragile (dynamically mutable) gene on the X chromosome[4]. The mechanism causing this chromosomal abnormality is yet unknown. Current thinking is that during the replication of DNA, one strand slips and makes multiple copies of small segments of the DNA[5]. The repeating segment in the fragile X site is $(CCG)_n/(GGC)_n$ triplet repeats. A biological characteristic related to this is that the $(CCG)_n$-strand shows asymmetric expansion during *in vitro* replication. It is reported[3] that both the C- and G-rich strands can independently form hairpins under physiological conditions. However, the C-rich strand has a much higher tendency to form hairpins than the G-rich strands and therefore is more likely to slip and be preferentially expanded during replication. Under 50 copies of the triplet unit is normal; 230 to several thousand copies are found in patients with disease. The individual strands of $(CCG)_n$ can

form hairpin loops and a double helix with every third base-pair being a C-C mismatch[6] (Figure 2).

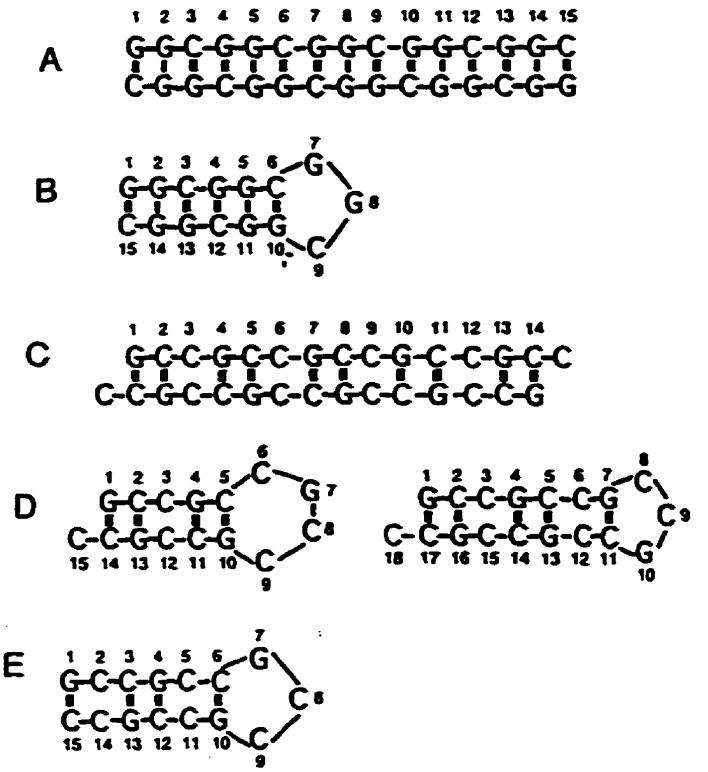


Figure 2. Schematic representation of self-assembled structures of $(GGC)_5$ duplex (A), $(GGC)_5$ hairpin (B), $(GCC)_5$ slipped duplex (C) and $(GCC)_6$ hairpins (D), and blunt $(GCC)_5$ hairpin (E). From Ref. 3.

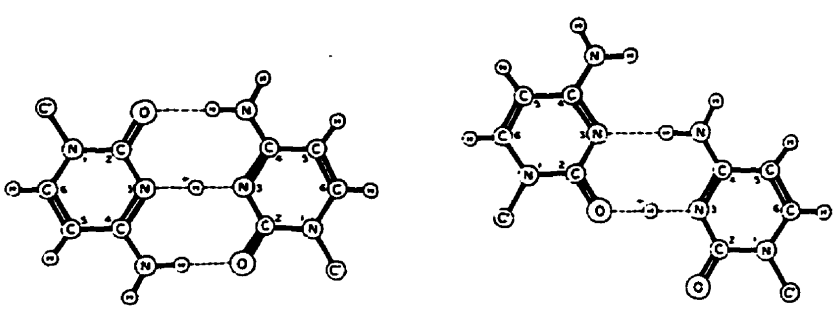
Why are some mismatches less efficiently repaired than others? Several hypotheses have been reported[7, 8], two of which are: (*i*) mismatches destabilize the duplex, result in local structure differences around the mismatch from that for canonical DNA, and therefore induce structural recognition problems for enzymes; (*ii*) the results from studying the structures and repair efficiencies of several mismatches indicate that the more a mismatch mimics a Watson-Crick base-pair, the less efficiently it is repaired. The question

of why a C-C mismatch is the least efficiently repaired of all eight possible mismatches remains unknown.

Determining the structures of DNA sequences with C-C mismatches is important from a chemical point of view, *i.e.* what is the structure of a C-C mismatch, how is it accommodated in duplex DNA and how does it affect interactions with other molecules?

All mismatches studied by crystallography, NMR, or circular dichroism have been found to be base-paired and stacked within the double-helical structure of the DNA fragment investigated. Results from these studies suggest that minimal perturbation of the helix occurs on introduction of a mismatch.

C-C mismatches undergo structural rearrangements as the pH is reduced[9-11]. The C-C pairs in low pH solution involve interaction between a protonated cytosine base and a neutral cytosine base[12], which could be of the types proposed by Langridge and Rich[13] and by Gray and Cui[14] (Figure 3). An additional hydrogen bond is introduced by the N3-protonation of cytosine.



(*a*) (*b*)

Figure 3. Types of C-C$^+$ base-pairs proposed by (*a*) Langridge and Rich and by (*b*) Gray and Cui. The C-C$^+$ base-pair can be accommodated between the antiparallel strands without requiring the unfavorable *syn* conformation in one of the strand. From Ref. 13 and Ref. 14, respectively.

This result is consistent with the work of Bhattacharya and Lilley[15], who found that the reactivity of the C-C mismatch to hydroxylamine is suppressed at low pH, suggesting that a rearrangement of base-pairing occurs on protonation. In fact for all mismatches studied to date, when non-Watson-Crick bases find themselves opposed in duplexes, they stack in the helix much as conventional base-pairs, and are frequently hydrogen bonded[16].

Determining the structures of DNA sequences with C-C mismatches is also important from a biochemical point of view. It is important to learn how C-C mismatches are able to escape recognition by repair systems[16].

If the DNA repair machinery fails to rectify errors in DNA synthesis, the result will be the introduction of a mutation into a daughter molecule, with potentially fatal consequences. The conservation of genetic information requires efficient recognition enzymes and post-replication mismatch repair systems. Correction systems require the means to recognize a mispair. Two possibilities are proposed: direct and indirect sequence readout. Direct readout requires that contacts be made simultaneously between the paired bases, most probably in the major groove, while indirect readout involves recognition of local structure distortion around the mismatch.

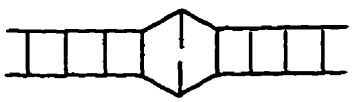
To date over 20 mismatch-containing crystallographic structures have been reported[17]. The small degree of distortion associated with the mismatches implies that recognition of gross structural perturbation will be difficult, and argues for direct recognition of the mispaired bases. In addition to the eight possible mispairings, many observations imply that each mismatched structure is obviously affected by the flanking sequences, making it more complicated for the mismatches to be recognized. The following conclu-

sions have been reached with respect to these problems[18]:

- first, *all* mismatches are associated with minimal, not global structural changes.
- second, despite no noticeable helical distortion around the mismatch, the mispairing of bases can lead to chemical reactivities that are not present in conventional base-pairs; however, the resulting chemical reactivity from mismatches can be dependent on the flanking sequences. Thus, even though some of the mismatched base-pairs contain the potential to be recognized as chemically different, this may be diminished by sequence context.

Werntges *et al.*[18] studied the relationship between repair efficiency and thermodynamic stability of 16 kinds of mispairs. Their results suggest that the repair efficiencies may be thought of in terms of two alternatives. Either the repair enzymes recognize the mismatched bases in their unpaired, *i.e.* locally denatured state, or the mismatch may be recognized for the repair action when it forms a base-pair of non-Watson-Crick type. They found that repair enzymes seemed to check the DNA for rigid local deformations rather than local instabilities of the double strand. Perhaps a mismatch made up of two small pyrimidines (like C-C) is able to escape recognition by the repair system by 'swinging into the helix' as Figure 4 suggests. The mismatches are most likely to occur stack inside the double strand and therefore behave like regular base-pairs.

In summary, all the studies concerning C-C mismatches have suggested that single base mismatches introduced into a DNA molecule create no detectable change in the global conformation of the DNA molecule, but very specific chemical reactions which are not found in Watson-crick base-pairs may be observed in mismatched bases. Even though tremendous efforts have been made in studying the effects of C-C mismatches on biological



wobble type: recognized

open mismatch type: not recognized

Figure 4. Hypothesis to explain the observed correlation between repair efficiency and postulated mismatched structure. From Ref. 18.

functions, many questions are still unanswered. No crystal structures of DNA containing C-C mismatches are available so far. It is important to compare the available results with the individual C-C mismatch structures deduced from crystallographic or NMR investigations. The crystallographic description of a C-C mismatch will provide an incomparably clear view of its structure and assist us in elucidating its biological functions at a molecular level.

1.2 DNA Structure: d(CGCTGGCCACCG) and d(CGGTGGCCACCG)

The sequence d(CGCTGGCCACCG) is pseudo-self-complementary with two C-C mismatches if duplexed, while the sequence d(CGGTGGCCACCG) is self-complementary.

With the development of biotechnological techniques and macromolecular engineering, the elucidation of the three-dimensional structures of proteins, nucleic acids, and macromolecular assemblies has become almost commonplace. The most important technique used for this purpose is single crystal x-ray crystallography. Diffraction-quality crystals are essential for x-ray studies.

The elucidation of structures of nucleic acids remains one of the great challenges in structural biology because single crystals of DNA segments are very hard to obtain. Examples have shown that the success of crystallization depends particularly on the choice of length and sequence of DNA segments. Most DNA oligonucleotide molecules that have been studied so far by crystallography range from 4 to 12 nucleotides long and have self-complementary sequences[19, 20]. This is mainly due to the fact that molecules longer than roughly one helical turn are capable of adopting many conformations, making them difficult to crystallize[21]. A self-complementary sequence is such a sequence that the second half of the base sequence is the complement of the first half. Two such identical molecules can form a duplex by typical-Watson-Crick base-pairing and therefore is usually selected because a crystallizing solution with an exact 1:1 ratio of each strand is readily obtained; such is not the case with non-self-complementary DNA. For a non-self-complementary sequence, the strand in slight excess may act as an impurity which hampers growth of single crystals, especially when the impurity shares a structural resemblance to the molecules being crystallized[22].

The sequence d(CGGTGGCCACCG) contains biologically important trimers —— GTG/CAC.

Nucleic acid molecules are able to adopt a wide variety of conformations, which are dictated by their base sequence and solution conditions[23-27]. The variability and flexibility of DNA facilitate its interaction with other molecules to transfer its genetic message. Stretches of cytosine-rich, self-complementary sequences are frequently located near or within regions of functional and/or regulatory importance[24]. They serve as recognition sequences for restriction enzymes[28]. A clear structural understanding of the DNA fragments at or around regulatory regions is necessary to elucidate the regulatory mechanisms. Therefore, many DNA sequences were designed for structural studies as analogs of the binding sites of regulatory proteins. With the ultimate goal to visualize the structure of C-C mismatches by x-ray crystallographic techniques, a DNA fragment d(CGCTGGCCACCG), containing C-C mismatches if duplexed, has been selected. The almost self-complementary sequence satisfies this requirement with mismatches possible at positions 3 and 10.

The GTG/CAC trimer is a frequently occurring sequence in regulatory regions. This trimer may act as a signal to a searching regulatory protein. In previous work completed in our lab by Tari[16, 29] on the hexamer, CGGTGG/CCACCG (a fragment of *gal* operon), significant structural anomalies from the typical helix conformation were found at the GTG site. The A-T is opened or sheared resulting in the T being pushed into the major groove and A into the minor groove and becomes a recognizable feature. The question remains as to whether the anomaly is inherent to GTG, or the result of some other external effects, such as crystallizing conditions.

One duplexed self-complementary dodecamer d(CGGTGGCCACCG) comprises two copies of the above duplexed hexamer. It has been shown[16] that the hexamer crystallizes with two duplexes related by a two-fold axis (perpendicular to the helix axis) such that one duplex stacks on the other. The difference between two such stacked hexamer molecules and one single dodecamer molecule is the covalent link in the backbone of the dodecamer. The goal of studying d(CGGTGGCCACCG) structure is to check for base-pair opening in a similar, but not identical, environment. If there exists any A-T opening in the dodecamer, it should be seen at both symmetry related sites. The solution of the dodecamer $d(CGGTGGCCACCG)_2$ structure will provide knowledge of the structure in the vicinity of the C-C mismatches in the duplex $d(CGCTGGCCACCG)_2$ and provide an excellent basis for the structural comparison of how C-C mismatches affect helix conformation.

Chapter 2

DNA Dodecamer d(CGCTGGCCACCG): Studies of Structure and Properties

"Completely or partly self-complementary sequences in the DNA genome very often serve as recognition sites for regulatory proteins. Under different conditions, these fragments may exist in double helix or hairpin forms. A good understanding of the formation principles is important to gain further insight into the rules governing the three-dimensional folding of nucleic acids."[25]

2.1 Crystallization

2.1.1 Introduction

Single crystal x-ray crystallography is the most powerful technique for visualizing the three-dimensional structures of biological macromolecules. This method has been widely used in the structural analysis of DNA oligonucleotides because of (*i*) the great advances in synthetic oligonucleotide chemistry and (*ii*) the fact that many of the DNA fragment crystals diffract x-rays to a resolution better than 3 Å.

In the crystallization of biological macromolecules, the quality and quantity of the required material is very important. Difficulties in crystal growth may be linked to the nature or source of the biological material. Proper purification, stabilization, storage, and handling of macromolecules are essential prior to crystallization. Normally, a few milligrams should be available for the initial crystallization trials. Following this, additional material with at least the same quality is needed to improve the quality and size of the crystals. Advances in synthetic oligonucleotide chemistry enable us to obtain pure oligonucleotide molecules of defined sequence in large amount for crystallization experiments.

The success of any structural work of oligonucleotides depends critically on whether diffraction-quality crystals can be obtained. Unfortunately, the crystallization of DNA molecules still remains at the empirical stage.

It is often said that crystal growth of biological macromolecules is more an art than a science. This may sometimes be true from a practical point of view because each crystallizing protocol is unique to a specific macromolecule or a macromolecular fragment. However, it is incorrect in principle. In the past twenty years, many efforts have been made in understanding the physical parameters involved in crystallization. To date, more

than 1,500 macromolecules have been crystallized and their structures solved. The information from these has facilitated the macromolecular crystallization as well as the understanding of the relationships between the structures and functions and therefore has greatly accelerated the advance of biological science.

Macromolecular crystallization, like any crystallization, is a multiparametric process mainly involving three steps: nucleation, growth, and cessation of growth. The significant difference from small molecules is that macromolecules are extremely complex physical-chemical systems whose properties vary as a function of many environmental influences such as temperature, pH, ionic strength, contaminants and solvent composition, and so on. Instead of the precise and reasoned approaches that are commonly applied to scientific problems, for the time being at least, an empirical methodology has to be employed. Macromolecular crystallization is, thus, a matter of searching the ranges of the individual parameters that impact upon crystal formation, finding a set or multiple sets of these factors that yield some kind of crystals, and then optimizing the variable sets to obtain the best possible crystals for x-ray analysis. This is done by conducting a long series, or establishing a vast array of crystallization trials, evaluating the results, and using information obtained to improve matters in successive rounds of trials. Because the number of trials is so large, and their ranges are so broad, intelligence and intuition in designing and evaluating the individual and collective trials become essential.

The conformation of DNA in solution is a function of temperature, pH, salt concentration, and solvent. Some special DNA sequences, such as self-complementary and pseudo-self-complementary sequences, are capable of adopting different conformations under different conditions. Two identical self-complementary molecules can either form a

duplex by typical-Watson-Crick base-pairing, or one molecule (*i.e.* one strand) may twist back on itself and form an intramolecularly base-paired structure resembling a hairpin. Oligonucleotide duplexes crystallize in right-handed A and B, and left-handed Z conformations and short self-complementary oligonucleotides crystallize mainly as normal B-type duplexes. However, only one crystallographic study has been reported on a hairpin structure[30], which suggests that hairpinned DNA molecules are extremely hard to crystallize.

A hairpin structure is composed of a stem of Watson-Crick base-pairs and a loop. Results from some optical and NMR studies[31-35] indicate that hairpin structures are common features not only of the sequences that can afford a perfect Watson-Crick duplex structure, but also of partly self-complementary or mismatched sequences. Under suitable solution conditions, those kinds of sequences may establish an equilibrium between a duplex, a hairpin and random coil conformation (Figure 5).

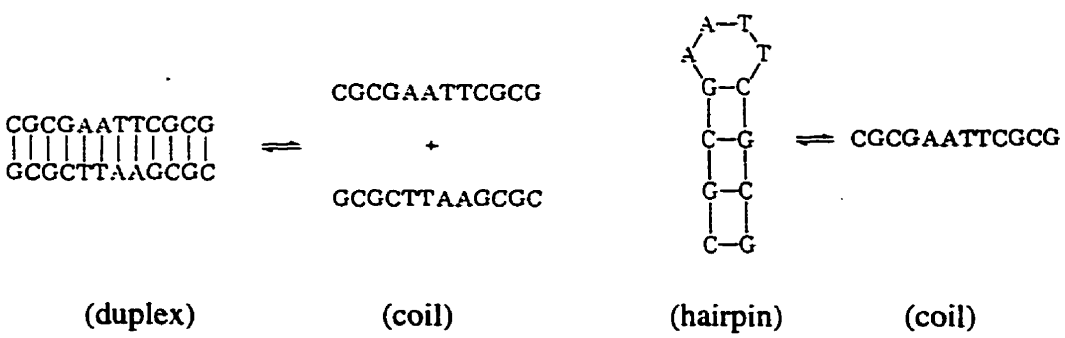


Figure 5. Duplex-coil and hairpin-coil equilibria occurring in d(CGCGAATTCGCG). From Ref. 35.

Hairpins are present as a dominant secondary structure element in RNA. Stable RNA hairpins define nucleation sites for folding, determine tertiary interactions in RNA enzymes, protect mRNAs from degradation, and are recognized by RNA-binding proteins. The biological importance of DNA hairpins is less obvious. Palindrome and self-complementary DNA sequences are frequently found to form hairpin structures at replication origins or operator sequences, which suggests a functional role of hairpin structure in the regulation of gene expression[30, 36, 37].

Because DNA molecules are labile macromolecules, they readily lose their native structures. Hence, the only conditions that can support crystal growth are those that cause little perturbation of the molecular structures. Thus DNA molecules can only be crystallized from a solution within a narrow range of pH, temperature and ionic strength. Crystallization is a process that is not really controlled. Many macromolecules are reluctant to crystallize, especially hydrophobic membrane proteins, nucleic acids and their complexes[38-46].

Since DNA crystals are very sensitive to their environment and the "winning combination" of solution conditions may be rather hard to find, one may identify potentially important factors by setting up a wide selection of screening trials. However, the screening ranges are generally so wide and the parameters affecting the crystallization are so many that the trials are too numerous to be practically performed. Therefore, an efficient, accurate and complete design is essential in solving such a "needle-in-the-haystack" problem. The best experimental method for this purpose is the factorial method. Factorial experiments help provide answers to a problem in a statistical manner with the emphasis on uncovering contributing factors, not on elucidating the underlying causes leading to the

experimental results, *i.e.* what factors are important, not why they are important. The 'contributors' to the crystallization of a DNA molecule are usually a rather small number of specific factors. Generally the factors act independently so that no simultaneous adjustments on two or more factors are necessary[38, 46]. A more detailed account of the factorial method is given in section of 2.1.1.2.

2.1.1.1 Crystallization Factors

The formation of DNA crystals depends critically upon several factors such as the relative concentrations of the chemicals, rate of diffusion, pH value, temperature, and purity of the DNA. Usually, diffraction-quality crystals can only be grown from one set of conditions, not a range of conditions because DNA conformation is very sensitive to the environment. Before starting crystallization experiments, it is essential to know some details of the main factors involved in crystallization.

Salts

Metal ions are required for virtually all biological processes in which nucleic acids are engaged. Positively charged ions are essential for maintaining the neutrality of nucleic acids in solution. Various metal ions, particularly alkaline earth metal ions, divalent metal ions of the transition series, and positively charged organic ions are found to induce or contribute to macromolecular crystallization and therefore are widely used in crystallization experiments[40].

Magnesium ions have profound effects on DNA conformation. In many cases the biological functions of nucleic acids can only be observed in the presence of Mg^{2+} ions.

Most DNA crystals are obtained in the presence of magnesium. It is thought that positively charged magnesium ions bond to the negatively charged phosphates on the backbone to make the helix rigid and facilitate the regular packing of molecules which will lead to crystals[38]. Other alkaline earth metal ions like barium and strontium can be used to replace magnesium and may produce crystals with different lattices[41]. Some other divalent cations have also been tried, such as manganese, calcium, cobalt, nickel, and mercury[39, 47, 48], and as well as complex ions such as cobalt hexammine[49], with limited success in aiding crystallization. Generally, higher magnesium concentrations are used in mismatched DNA crystallization probably because higher ionic strength is required to stabilize the deformation induced by mismatches[50-57].

Na^+ ions are also frequently selected in addition to Mg^{2+} to facilitate DNA crystallization. The facilitating mechanism of Na^+ ions is not very clear. Na^+ ions can increase the ionic strength of the solution as well as stabilize the DNA helices. Such effects can not be obtained by simply adding more Mg^{2+}. Na^+ only, without Mg^{2+}, was used for the crystallization of $d(Br^5CG)_3$ in the Z-form[58]. The structure shows that Na^+ ions stabilize the DNA helices in a different way from Mg^{2+} ions by bridging the N7 of guanine G2 on the neighboring molecules.

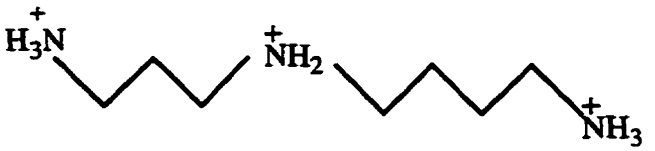
Organic Ions

Organic cations like polyamines are found in all cells. Polyamines may be important in maintaining cellular DNA in a compact state, and facilitating packing of DNA in viruses. Because of their high positive charge, polyamines exert many of their effects

through (*i*) charge-charge interactions, (*ii*) hydrogen bonds, and (*iii*) hydrophobic interactions with negatively charged DNA and RNA[59, 60].

Among the various polyamines found in cells, the most generally used in DNA crystallization are spermine $[NH_3\text{-}(CH_2)_3\text{-}NH_2\text{-}(CH_2)_4\text{-}NH_2\text{-}(CH_2)_3\text{-}NH_3]^{4+}$, an essentially linear molecule with four positive charges at neutral pH and spermidine $[NH_3\text{-}(CH_2)_3\text{-}NH_2\text{-}(CH_2)_4\text{-}NH_3]^{3+}$, an asymmetric molecule with three positive charges at neutral pH (Figure 6).

(*a*)



(*b*)

Figure 6. Structural formulae for (*a*) spermine and (*b*) spermidine.

The two polyamines can interact strongly with nucleic acids due to their highly positive charge. They also have functions similar to the metal ions discussed above in the crystallization of DNA. In many cases crystals can be obtained only in the presence of polyamines, even though very few DNA crystal structures reveal bound spermine molecules in the final crystal structure[29]. Where spermine molecules are found, they twist to

adapt to the surface contour of the helices and form hydrogen bonds with DNA molecules helping to "knit" the molecules together[61].

Precipitating Agents

Precipitating agents in DNA crystallization are usually nonvolatile organic compounds or low molecular weight polymers. The most commonly used agents are 2-methyl-2,4-pentanediol, $CH_3CH(OH)CH_2C(CH_3)_2OH$ (MPD) and low molecular weight polyethylene glycol (*e.g.* PEG 400, $H(OCH_2CH_2)_{400}OH$).

The organic solvent MPD is viscous and is completely miscible with water. In DNA crystallization experiments, a droplet containing crystallizing agents and DNA is sealed against a reservoir with a higher concentration of MPD (see Figure 7 for details). Even though MPD is abundant in the solution, no MPD molecules have been found in any of the crystal structures of nucleic acids[39]. Increasing the MPD concentration in the droplet will (*i*) decrease DNA solubility, (*ii*) increase crystal growing speed if the concentration gradient of MPD between the droplet and the reservoir remains the same, and (*iii*) produce poor quality crystals if the MPD concentration is too high. Spike-like growth can occur from the original crystal surfaces if the growing speed is too high. Sometimes, this may appear weeks after the droplet is set up as the MPD concentration in the droplet becomes high enough (see Chapter 3).

The other common precipitating reagent is PEG which is also completely miscible with water. It is a polymer produced in various lengths, from several to many hundred units. PEG forms a complex network in water which contains both water and itself in the network structure. This reorganization of solvent molecules results in macromolecules

tending to be excluded and leading to phase separation[45, 62]. In addition to PEG's volume exclusion property, PEG probably competes for water and produces dehydration.

Buffer

A buffer is important to DNA because native DNA molecules are only stable in a proper pH range. The buffer selected should not interact strongly with the DNA. The most commonly used buffer is an aqueous solution of sodium cacodylate (dimethyl arsenate) (pK_a = 6.19), which has a favorable buffering strength in the range pH 5 to pH 7[39]. Sodium cacodylate (abbreviated as NaCac) also has an advantage of preventing bacterial growth which may be a problem where PEG is used as precipitating agent.

The pH of a crystallization solution can affect molecular packing. Different crystal forms can be produced depending upon the pH[48, 63]. pH also influences crystal growth kinetics. A slight change in pH, usually 0.5 pH units lower, can accelerate the crystal growth[39].

Cytidines are protonated at low pH. At pH 5 crystallization of the mismatched DNA might be promoted by the introduction of an additional hydrogen bond between every C-C mispair when there is an accessible cytidine in the sequence[38]. Since fifty percent of the bases in the 'mismatched'* DNA, d(CGCTGGCCACCG), are cytidine bases, taking pH into account in crystallization is important.

---

* Referred to as mismatched DNA because of the potential for C-C mismatches in a duplex of two such strands. Strictly speaking, a single strand cannot be mismatched. Therefore quotes are used to indicate the unconventional usage of the term, mismatch.

2.1.1.2 Factorial Method[38, 46]

Experiments specifically designed for screening crystal growth conditions can greatly reduce the number of overall experimental trials and produce adequate results with less effort and sample. This screening can be accomplished by the factorial method[38, 46]. In factorial designs, the crystallizing factors are varied and crystallization results are compared and evaluated, which leads to the uncovering of some effective crystallizing conditions or, at least, a useful database for improving crystallizing conditions.

A complete factorial design starts with formulating an experimental matrix to test the effects of different crystallizing factors at different levels at the same time. The total number of trials involved in a complete factorial design is $m^N \times N$, where m is the number of testing levels for each factor and N is the number of separate factors. For reasons of simplicity and efficiency, the levels are often limited to just two, high (+) and low (-), in which the optimal values are located by two different experiments in each dimension. The total number of trials is then reduced to $2^N \times N$. For example, for four separate factors, the number of trials is $2^4 \times 4 = 64$.

For many more than four separate factors, a full factorial design becomes difficult to carry out because too many trials are required. Selecting subsets of experiments from a full factorial design matrix is essential. This is done by systematically deleting some of the individual experiments, leaving only those necessary to identify main effects. This method of selecting such subsets is called an *incomplete* factorial method.

To further simplify the preliminary test, it was assumed in the design that the interactions between any two factors are very weak, *i.e.* that each factor affects crystallization independently even though the other factors are present. Therefore only main effects, no

interactions, are considered in practice.

Generally, the working matrix for initial crystallization screening by the incomplete factorial method may be designed as follows:

1. A list of possible crystallizing factors that may influence crystal growth is prepared, and from this list a decision on the factors to be screened is made. Normally, five separate factors should be considered at the initial stage: macromolecule, salt, polyamine, and precipitating agent concentrations, and pH. The concentration of the buffer, sodium cacodylate, is kept constant and is not considered in a factorial design.
2. The number of tests is decided. In general, this number should be somewhat more than the number of factors to be screened. There is no rigid rule or short-cut. However, the number of tests should be manageable in terms of carrying out and evaluating the different tests.
3. The experimental matrix itself is designed. The level of each factor is balanced so that there are equal or as nearly equal numbers of tests at each level, *i.e.* equal '+'s and '-'s in each column.
4. The individual tests as specified by the experimental matrix are carried out.
5. The experimental results are examined and evaluated by scoring each test. Observation is very important during crystallization experiments. Often an arbitrary scale such as that in Table 1 is selected and significant and useful information can be obtained by regression analysis. Scoring is difficult to standardize. However, it should be borne in mind that nucleation requires a greater supersaturation than growth, and that crystallization rates increase when supersaturation

increases.

Table 1: Crystal quality scale[38]

Result	Score
Prisms	6.0
Plates	5.0
Needles	4.0
Spherulites	3.0
Gelatinous or participate precipitates	2.0
Cloudy precipitates	1.0

An example of an incomplete factorial method experimental matrix for the crystallization screen of the DNA dodecamer d(CGCTGGCCACCG) is designed as shown in Table 2.

A consideration which deserves special attention while designing the experimental matrix is the relative concentrations of spermine, $MgCl_2$ and DNA when all these are used. For magnesium, a rule of thumb for the first trials is 0.5 - 2.0 Mg^{2+} per phosphate; for spermine, it is one spermine molecule for every 10 - 12 base-pairs. The relative concentrations of spermine to Mg^{2+} also needs to be considered; at high Mg^{2+} concentration, high spermine concentration should be used.

Table 2: Incomplete factorial method experimental matrix for the 'mismatched' DNA d(CGCTGGCCACCG) crystallization screen

trial No.	macromolecule	salt	polyamine	precipitating agent	pH
1	+	+	+	+	+
2	+	+	+	+	-
3	+	+	+	-	+
4	+	+	+	-	-
5	+	-	-	+	+
6	+	-	-	+	-
7	+	-	-	-	+
8	+	-	-	-	-
9	-	+	+	+	+
10	-	+	+	+	-
11	-	+	+	-	+
12	-	+	+	-	-
13	-	-	-	+	+
14	-	-	-	+	-
15	-	-	-	-	+
16	-	-	-	-	-

2.1.1.3 Crystallization Method

There are many methods to crystallize biological macromolecules. Vapor phase equilibrium is the method most favored by crystallographers[38]. Sitting and hanging drop methods are more popular than the other vapor phase equilibrium methods because they are easy to perform, require a small amount of sample, and allow a large amount of flexibility during screening and optimization. The hanging drop method uses less sample than

the sitting drop method and therefore is the method chosen for crystallization of d(CGCTGGCCACCG) and d(CGGTGGCCACCG). In this method, a droplet containing DNA and crystallizing agents which usually include salts, polyamines, precipitating agent, and buffer is sealed against a reservoir containing a solution of precipitating agent at a higher concentration than the droplet. Over time the reservoir will pull water from the droplet via the vapor phase such that an equilibrium will exist between the drop and the reservoir. During this equilibrium process, the dehydration of water in the droplet leads to a shrinkage in droplet volume. Consequently, the concentration of all constituents in the droplet will increase. DNA supersaturation is eventually obtained and DNA precipitate or microcrystals will come out of solution.

2.1.2 Experimental

Materials

The 'mismatched' DNA d(CGCTGGCCACCG) was purchased from the Midland Certified Reagent Company of Midland, Texas. The molecule was synthesized by phosphoramidite chemistry, purified by reverse-phase high performance liquid chromatography, separated by gel filtration, and converted to the ammonium salt.

(1) DNA crystallization working solution

The most common method to determine DNA concentration is to dilute a small amount of DNA with water and measure its UV absorbance at 260 nm. Approximately 7.5 μmoles of the dodecamer was diluted with 100 μL of very pure water (all water used in preparing solutions for use with DNA and its crystallization is ultra-pure from a Barnstead

NANOpure II water purification system) to obtain a concentrated DNA stock solution referred to as "DNA stock 1". Approximately 0.5 μL of "DNA stock 1" was taken and diluted a second time with 1.0 mL of water. The absorbance of this solution is close to 1. The DNA concentration can be calculated from Beer's law $A = \varepsilon Cl$, where A is absorbance, ε is extinction coefficient in units of M^{-1} cm^{-1}, C is concentration in M, and l is the cell length in cm. The extinction coefficients for oligomers can be calculated from the tabulated values of monomer and dimer extinction coefficients with the assumption that absorbance is a nearest neighbor property[64, 65].

Approximate extinction coefficients for single-strand oligonucleotides of base sequence, DpEpFpGp...KpL, may be calculated according to the formula

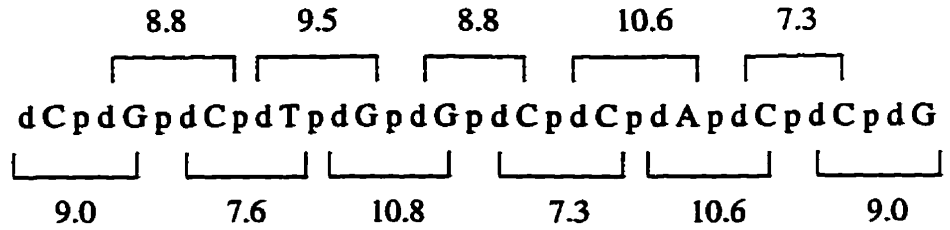
$$\varepsilon_{DpEpFpGp...KpL} = 2(\varepsilon_{DpE} + \varepsilon_{EpF} + \varepsilon_{FpG} + ... + \varepsilon_{KpL}) - \varepsilon_E - \varepsilon_F - \varepsilon_G ... - \varepsilon_K$$

This approximation yields extinction coefficients accurate to about 10% for most single-strand sequences.

The relevant measured and calculated extinction coefficients at 260 nm, 25°C and neutral pH for single-strand DNA (l/mmol cm) are

DNA	pdA	pdC	pdG	pdT	dApdC	dCpdA
ε_{260}	15.4	7.4	11.5	8.7	10.6	10.6
DNA	dCpdC	dCpdG	dCpdT	dGpdC	dGpdG	dTpdG
ε_{260}	7.3	9.0	7.6	8.8	10.8	9.5

For the single-strand DNA d(CGCTGGCCACCG), the molecular weight is 3580.48 g/mol for the polyanion state. The extinction coefficient was calculated as follows:



$$\varepsilon_{260} = 2\,(9.0 + 7.6 + 10.8 + 7.3 + 10.6 + 9.0 + 8.8 + 9.5 + 8.8 + 10.6 + 7.3)$$
$$- 3 \times 11.5 - 5 \times 7.4 - 8.7 - 15.4 = 103 \text{ (l/mmol cm)} = 103{,}000 \text{ (l/mol cm)}$$

With $\varepsilon = 103{,}000\ M^{-1}\ cm^{-1}$, a path length of the cuvette $l = 1$ cm, and an experimental absorbance A = 1.1311, the single-strand concentration of the DNA in the UV solution labelled C_{uv} was

$$C_{uv} = \frac{A}{\varepsilon l} = \frac{1.1311}{103000 \times 1} = 1.098 \times 10^{-5}\,(M)$$

and the concentration of "DNA stock 1" was

$$C = C_{uv} \times \frac{1000.5}{0.5} = 2.20 \times 10^{-2}\,(M) = 22.0\,(mM)$$

It is worth noting that the concentration obtained by this method is the single-strand concentration because DNA is single stranded in pure water and also the single-strand extinction coefficient is used in the calculation. Once the concentration of "DNA stock 1" is known, 1.5 - 5.0 mM (single-strand concentration) DNA working solution "DNA stock 2" for crystallization setups may be accurately prepared.

(2) stock solvent solutions

Throughout this thesis, "stock solvent solution" is defined as the solution containing all crystallization chemicals (salts, polyamines, precipitating agent, buffer) other than

DNA. In order to decrease the concentration errors for chemicals and to ensure that further setups based on the same protocol are reproducible, 50 - 100 μL of stock solvent solution were prepared.

Method

An incomplete factorial method experimental matrix for the crystallization screen was designed exactly as shown in Table 2. The macromolecule was the 'mismatched' DNA dodecamer d(CGCTGGCCACCG). The source of Mg^{2+} was magnesium chloride. With the addition of spermine, less DNA is needed and the crystallizing time is reduced. Therefore spermine was selected as the polyamine in the initial trials. MPD was used as the precipitating agent. NaCac and $Mg(CH_3COO)_2$ were selected as buffers to keep pH values at 6.9 and 5.5, respectively.

Some practical considerations when setting up the 'mismatched' dodecamer were:

(*1*) [DNA] = 0.1 - 1.0 mM (double-stranded concentration) in the droplet.

(*2*) $[Mg^{2+}]$ / [DNA] = 2 - 30

(*3*) [spermine] / [DNA] = 0.2 - 6

(*4*) [NaCac] = 25 mM and $[Mg(CH_2COO)_2]$ = 25 mM

Stock solvent solutions were prepared according to the designs based on the incomplete factorial experimental matrix as shown in Table 2. The stock solvent solutions were stored in the refrigerator. However, any DNA stock solutions were stored in the freezer at -20°C and never kept at room temperature for extended periods.

The procedure for setting up hanging drops on Linbro plates was as follows:

(*1*) A standard 24 well tissue culture plate (Linbro plate, ICN Biomedicals, Inc., CAT No. 76-033-05) was prepared by greasing the rim of the well using a 10 cc syringe filled with high vacuum silicone grease (Dow Corning).

(*2*) The reservoir solution was obtained by diluting the concentrated MPD with water to 50%(v/v). The well was filled with 500μL of this diluted MPD solution as the reservoir.

(*3*) 9 μL stock solvent solution was applied to the center of a clean, siliconized glass cover slip (22 mm square, Hampton Research, CAT No. HR3-215).

(*4*) 1 μL DNA working solution was added and the solution was drawn back and expelled with the pipet several times to ensure that DNA was quantitatively transferred and thoroughly mixed. This step was performed quickly to minimize evaporation.

(*5*) The cover slip was inverted (resulting in the drop being suspended from the underside of the slip) and sealed to the top of the well.

A sample plate with an array of crystallizing droplets is shown in Figure 7.

Usually the total volume of droplets varied from 3-15μL. A smaller volume was hard to pipet quantitatively and larger droplets made it difficult to invert the cover slips without losing the droplets.

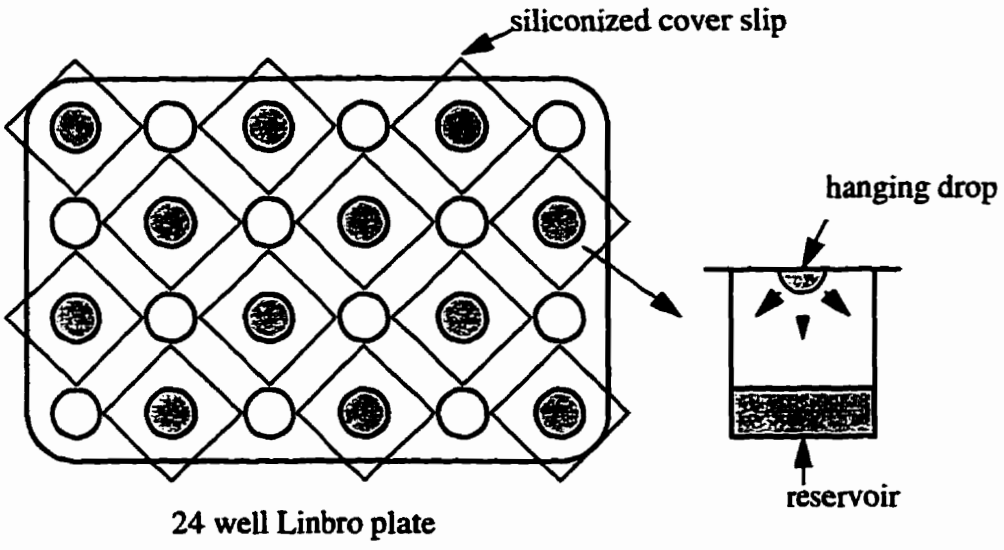


Figure 7. Vapor diffusion method using hanging drops in Linbro plate.

2.1.3 Results

Despite extensive efforts to crystallize this 'mismatched' DNA sequence*, only microcrystals have been observed (Figure 8) under the conditions of:

(*1*) 0.1 mM DNA, 4.0 - 8.0 mM $MgCl_2$, 0.25 - 0.50 mM spermine, 5 - 15% MPD, 25 mM NaCac, pH 6.9, temperature 4°C, and 50%(v/v) MPD in the reservoir.

(2) 0.1 mM DNA, 4.0 - 8.0 mM $MgCl_2$, 0.25 - 0.50 mM spermine, 5 - 15% MPD, 25 mM $Mg(OAc)_2$, pH 5.5, temperature 4°C, and 50%(v/v) MPD in the reservoir.

---

* The crystallizing conditions with $ZnCl_2$ (substituting for $MgCl_2$ as the salt) which were successfully used to crystallize the self-complementary DNA d(CGGTGGCCACCG) (see Chapter 3) were also tried on this 'mismatched' DNA as well as on a 1:1 ratio mismatched: self-complementary dodecamers. No crystals have been obtained thus far.

Figure 8. Small crystals of the 'mismatched' DNA dodecamer.

The crystals in Figure 8 are too small for diffraction analysis. The question arises: why is the 'mismatched' DNA d(CGCTGGCCACCG) so reluctant to form suitable crystals even though numerous conditions have been tested? It even seems to have resisted a marriage with the strand d(CGGTGGCCACCG) where only one mismatch would occur. The answer likely rests on the fact that crystallization of DNA is a rather complicated process. A suitable crystallizing condition is the condition that is able to promote the formation of duplexes from single strands and facilitate the duplexes in packing regularly in three dimensions, *i.e.* to form crystals. The above question prompts the more fundamental query as to whether the 'mismatched' DNA fragment is duplexed, or more directly, what structure it possesses in the crystallizing solutions.

Since DNA molecules can adopt single stranded, hairpinned and duplexed conformations in solutions and hairpinned DNA molecules are extremely hard to crystallize, it is conceivable that under the crystallizing conditions suitable for most DNA fragments, this 'mismatched' DNA predominantly forms hairpins, not duplexes, which may well account for the difficulty in obtaining crystals. Prior to the appearance of the microcrystals shown in Figure 8, it was decided that ultraviolet and circular dichroism spectroscopies, and gel electrophoresis experiments would be used to study the conformation of the 'mismatched' DNA in solution. It was anticipated that a better understanding about the conformation of the 'mismatched' DNA in the crystallizing solution would provide some insight and constructive suggestions on the feasibility of further crystallization experiments.

2.2 UV Study

2.2.1 Introduction

All molecules absorb light at some wavelength. However, for any particular wavelength, certain types of chemical groups usually dominate the observed spectrum. The groups are called chromophores. Typical chromophores found in nucleic acids absorb light only at wavelengths less than 300 nm, *i.e.* in the near ultraviolet range. The strong UV absorption of nucleic acids resides almost exclusively in the purine and pyrimidine bases. The sugar phosphate backbone is insignificant in the UV absorption at wavelengths less than 200 nm[66].

Nucleotide chromophores

All four bases in DNA, A, G, C, and T are heterocyclic chromophores involving π-electrons and are planar. The isolated chromophores are rarely optically active. Chromophores are made optically active by their asymmetric environment. Optical activity of nucleic acids arises from two sources: (*i*) the presence of the chiral sugar at C1' which destroys the mirror plane found in the isolated base and (*ii*) the chirality arising from the secondary and tertiary structures. The bases in native DNA are stacked in such a way that the strong interactions of their π electrons suppress the absorption of photons.

Base stacking and hydrogen bonding are the two main double helix stabilizing forces. At high temperature, the hydrogen bonds and base stacking in double-stranded DNA are disrupted, resulting in two single-stranded random coils. This denaturing process is referred to as DNA melting. As the molecules are denatured, the bases become unstacked with a concomitant increase of absorption, called the hyperchromic effect. The hypochromic effect is common to systems with stacked chromophores.

The melting of DNA is either a cooperative "unzippering" procedure, *i.e.* the collapse of one part of the structure destabilizes the remainder[1, 67], or a non-cooperative procedure in which the breaking of one base-pair only affects the stacking of the nearest neighboring bases[68-72]. Generally, a sharp increase in absorbance in a few degrees results from the cooperative nature of the transition, whereas, a slow increase in absorbance in a wide temperature range demonstrates noncooperative melting.

Thermal denaturation studies on DNA molecules can be performed using a technique called ultraviolet melting measurement which is simple, sensitive and reasonably accurate. In this experiment the DNA duplex is dissolved in an aqueous buffer and heated slowly. The absorbance at a given λ is monitored and plotted as function of temperature. The temperature corresponding to the inflection point of the curve is characteristic of each species of DNA and is defined as the melting point T_m. The value of T_m reflects the stability of the DNA[23, 73]. The absorbance differences between the low and high temperature result from (*i*) the change due to the breakage of the hydrogen bonds between the Watson-Crick base-pairs, (*ii*) the change due to unstacking of the single strands with temperature, and (*iii*) the change accompanying the unstacking of the loop residues if the DNA molecules form hairpin structures[74]. The enthalpy of the overall transition is calculated as the summation of all three contributions.

Generally, the melting curves for most oligonucleotides are sigmoidal and monophasic, *i.e.* there is only one transition in the whole melting procedure. The transition could be (*i*) the denaturation of "single-strand helix"[75-77] (if the single strand is the only species in the solution), which is characterized by a very low 'melting' temperature, broad "melting" temperature range, and very low hyperchromicity (roughly six to ten percent for

oligomers); (*ii*) the transition from hairpin structure to random coil, which shows a sharper transition, higher melting point, and higher hyperchromicity (approximately ten to fifteen percent for oligomers); (*iii*) the transition from duplex to random coil, which is distinguished from a hairpin to coil transition if the oligomer is able to adopt both hairpin and duplex structures in much sharper transiting, higher melting point, and higher hyperchromicity (ranging from fifteen to forty percent depending on the length of the oligomer).

2.2.2 Experimental

Materials

Two DNA dodecamers, 'mismatched' d(CGCTGGCCACCG) and self-complementary d(CGGTGGCCACCG), were studied by the UV experiments. In addition to pure water as solvent, four aqueous buffers were prepared:

(*1*) water

(*2*) $MgCl_2$ (16 mM), and NaCac (25 mM), pH 6.9

(*3*) $MgCl_2$ (16 mM), NaCac (25 mM), and spermine (1.0 mM), pH 6.9

(*4*) $MgCl_2$ (16 mM), NaCac (25 mM), and MPD (2%), pH 6.9

(*5*) $MgCl_2$ (16 mM), NaCac (25 mM), spermine (1.0 mM), and MPD (2%), pH 6.9

10 μL 1.5 mM (single-strand concentration) DNA was added into each above solution respectively, with total volume of 1.2 mL.

Methods

Absorption spectra were collected at 300 nm - 220 nm at different temperatures using a Shimazu UV-2101PC scanning spectrophotometer. The spectrophotometer was

equipped with a thermostated cell holder and the temperature was controlled by a HAAKE G water bath and HAAKE D1 pump. The scan speed was 'slow' with a slit width of 0.5 nm and sampling interval of 0.2 nm. Each scan was performed after an incubation of 10 minutes to ensure that the solution was equilibrated at the desired temperature. The absorbance was found to be constant after 9 minutes. Reference solutions were either buffer or water depending on the experiment.

2.2.3 Results

Treatment of melting data

(1) Absorbance normalization

The UV absorbance at 260 nm was recorded at each experimental temperature. The normalized absorbance was obtained by dividing every experimental absorbance by the absorbance at 100°C, a value obtained by extrapolation of the experimental absorbances at lower temperatures. The melting curves were obtained by plotting the normalized absorbances against the temperatures and these curves are illustrated in Figures 9 to 16.

(2) Thermodynamic parameters determination

The melting curves for the two DNA molecules in water, Figures 9 and 14, are characteristic of a single-strand helix denaturation. The reaction can be expressed as:

$$\text{CGCTGGCCACCG} \leftrightarrow \text{CGCTGGCCACCG}$$

and $$\text{CGGTGGCCACCG} \leftrightarrow \text{CGGTGGCCACCG}$$

(coil) (single-strand helix)

The melting curves for the 'mismatched' DNA in the four buffers, Figures 10 to 13, exhibit the features for the transition from hairpin to random coil. Denaturation of a hairpin structure involves the breakage of the hydrogen bonds on the hairpin stem and the unstacking of bases in each strand to form a random coil. This can be written as

$$\underset{\text{(coil)}}{\text{CGCTGGCCACCG}} \leftrightarrow \underset{\text{(hairpin)}}{\text{CGCTGGCCACCG}}$$

For simplicity, it was assumed that the thermal denaturing of DNA oligomers was a two-state equilibrium, *i.e.* the reaction took place in one step. Each temperature point represented equilibrium between duplexes and random coils or between hairpins and random coils. At 0°C, the oligomer was assumed to be entirely in duplex or hairpin conformation; at 100°C, it was assumed to be random coils. Both "single-strand helix" to coil and hairpin to coil transitions are intramolecular first-order reactions with an equilibrium constant K given by

$$K = \frac{f}{1-f} = \exp\left(-\frac{\Delta H}{RT} + \frac{\Delta S}{R}\right)$$

or

$$\ln K = -\frac{\Delta H}{RT} + \frac{\Delta S}{R}$$

where f is the hairpin or single-strand helix fraction.

The thermodynamic parameters which characterize the reaction are T_m, the temperature where $f = 0.5$ ($K = 1$); ΔH, the enthalpy change of the reaction; ΔS, the entropy change of the reaction. These parameters are obtained from a van't Hoff plot. A van't Hoff plot is constructed by taking the logarithm of the equilibrium constant at each point and plotting it against the inverse of the temperature, *i.e.* $\ln K$ *vs.* $\frac{1}{T}$. A least-squares line is cal-

culated to fit the experimental points. The values of ΔH and ΔS are related to the slope and intercept of the plot according to

$$slope = -\frac{\Delta H}{R} \quad \text{and} \quad intercept = \frac{\Delta S}{R}$$

At T_m, the free energy ΔG equals 0, *i.e.* $\Delta G = \Delta H - T_m \Delta S = 0$. Therefore, the melting point T_m is given by

$$T_m = \frac{\Delta H}{\Delta S}$$

The melting curve of self-complementary DNA in buffer, Figure 15, manifests the typical transition from duplex to random coil. This reaction is an intermolecular second-order reaction which can be expressed as

$$2(\text{CGGTGGCCACCG}) \leftrightarrow (\text{CGGTGGCCACCG})_2$$

(single strand) (double strand)

The equilibrium constant, K, is given by

$$K = \frac{f}{2c(1-f)^2} = \exp\left(-\frac{\Delta H}{RT} + \frac{\Delta S}{R}\right)$$

or

$$\ln K = -\frac{\Delta H}{RT} + \frac{\Delta S}{R}$$

where c is the total strand concentration and f is the fraction of strands in the double helix state.

The experimental data were processed by Mathcad 6.0[78]. The data processing details are presented in Appendix A. The UV melting curves and the corresponding van't Hoff plots for 'mismatched' DNA are shown in Figures 9 to 13, and for self-complementary DNA are shown in Figures 14 and 15. Figure 16 plots all the melting curves on one graph. The thermodynamic parameters are tabulated in Table 3.

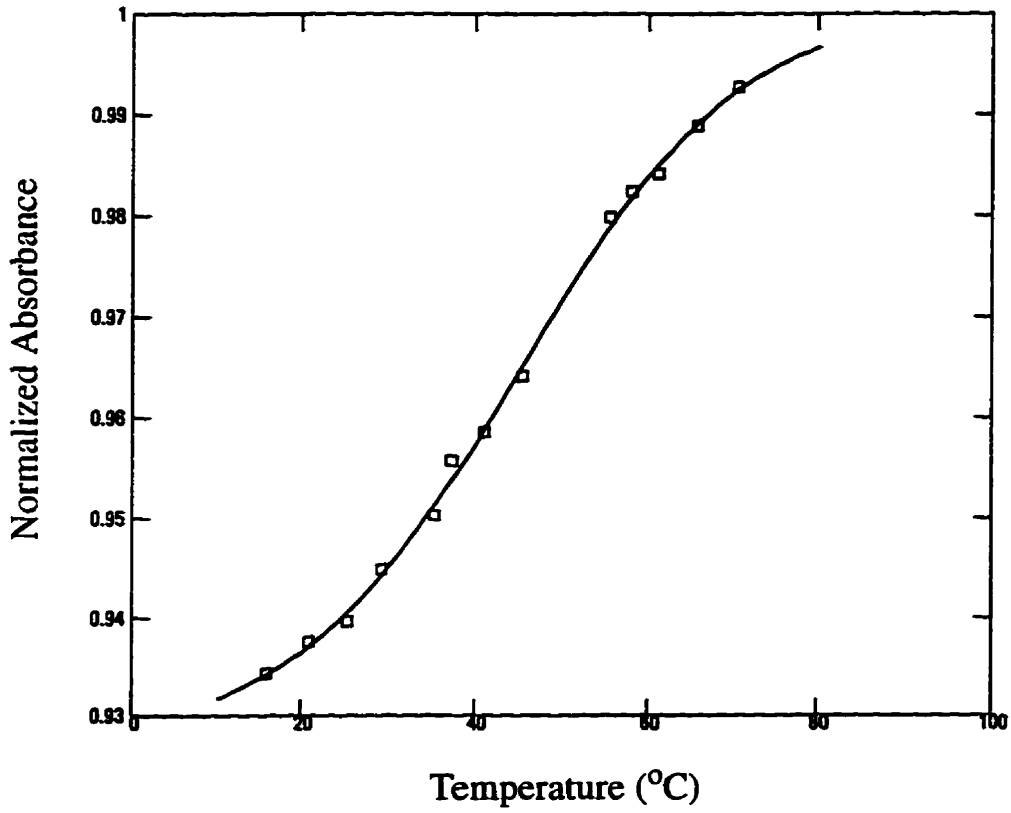


Figure 9. (*a*) Melting curve of 'mismatched' DNA in water.

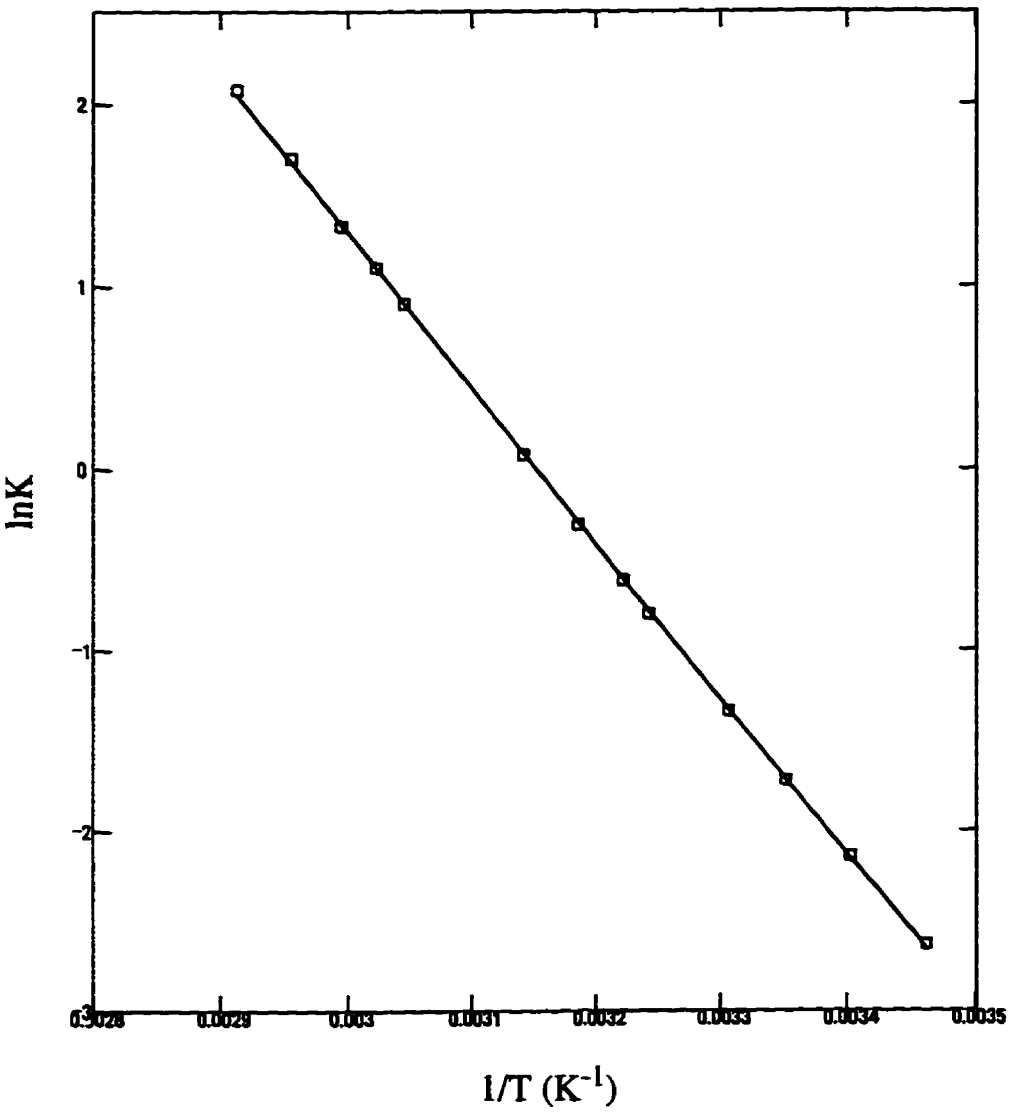


Figure 9. (*b*) van't Hoff plot of the thermal denaturation of 'mismatched' DNA in water.

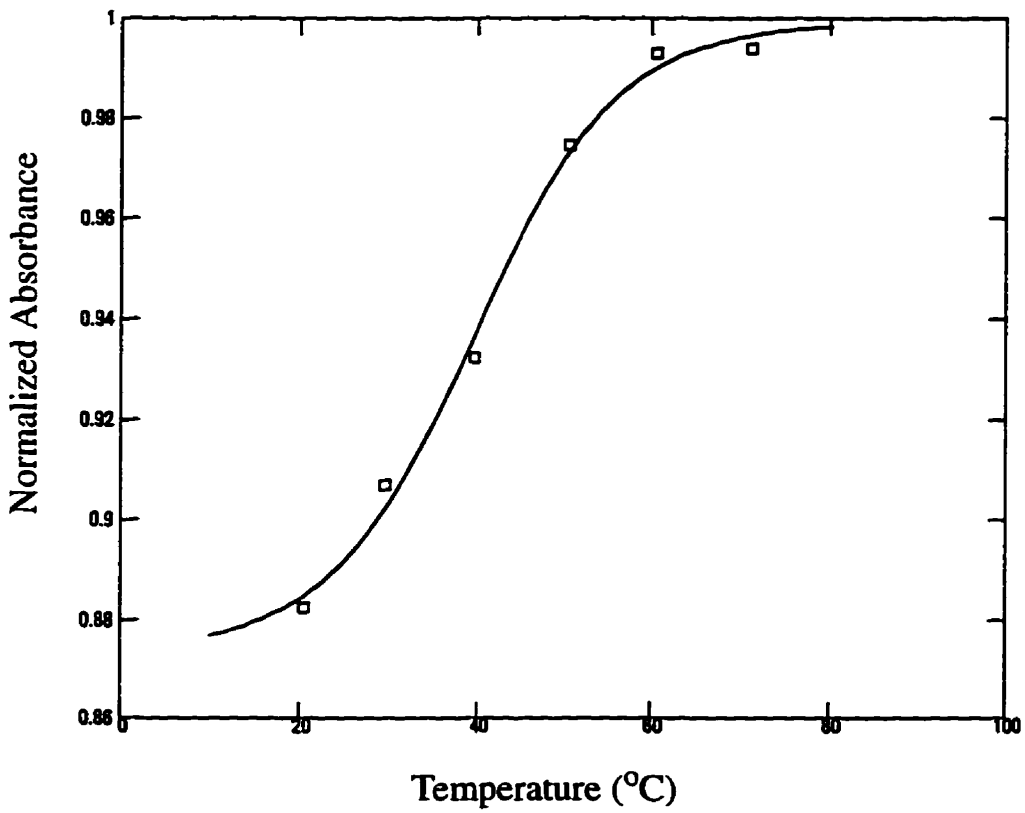


Figure 10. (*a*) Melting curve of 'mismatched' DNA in $MgCl_2$ and NaCac solution.

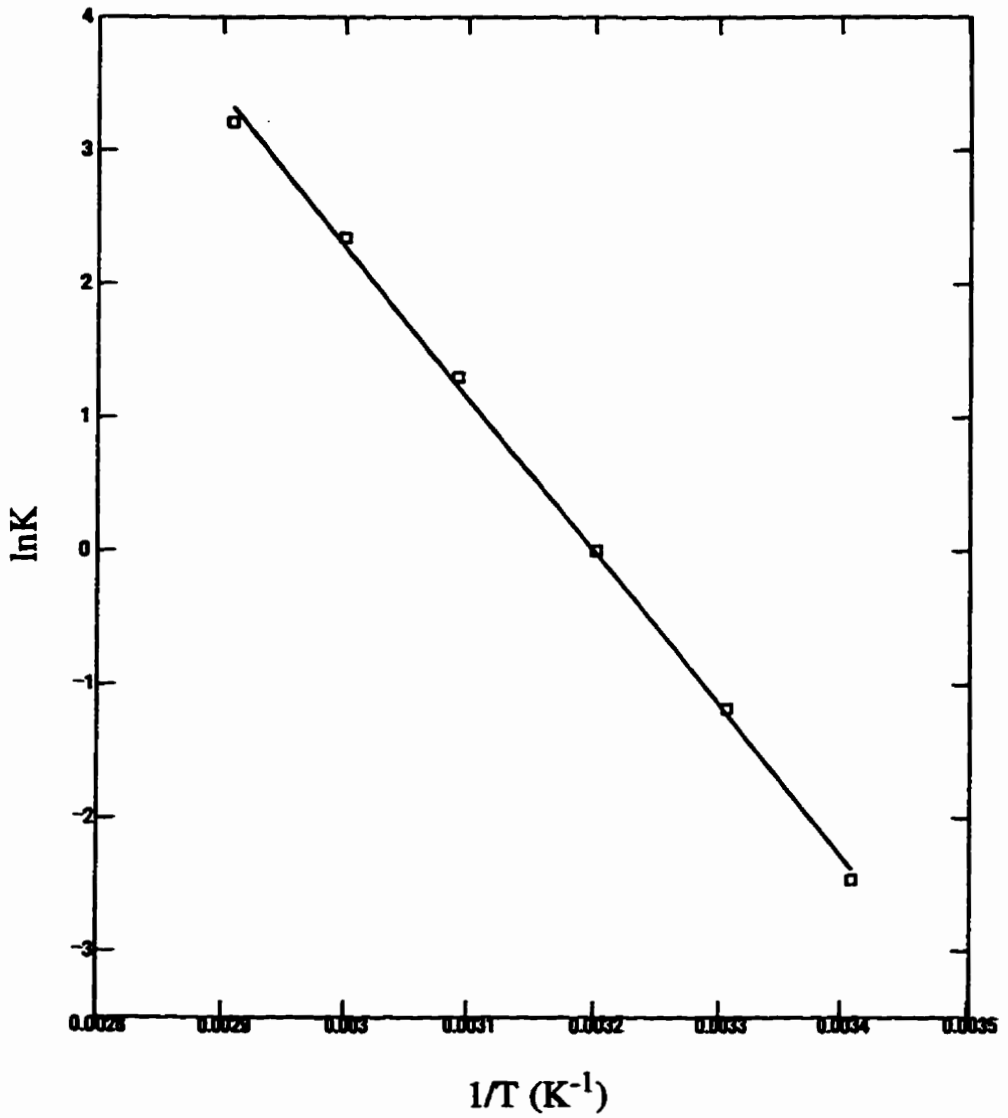


Figure 10. (*b*) van't Hoff plot of the thermal denaturation of 'mismatched' DNA in $MgCl_2$ and NaCac solution.

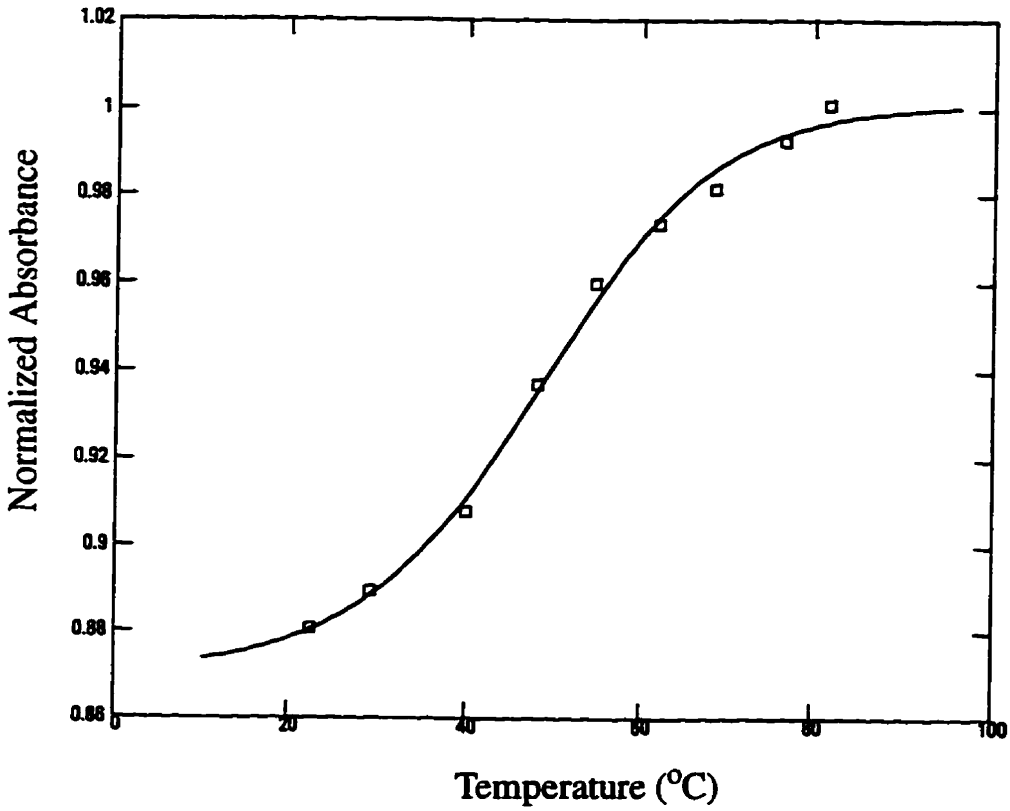


Figure 11. (*a*) Melting curve of 'mismatched' DNA in $MgCl_2$, NaCac, and spermine solution.

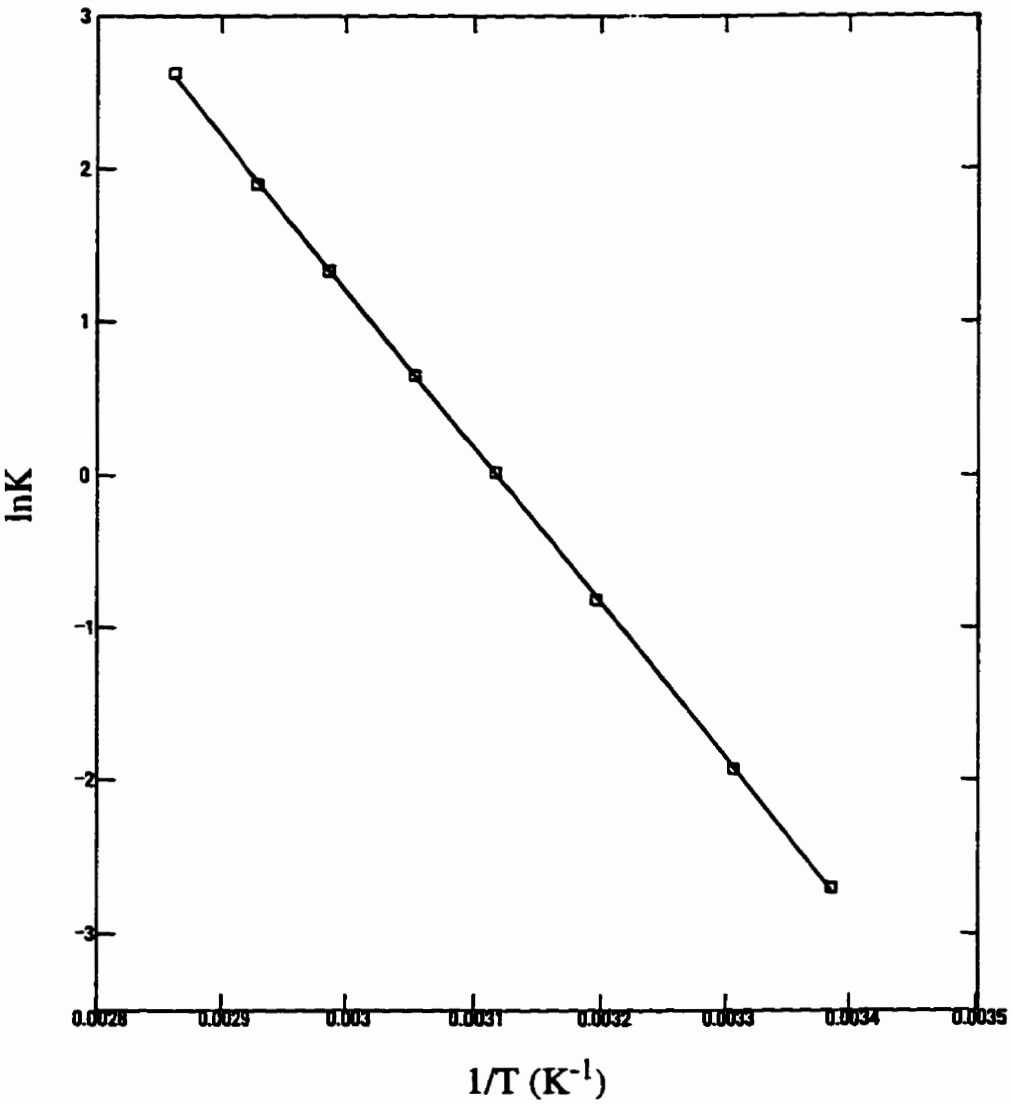


Figure 11. (*b*) van't Hoff plot of the thermal denaturation of 'mismatched' DNA in $MgCl_2$, NaCac, and spermine solution.

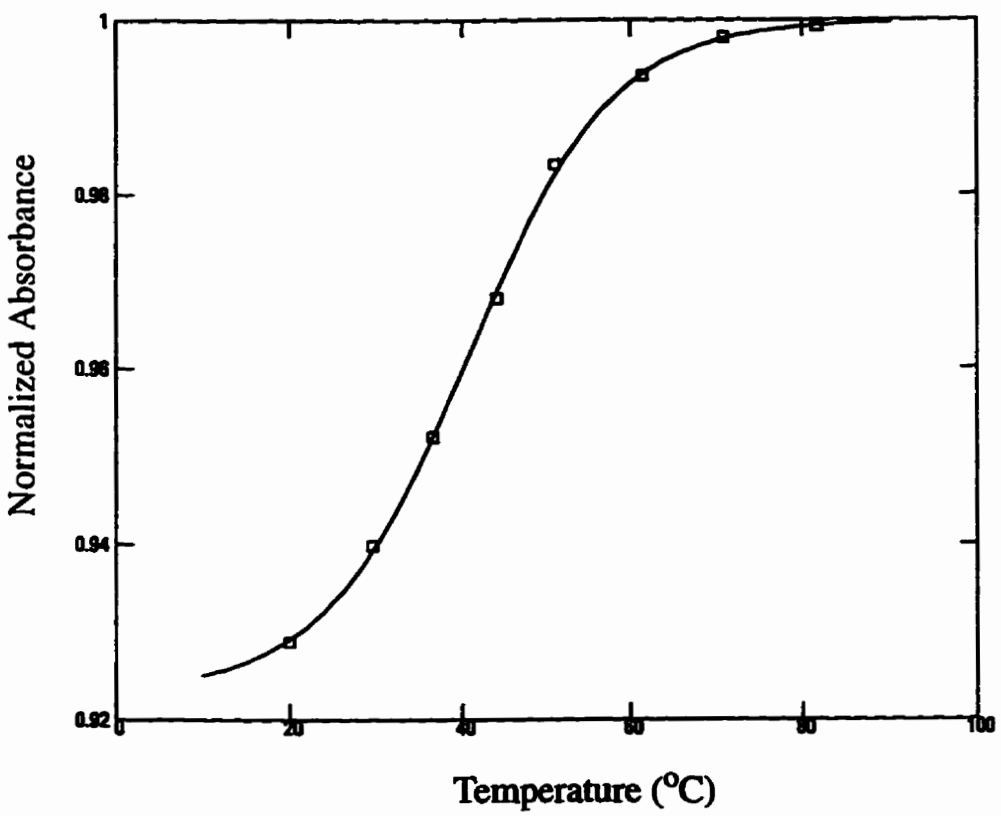


Figure 12. (*a*) Melting curve of 'mismatched' DNA in $MgCl_2$, NaCac, and MPD solution.

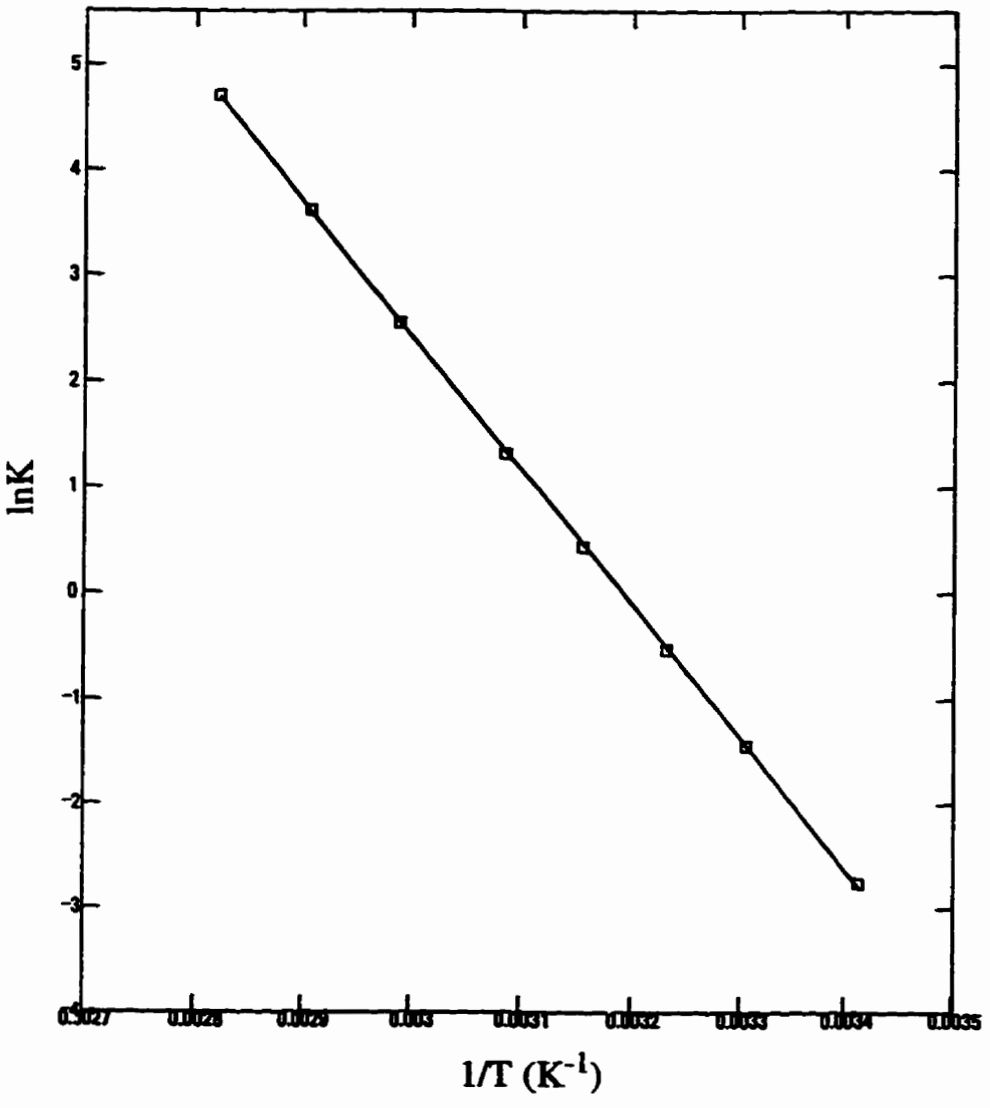


Figure 12. (*b*) van't Hoff plot of the thermal denaturation of 'mismatched' DNA in $MgCl_2$, NaCac, and MPD solution.

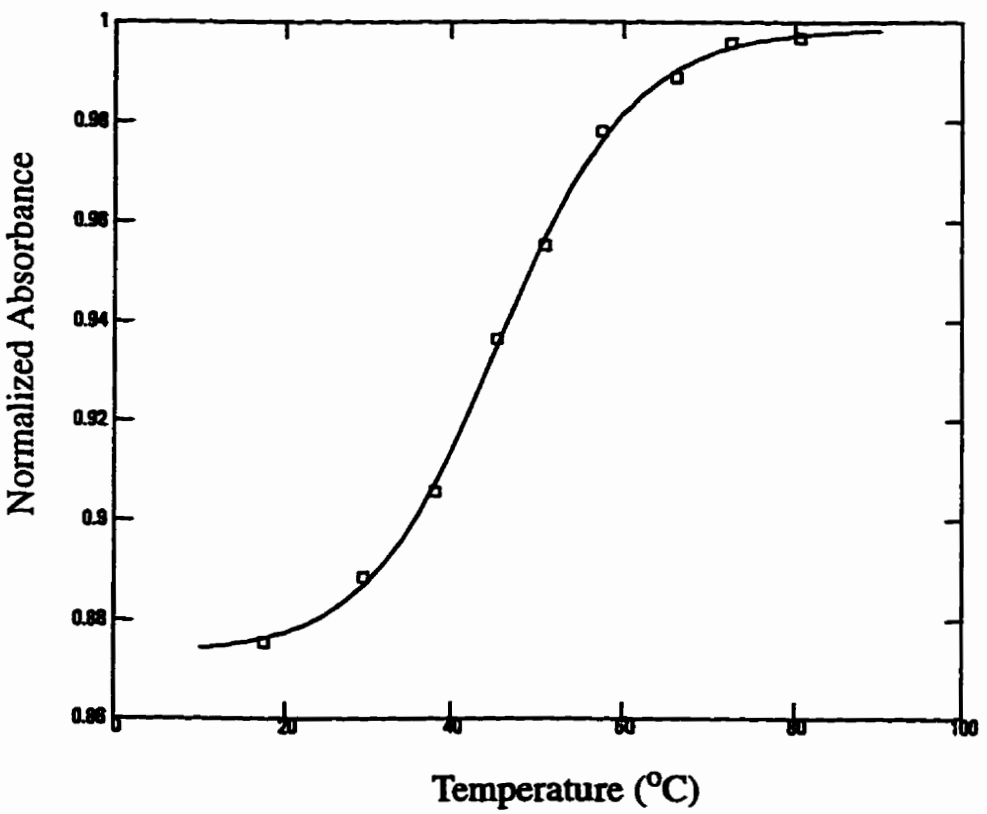


Figure 13. (*a*) Melting curve of 'mismatched' DNA in $MgCl_2$, NaCac, spermine, and MPD solution.

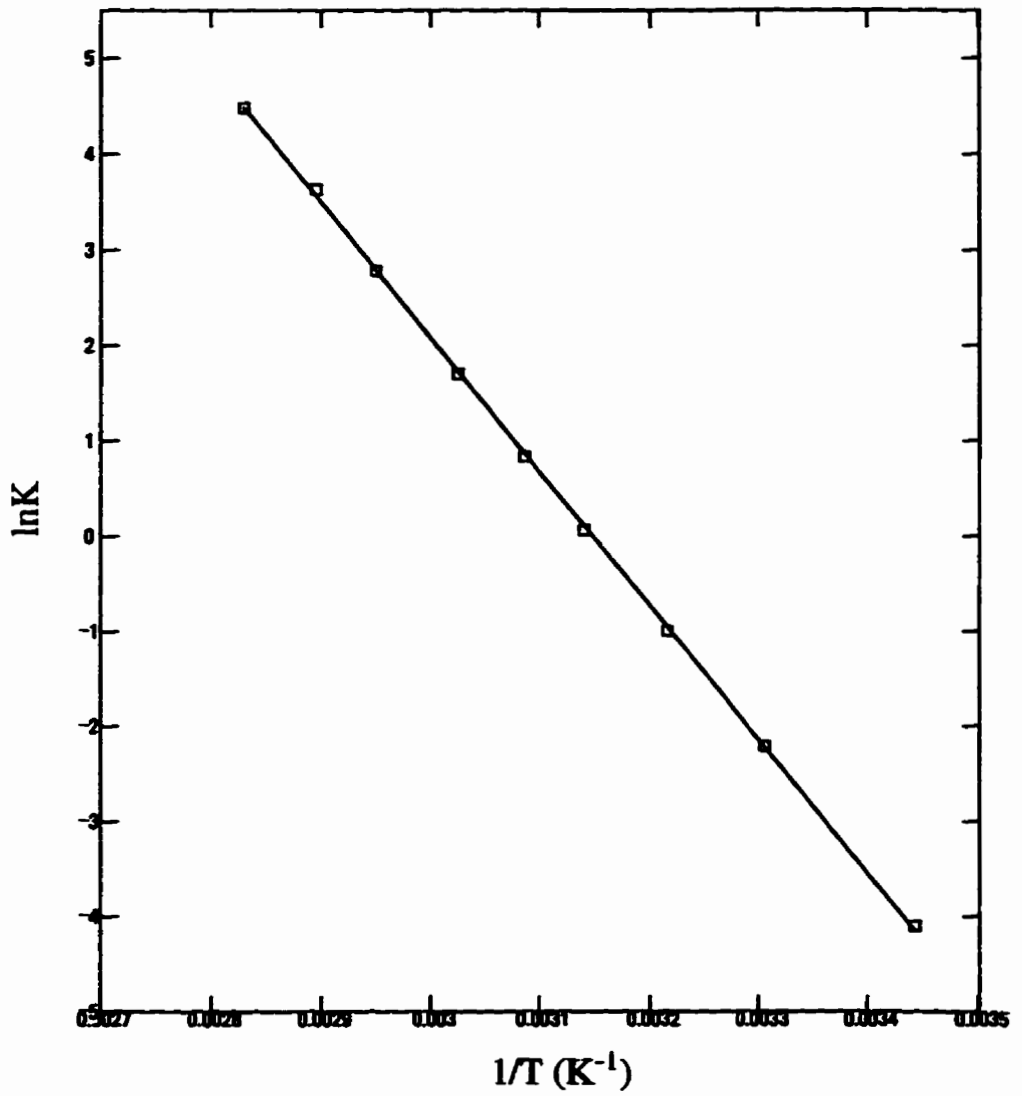


Figure 13. (*b*) van't Hoff plot of the thermal denaturation of 'mismatched' DNA in $MgCl_2$, NaCac, spermine, and MPD solution.

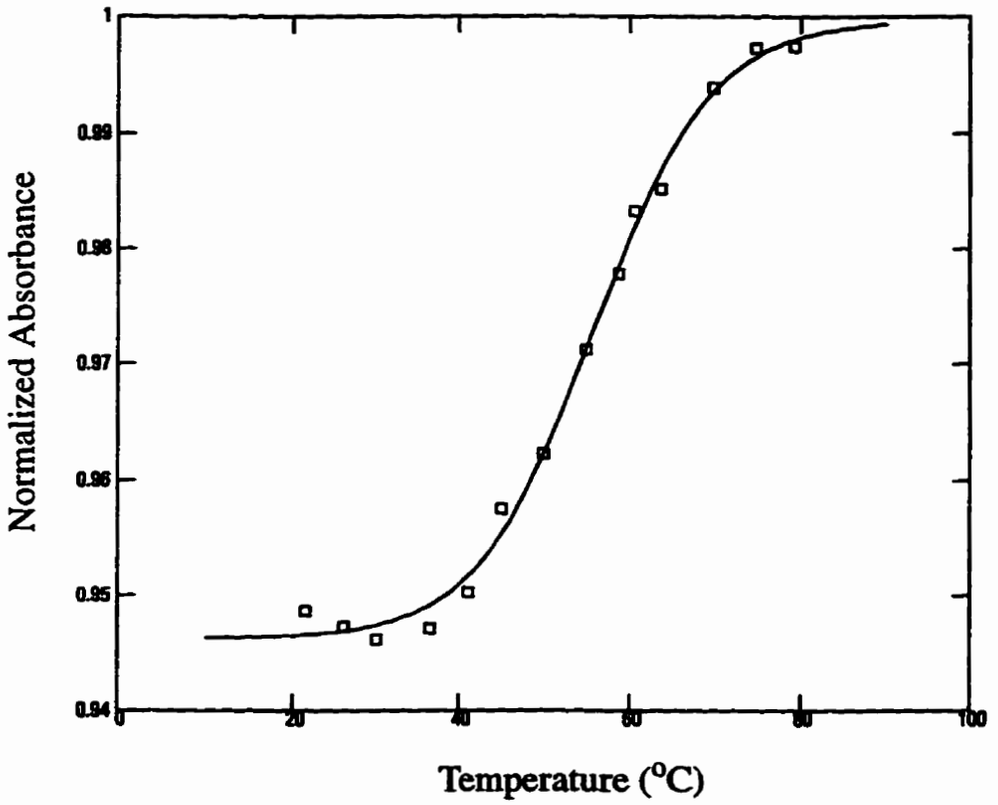


Figure 14. (*a*) Melting curve of self-complementary DNA in water.

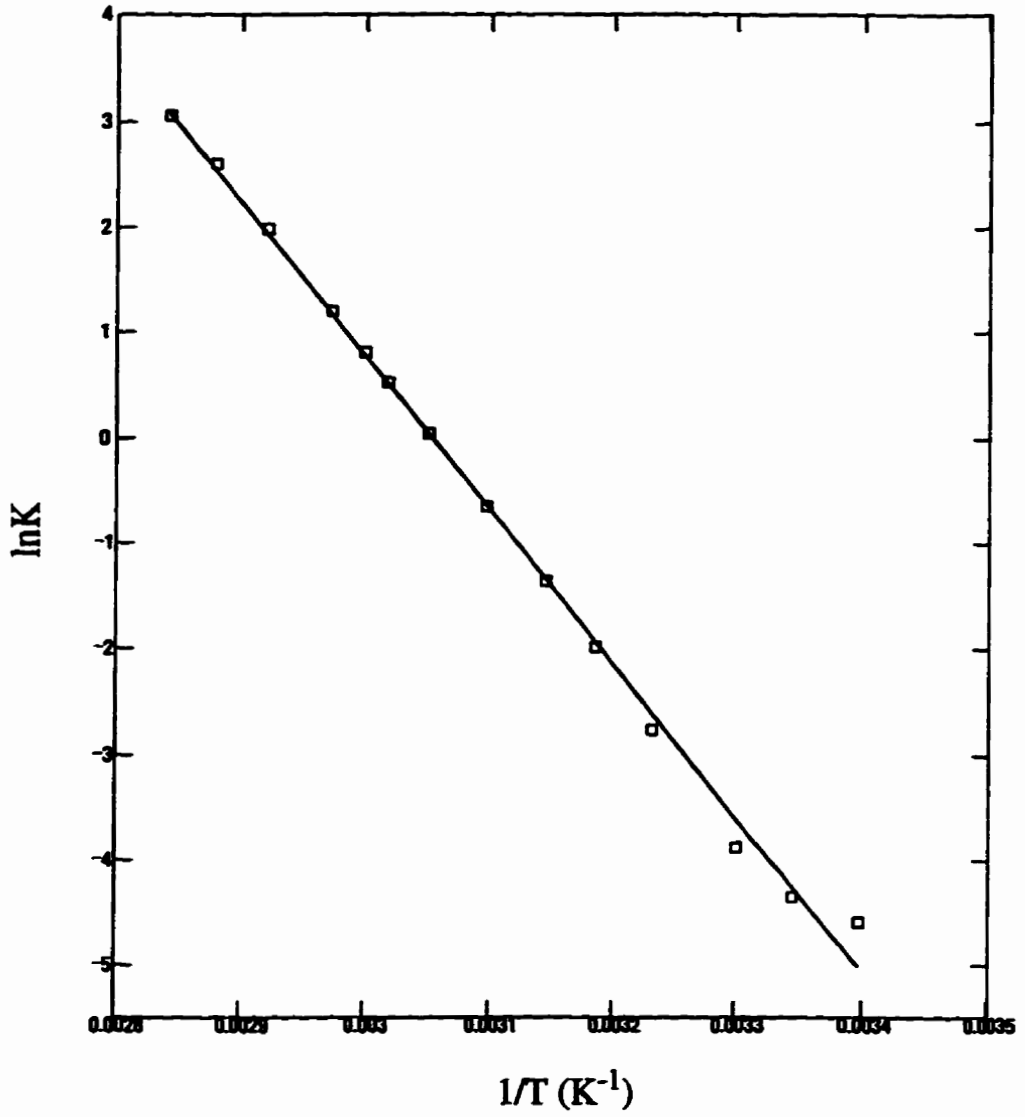


Figure 14. (*b*) van't Hoff plot of the thermal denaturation of self-complementary DNA in water.

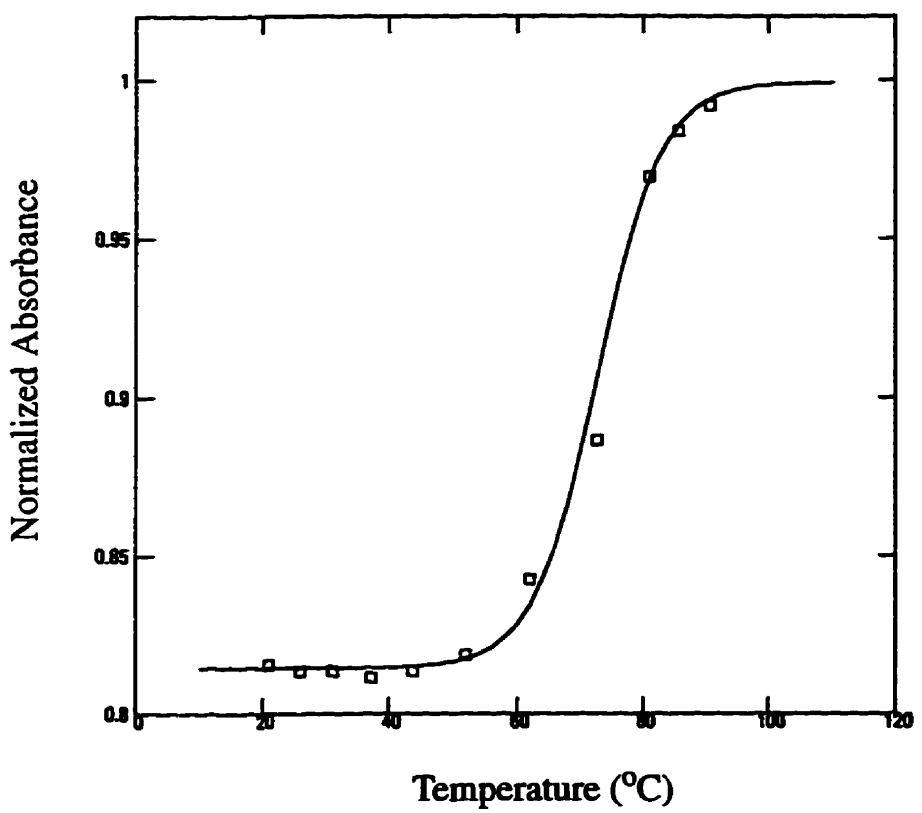


Figure 15. (*a*) Melting curve of self-complementary DNA in $MgCl_2$, NaCac, spermine, and MPD solution.

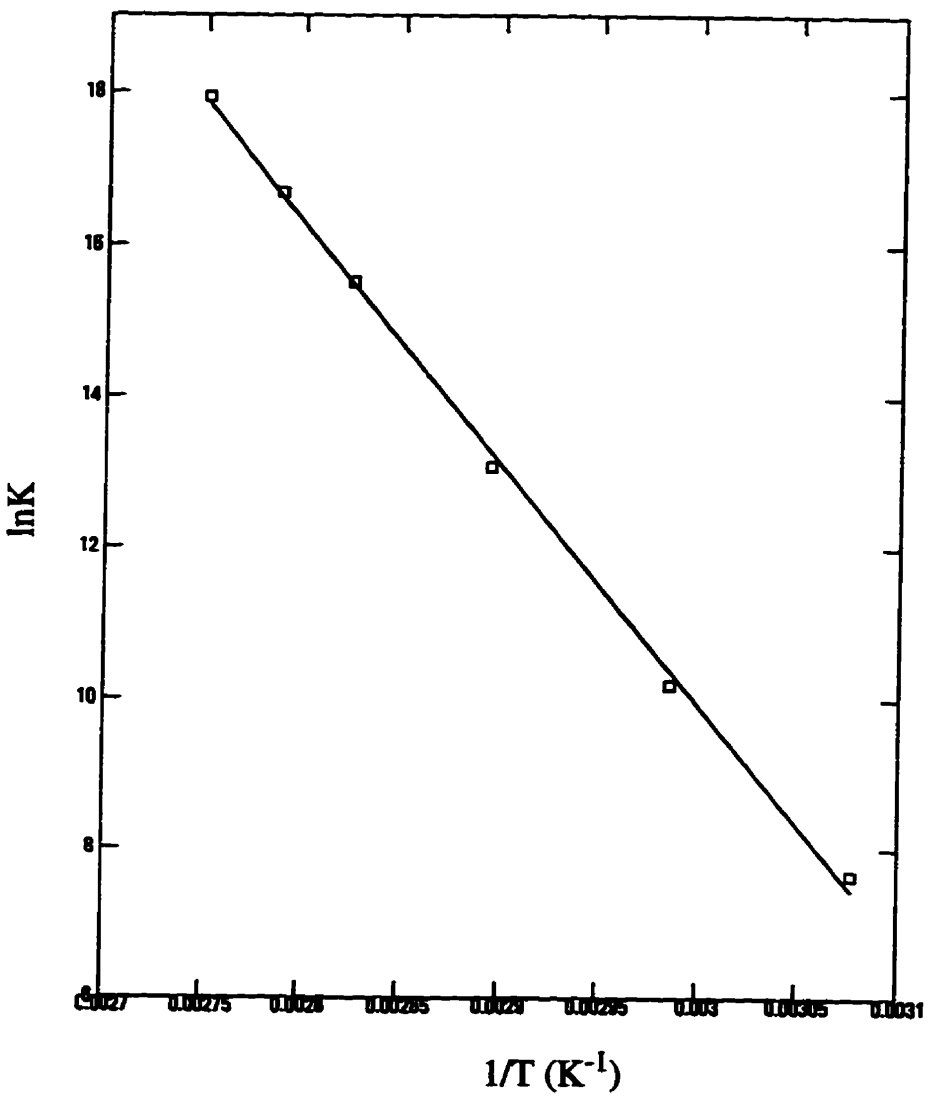


Figure 15. (*b*) van't Hoff plot of the thermal denaturation of self-complementary DNA in $MgCl_2$, NaCac, spermine, and MPD solution.

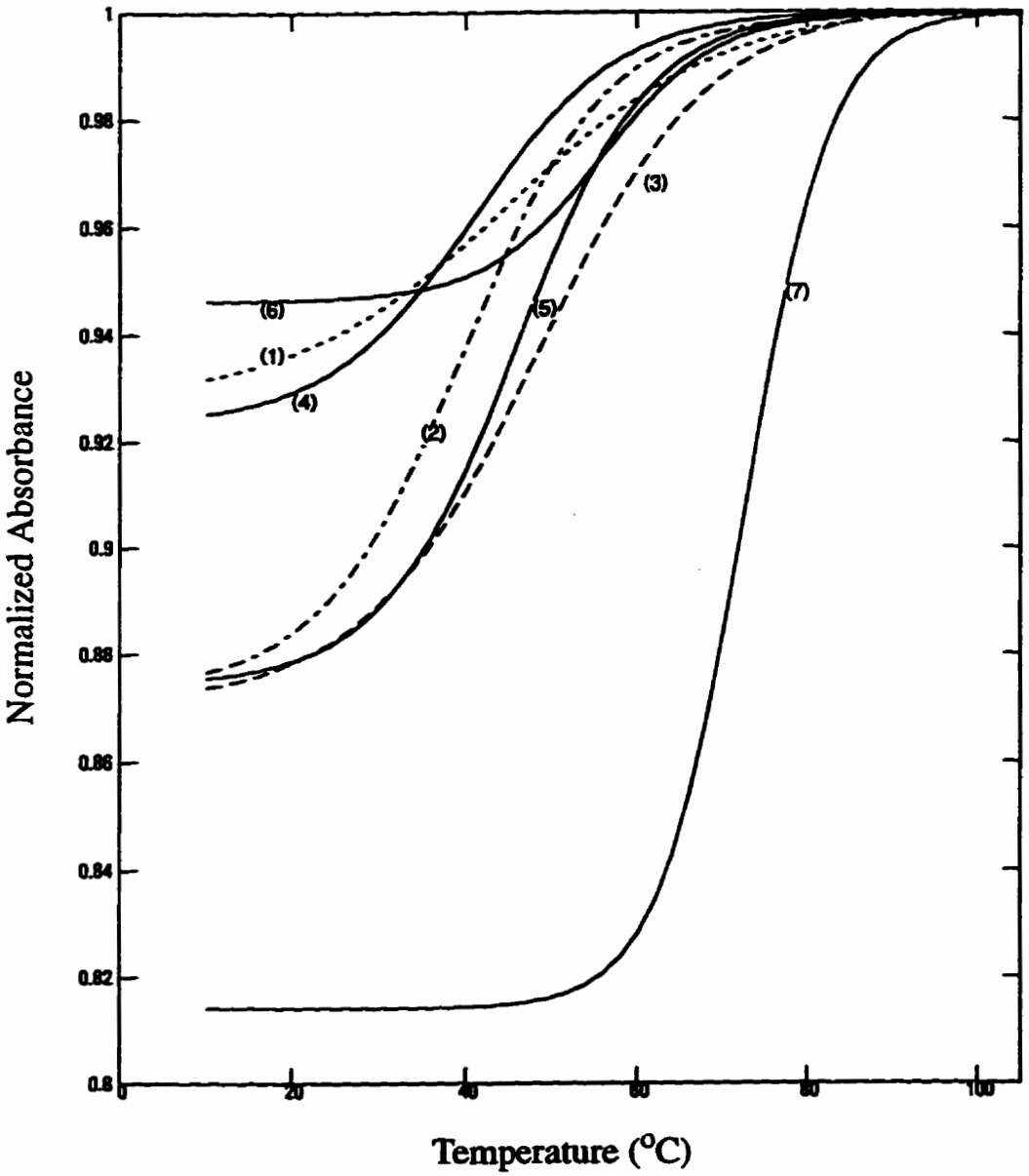


Figure 16. Melting curves of 'mismatched' DNA dodecamer in solutions: (1) H_2O; (2) $MgCl_2$ and NaCac solution; (3) $MgCl_2$, NaCac, and spermine solution; (4) $MgCl_2$, NaCac, and MPD solution; (5) $MgCl_2$, NaCac, spermine, and MPD solution. Melting curves for self-complementary DNA dodecamer in solutions; (6) H_2O; (7) $MgCl_2$, NaCac, spermine, and MPD solution.

Table 3: Thermodynamic parameters for d(CGCTGGCCACCG) and d(CGGTGGCCACCG) denaturations in different solutions

Sample	Solution	T_m (°C)	ΔH (kcal mol^{-1})	ΔS (cal mol^{-1} K^{-1})
d(CGCTGG -CCACCG)	(1) H_2O	-	17	54
	(2) $MgCl_2$, NaCac	39.7	23	73
	(3) $MgCl_2$, NaCac, spermine	47.9	20	63
	(4) $MgCl_2$, NaCac, MPD	40.3	25	80
	(5) $MgCl_2$, NaCac, spermine, MPD	44.9	28	88
d(CGGTGG -CCACCG)	(1) H_2O	-	29	88
	(5) $MgCl_2$, NaCac, spermine, MPD	72.7	63	210

The following observations may be made from Figure 16 and Table 3:

(*1*) Curve 1 and curve 6 are melting curves for the 'mismatched' and self-complementary DNA in water, respectively. It can be seen that the melting temperatures are low, 'melting' ranges are quite broad, and enthalpy changes are very small. The results indicate that the DNA molecules in water are unstable and flexible (indicated by the gradual conformational changes in a wide temperature range).

(*2*) Curve 2, which is the melting curve of 'mismatched' DNA in $MgCl_2$ and NaCac solution, suggests that adding counter ions (Mg^{2+}) and buffer (NaCac) can stabilize DNA as exhibited by the narrower melting range and higher enthalpy change. If spermine (only 1 mM) is added into the solution, a significant increase in T_m by 8.2 degrees is observed as shown in curve 3.

(*3*) Curve 4 reveals that the melting range is broadened and the melting tempera-

ture is lowered drastically by adding only 2% MPD into the $MgCl_2$, NaCac, 'mismatched' DNA solution. Such a small amount of MPD almost offsets the stabilizing effect of $MgCl_2$ and NaCac and the melting curve is similar to curve 1 which is for 'mismatched' DNA in water. This undermining effect of MPD is confirmed by curve 5 which corresponds to 'mismatched' DNA in the $MgCl_2$, NaCac, spermine, and MPD solution in which MPD decreases the T_m by 3 degrees compared with curve 3.

(*4*) Curve 7 is the melting curve for self-complementary DNA in $MgCl_2$, NaCac, spermine, and MPD solution. Curve 7 is a typical DNA double strand melting curve in that it shows a sharp increase in absorbance in a very narrow range, a high hypochromic effect, a much higher enthalpy change as well as a much higher melting point than its 'mismatched' counterpart.

2.3 Circular Dichroism Study

2.3.1 Introduction

While a UV study can only indicate an extensive stacking interaction between the bases in the helix, a circular dichroism (CD) study is able to provide more details on DNA conformations. Nearly all molecules synthesized by living organisms are optically active. The optical activity arises from the presence of chiral structures. A chiral structure is a structure that is not superimposable on its mirror image. DNA molecules are chiral mainly due to the helicity. Chiral structures can be distinguished and characterized by plane polarized light (PPL). Plane polarized light can be viewed as the superposition of right circularly polarized light (RCPL) and left circularly polarized light beams (LCPL). A medium with a chiral molecule interacts differently with the left and right circularly polarized light. After the PPL passes through the sample, each component remains polarized, but the radius of the RCPL circle is different from that of the LCPL circle. This phenomenon is called circular dichroism. When these two opposite circularly polarized light waves are combined, the result will be an elliptically polarized light. If the right circularly polarized light is absorbed more strongly than the left circularly polarized light, the spectrum shows a negative band; if the left circularly rotated light is absorbed more strongly than right circularly rotated light, the spectrum shows a positive band.

CD data are usually expressed in ellipticity (Θ) values:

$$\Theta \text{ (degrees)} = 2.303\,(A_L - A_R)\,180^{\circ}/4\pi = 32.98(\varepsilon_L lC - \varepsilon_R lC) = 32.98\Delta\varepsilon lC$$

where Θ is the ellipticity in degrees, A is the absorbance, ε is the extinction coefficient in units of $M^{-1}\,cm^{-1}$, l is the cell length in cm, C the is the optically active molecule concentration in M.

Molar ellipticity is defined as $[\Theta] = 100\Theta / lC$, and therefore

$$[\Theta] = 3298(\varepsilon_L - \varepsilon_R)$$

Circular dichroism is an appropriate method for investigating conformational differences of different forms of nucleic acids in solutions. It is extremely sensitive even to small changes in relative orientation of the neighboring bases in a polynucleotide. In general, the magnitude of the molar ellipticity [Θ] reflects the degree of the helicity, or the extent of base stacking; the better the bases stacked, the higher the ellipticity. The differences between the low and high temperature circular dichroism spectra are due to the conformational changes originating mainly from (*i*) the breakage of hydrogen bonds between the Watson-Crick base-pairs, (*ii*) the unstacking of the single strands with temperature, and (*iii*) the unstacking of the loop residues[67, 73].

2.3.2 Experimental

Materials

Two dodecamers, 'mismatched' and self-complementary, were studied by CD experiments. Four buffers were prepared:

(1) $MgCl_2$(16 mM), and NaCac (25 mM), pH 6.9

(2) $MgCl_2$(16 mM), NaCac (25 mM), and spermine (1.0 mM), pH 6.9

(3) $MgCl_2$(16 mM), NaCac (25 mM), and MPD (2%), pH 6.9

(4) $MgCl_2$(16 mM), NaCac (25 mM), spermine (1.0 mM), and MPD (2%), pH 6.9

An experimental solution for CD was prepared by diluting 6.0 μL of 1.0 mM DNA (double-stranded) into 114 μL of each above buffer.

Method:

The CD experiments were performed on a Jasco J500A spectropolarimeter. Sample temperatures were controlled by a thermostated cell holder which was connected to a HAAKE G water bath and HAAKE D1 pump. The scan range was 220 nm - 300 nm and the chart speed was 2 cm min^{-1}. The wavelength expansion was 5 nm cm^{-1}, and therefore the scan speed was $2 \times 5 = 10$ nm min^{-1}.

The corresponding buffer scan (*i.e.* blank) was performed at the room temperature before each DNA sample run and was rescanned at the room temperature after all the DNA spectra from the same sample had been collected. The sample was incubated at each experimental temperature for ten minutes before collecting spectrum to ensure that the DNA solution temperature had thermally equilibrated. The temperature reading on the thermometer in the water bath was recorded as the temperature of the solution. All the experiments were started at room temperature. The CD spectra were measured at six temperatures spanning from room temperature to 70°C.

2.3.3 Results:

All the blank scans of buffers showed that they were transparent from 220 nm to 300 nm (data not shown). The blank scan of the buffer was subtracted from every CD spectrum. The recorded experimental data were voltage readings. Every voltage value was converted into ellipticity Θ (in millidegrees) by multiplying a factor 41.24 (conversion factor obtained by calibration) and multiplying the sensitivity of the equipment for that specific scan. Molar ellipticity $[\Theta]$ was calculated by $[\Theta] = \frac{100\Theta}{lC}$ and all the spectra were plotted as $[\Theta]$ *vs.* λ. All the voltage-wavelength data from the experiments were pro-

cessed by Mathcad 6.0[78] (see Appendix B for details). The spectra for the 'mismatched' DNA are shown in Figures 17 to 20 and the spectra for the self-complementary DNA are in Figure 21.

Figure 17 is the CD spectra of 'mismatched' DNA in $MgCl_2$, NaCac solution. It is clear that the molar ellipticity on both negative and positive bands decreases as the solution temperature increases, indicating the gradual loss in helicity. The general helicity is maintained in the range of 20°C to 37°C. The shape of the spectrum at 41°C is somewhat different from that at 37°C which implies some change occurred in secondary structure. The shape of the spectrum at 55°C is quite different from that at 37°C as almost all molecules are single stranded at 55°C and therefore the spectrum is for the single strands. Further increasing the temperature to 70°C results in a decrease in ellipticity for the negative band.

Figure 18 shows the effect of spermine on DNA ellipticity. It can be seen that Figure 18 is quite similar to Figure 17 except that the large change in ellipticity occurs at higher temperature, between 41°C to 55°C.

Figure 19 shows the effect of adding MPD on ellipticity. The remarkable change in ellipticity in a wide temperature range refers to the significant conformation change of DNA in the same range. The elliptical intensity changes are large for the negative band but small for the positive band. For the negative band, the ellipticity decreases consistently as the temperature increases except for the abnormality seen at 37°C. However, for the positive band, the ellipticity increases as the temperature increases from 23°C, reaches a maximum at 37°C and then decreases to a minimum at 41°C and then increases again. It should be noted that at 70°C the ellipticity for the positive band is high, but very low for

the negative band.

Figure 20 is the CD spectra of 'mismatched' DNA in $MgCl_2$, NaCac, spermine, and MPD solution. Both spermine and MPD effects on DNA stabilization are revealed in this figure. Compared with Figure 18, spermine diminishes the significant elliptical decrease for the negative band especially before the DNA is melted (melting point 44.7°C). However, the positive band still shows the same trend as shown in Figure 19. This experiment was repeated five times with the same or independently prepared samples (data not shown) and consistent results were obtained, as illustrated in Figure 20.

Figure 21 is the spectra of self-complementary DNA in $MgCl_2$, NaCac, spermine, and MPD solution. The result is fairly straightforward. Only very small elliptical changes are observed which are consistent with the UV experiment (melting point 72.7°C) in that the DNA molecule remains in its intact double-stranded helical conformation in the over-all experimental temperature range.

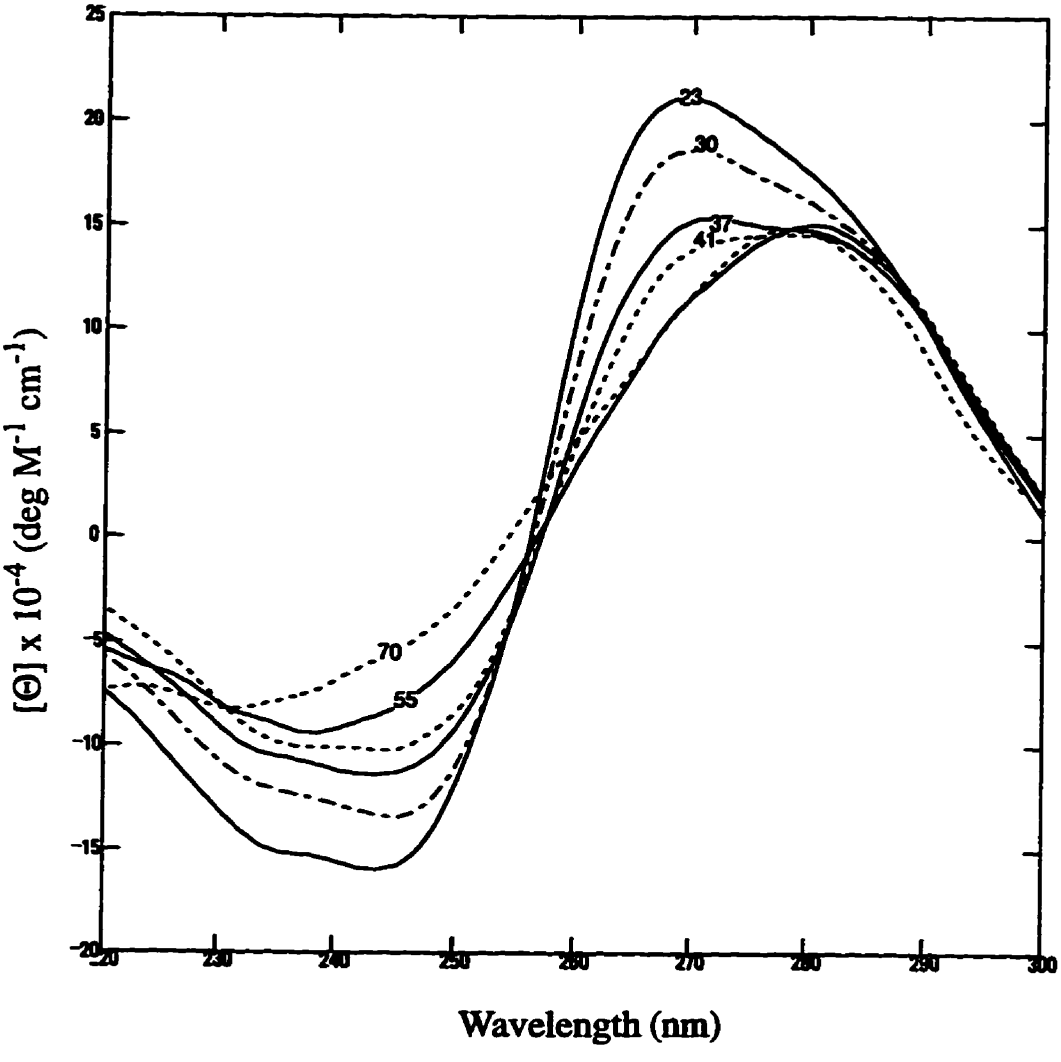


Figure 17. CD spectra as a function of temperature for 'mismatched' DNA in $MgCl_2$ and NaCac solution.

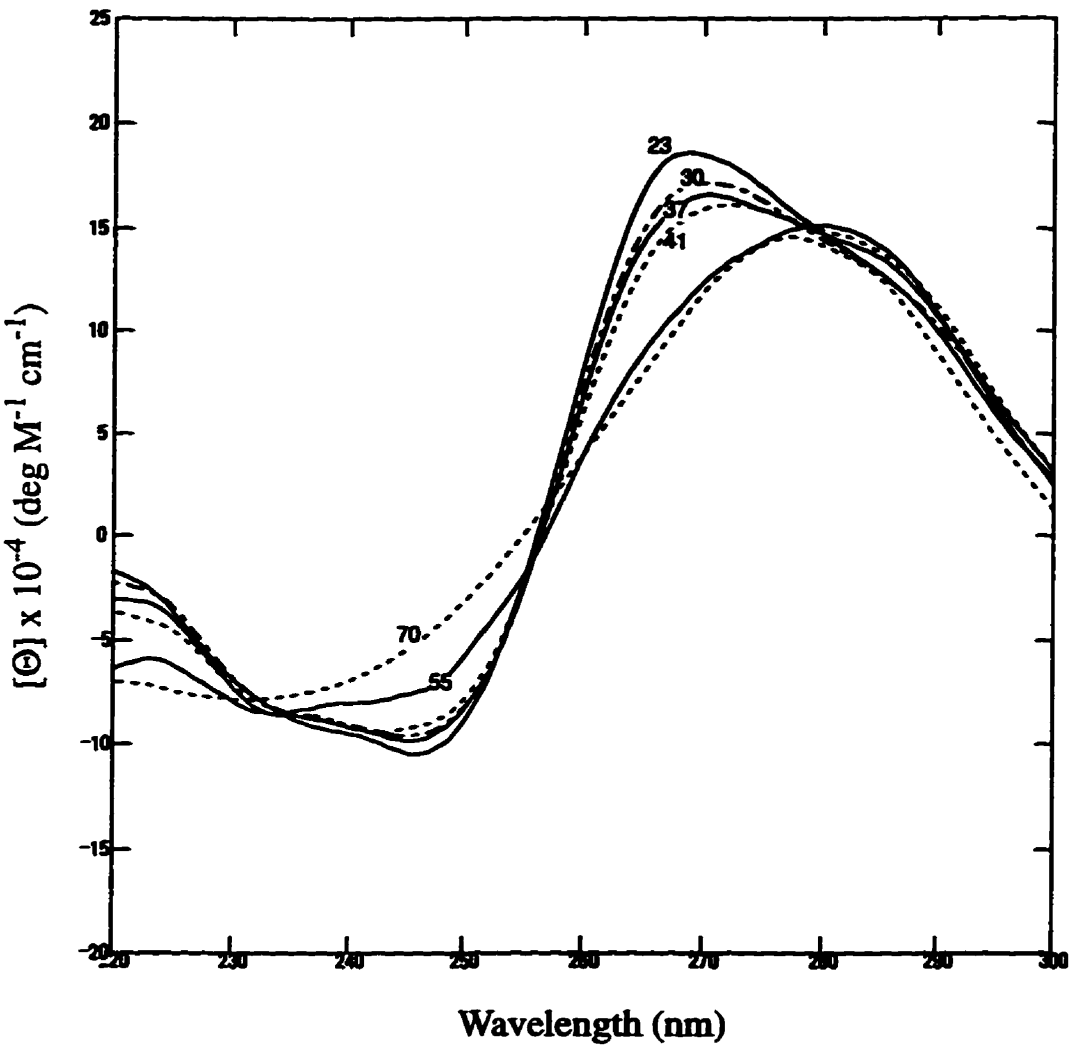


Figure 18. CD spectra as a function of temperature for 'mismatched' DNA in $MgCl_2$, NaCac, and spermine solution.

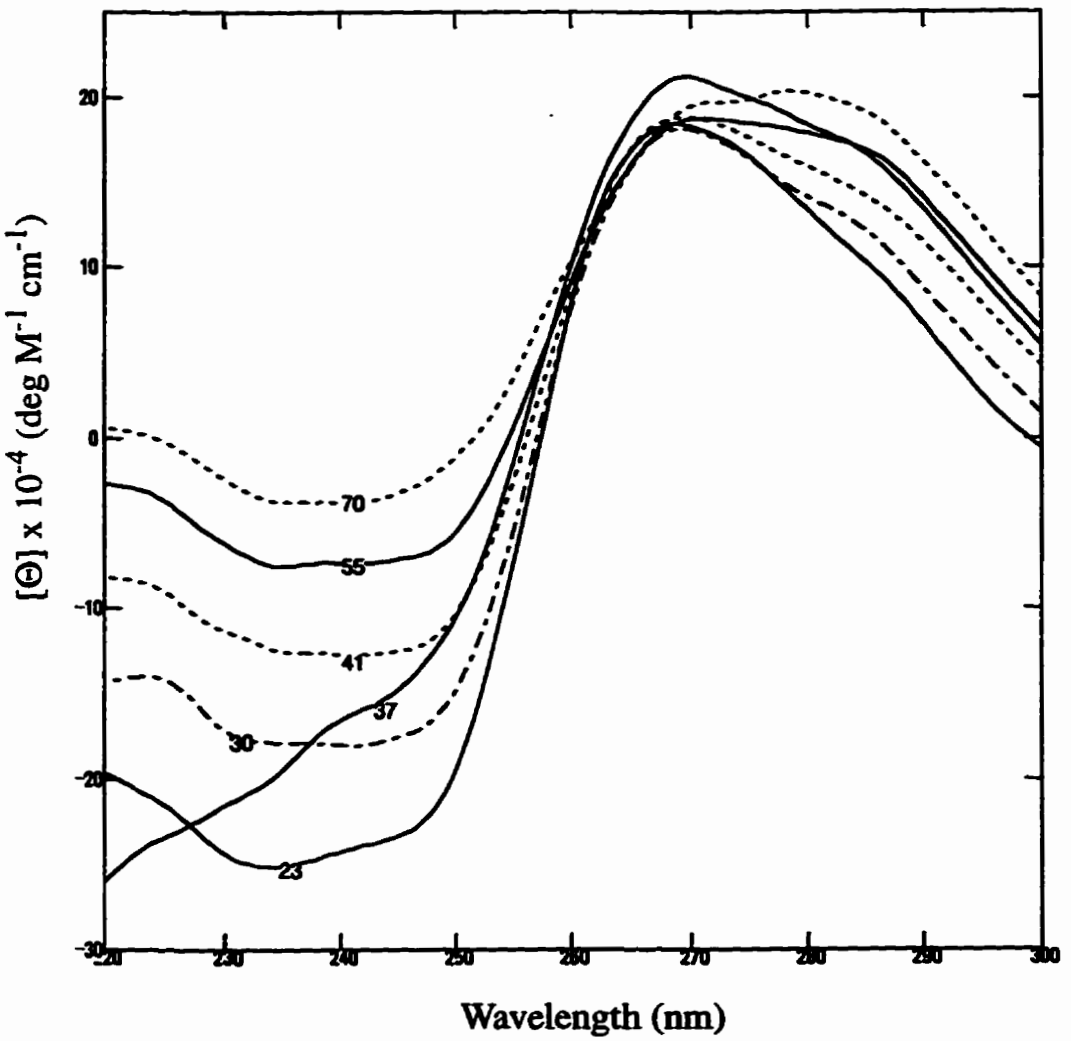


Figure 19. CD spectra as a function of temperature for 'mismatched' DNA in $MgCl_2$, NaCac, and MPD solution.

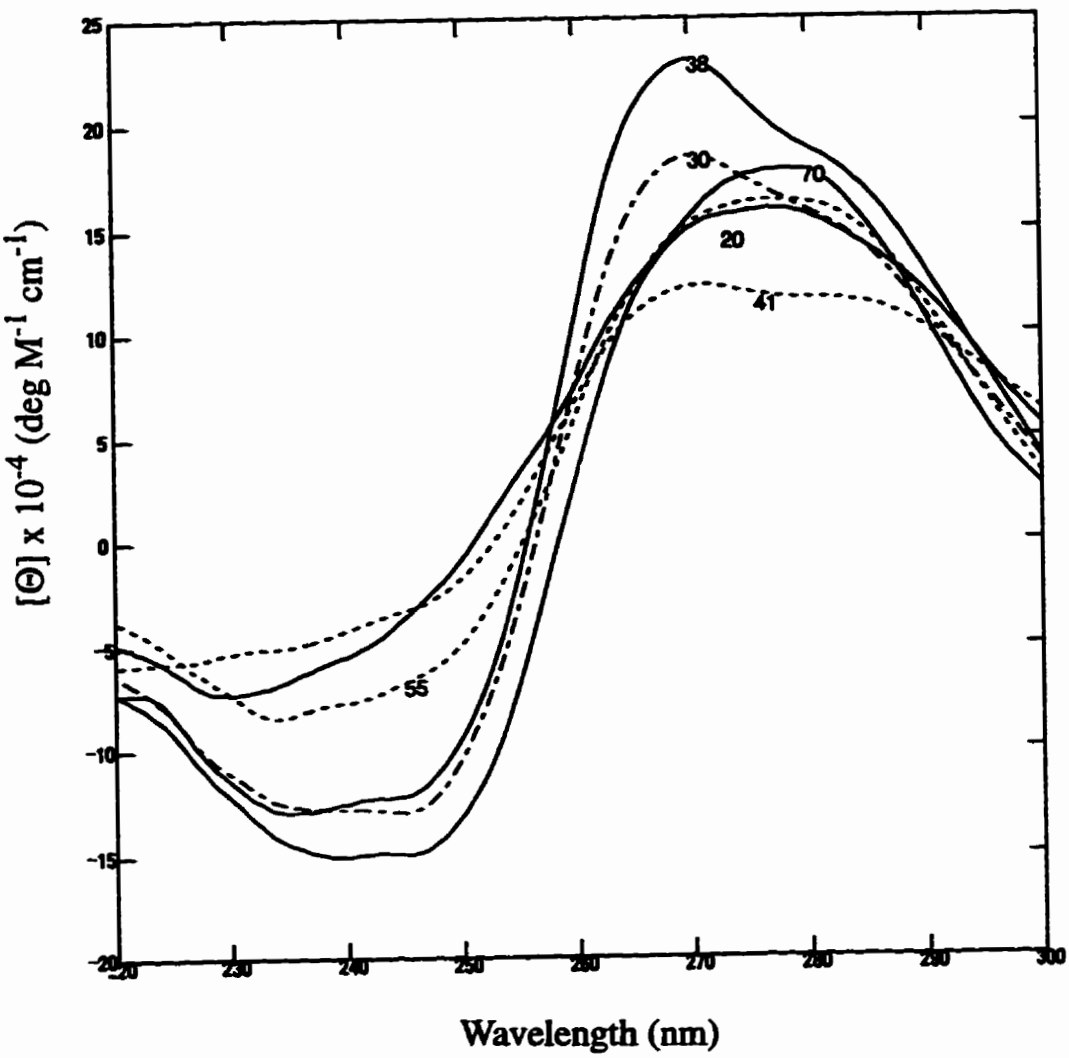


Figure 20. CD spectra as a function of temperature for 'mismatched' DNA in $MgCl_2$, NaCac, spermine, and MPD solution.

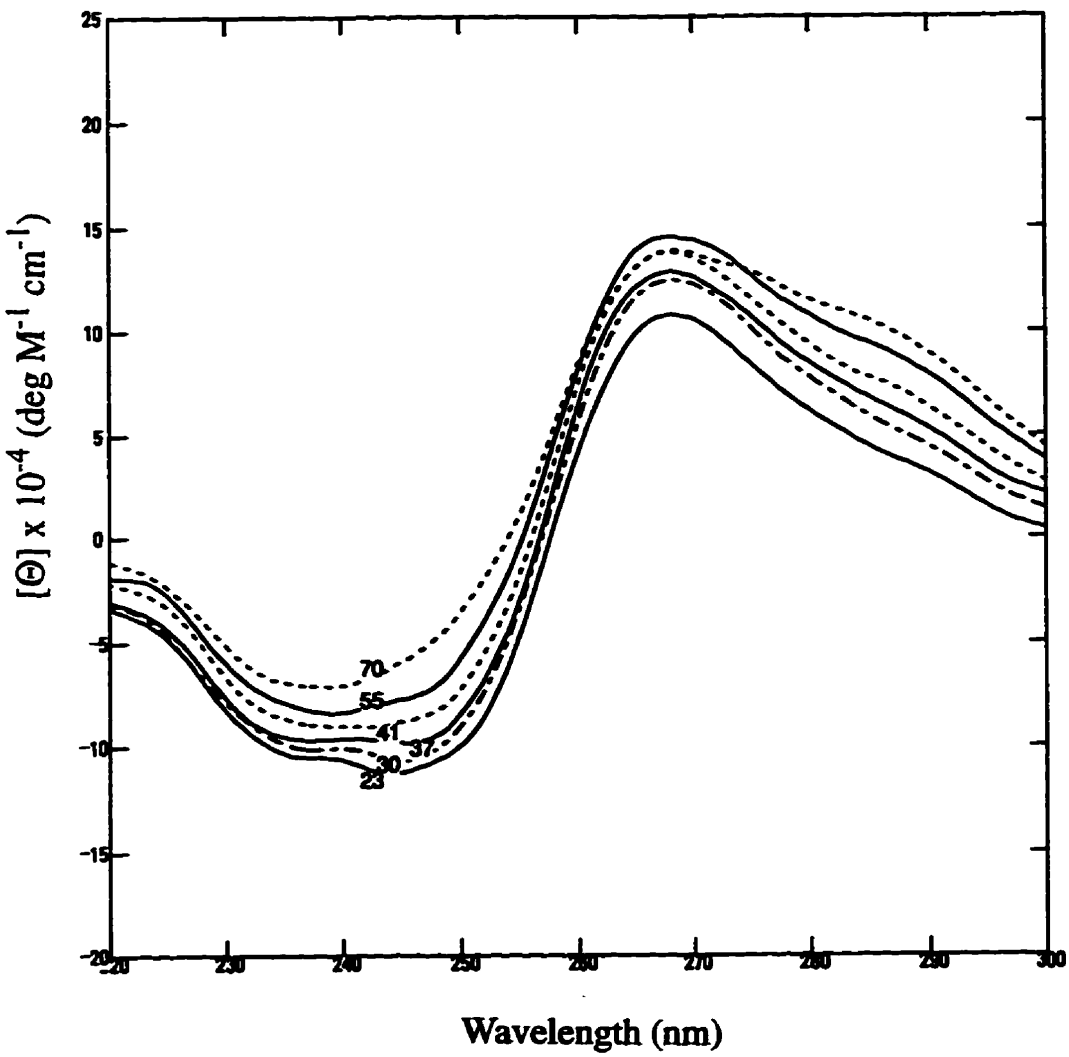


Figure 21. CD spectra as a function of temperature for self-complementary DNA in $MgCl_2$, NaCac, spermine, and MPD solution.

2.4 Gel Electrophoresis Study

2.4.1 Introduction

Macromolecules with different charge densities and conformations generally have different mobilities in an electric field. Electrophoresis is a technique that makes use of the charged property of macromolecules to analyze or separate the molecules. The most popular type of electrophoresis is gel electrophoresis. The gel is a three-dimensional network of the constituent polymers formed when the gelation takes place. It exists as random distributions of solid material and open spaces, or pores and can be thought of as a sieve that retards large molecules relative to small molecules. Different extents of polymerization of the gel correspond to different resistances to mobility. The resistance also varies with the size and shape of the macromolecule. In polyacrylamide gel electrophoresis (PAGE), gels are made by the polymerization of acrylamide and N,N'-methylene bisacrylamide. The pore size of the gel, which can be made of molecular dimensions, is controlled by the proportion of polyacrylamide in the gel, ranging from 3.5 to 30 percent, depending on the size of the fragments of interest. Negatively charged DNA molecules will move between the buffer-filled pores of the gel from the negative to the positive electrode. The ability of DNA to squeeze through the small-pore regions of the gel depends on the length and conformations of the molecules. The longer the DNA molecule, or the bulkier if the DNA molecules have the same length, the slower the migrating speed. Urea is able to disrupt hydrogen bonds and is therefore often added into the gel when it is necessary to maintain DNA in single-stranded form. High concentrations (7 - 8 M) of urea are required to dissociate hydrogen bonds.

Electrophoresis of DNA is done almost exclusively with continuous-buffer sys-

tems. Tris-borate electrophoresis buffer (TBE) is used for DNA polyacrylamide gel electrophoresis. Formamide, like urea, disrupts hydrogen bonding and is sometimes used as a denaturant for DNA molecules[79].

2.4.2 Experimental

Each DNA sample for gel electrophoresis was prepared by dissolving 10 μg of DNA in 20 μL of 20 mM $MgCl_2$ and 25 mM NaCac (pH 7.0) buffer. The mixture of a 1:1 ratio (moles) of two single-stranded hexamers, d(CGGTGG) and d(CCACCG), was carefully made with total amount of 20 μL and total mass of 10 μg.

The DNA samples were run in two different gels, a nondenaturing and a denaturing gel, respectively. The nondenaturing gel ($8 \times 10 \times 0.15$ cm^3) was prepared and run as follows:

(*1*) Two stock solutions were prepared:

5× (five times concentrated) Tris-borate electrophoresis buffer (TBE)

Tris base	27 g
boric acid	13.75 g
Na_2EDTA	2.33 g

dilute to 500 mL with H_2O

10% ammonium persulfate (APS)

ammonium persulfate	1 g
H_2O	10 mL

(*2*) Two clean glass plates which were held apart by spacers and sealed with tape along the two sides and bottom were prepared for pouring the gel.

(*3*) 2.9 g acrylamide, 1.0 g N, N'-methylene bisacrylamide, and 2 mL of 5× TBE were diluted to 9.9 mL with H_2O. This solution was mixed with 100 µL 10% APS and 30 µL TEMED (N, N, N', N' tetramethylethylene diamide). The mixture was poured immediately into the space between the glass plates. An appropriate comb was inserted and the gel was allowed to polymerize at room temperature for two hours. The comb was removed and the gel together with the glass plates were attached to the electrophoresis tank.

(*4*) 200 mL 5× TBE was diluted with 800 mL H_2O and was used to fill the reservoir of the tank. The air bubbles trapped in the wells were removed with a pasteur pipet. The DNA samples and dyes were loaded using a plastic micropipet. The samples loaded were as follows*:

lane 3 40 µL 0.05% bromophenol blue (fast blue) and 0.05% xylene cyanol (slow blue).

lane 5 20 µL 1:1 mixture (moles) of d(CGGTGG) and d(CCACCG) + 20 µL 40% sucrose

lane 6 20 µL 'mismatched' dodecamer d(CGCTGGCCACCG) + 20 µL 40% sucrose

lane 7 20 µL self-complementary dodecamer d(CGGTGGCCACCG) + 20 µL 40% sucrose

(*5*) The electrophoresis was performed at room temperature and run at a voltage of 100 V. At the end of the run, the plates were detached from the tank and

---

* Lanes 1, 2, and 4 were not used.

the two glass plates were removed. A photograph of the gel was taken under UV light (Figure 22(*a*)).

The denaturing gel ($20 \times 35 \times 0.15\,cm^3$) was prepared and run as follows:

(*1*) and (*2*) are the same as the above.

(*3*) 19.0 g acrylamide, 1.0 g N, N'-methylene bisacrylamide, 46 g urea, and 20 mL of 5× TBE were diluted to 100 mL with H_2O. This solution was mixed with 500 µL 10% APS and 150 µL TEMED. The gel was poured immediately.

(*4*) The DNA samples were heated to 90°C for two minutes and were cooled quickly in ice water.

(*5*) 200 mL 5× TBE was diluted with 800 mL H_2O and used to fill the reservoir of the tank. The air bubbles trapped in the wells were removed with a pasteur pipet. The DNA samples and dyes were loaded using a plastic micropipet. The DNA samples were redissolved in one of the solutions, (*i*) deionized formamide + 10 mM EDTA, and (*ii*) 40% sucrose. The samples loaded were as follows:

lane 1 20 µL 'mismatched' dodecamer d(CGCTGGCCACCG) + 20 µL solution (*i*)

lane 2 20 µL 'mismatched' dodecamer d(CGCTGGCCACCG) + 20 µL solution (*ii*)

lane 3 20 µL 1:1 mixture (moles) of d(CGGTGG) and d(CCACCG) + 20 mL solution (*i*)

lane 4 20 μL single-stranded hexamer d(CGGTGG) + 20 μL solution (*i*)

lane 5 20 μL single-stranded hexamer d(CCACCG) + 20 μL solution (*i*)

lane 6 40 μL 0.05% bromophenol blue

(*6*) The electrophoresis was performed at room temperature and run at a voltage of 400 V. At the end of the run, the plates were detached from the tank and the two glass plates were removed. A photograph of the gel was taken under UV light (Figure 22(*b*)).

2.4.3 Results

There are two bands in Figure 22(*a*) lane 5. The fast band is faint while the slow band is dark. The slow migrating band corresponds to duplexed hexamer d(CGGTGG)/ d(CCACCG) which is the main species in the solution. The UV melting curve (data not shown) indicates that at room temperature, about 25% of the duplexed d(CGGTGG)/ d(CCACCG) has melted to single strands. Consequently, it is concluded that the faint band corresponds to the slower migrating species of the two single strands, d(CGGTGG) (see the results for Figure 22(*b*) below). The faster migrating single strand has migrated into the bottom buffer and is not visible.

The band in Figure 22(*a*) lane 6 is for the 'mismatched' DNA. It is clear that the migrating speed of this dodecamer is closer to the duplexed hexamer than to the duplexed self-complementary dodecamer* (lane 7). If the 'mismatched' DNA were duplexed, it should comigrate with duplexed self-complementary dodecamer. The mobility difference

---

* From the UV melting curve, it is known that the self-complementary dodecamer is duplexed at room temperature.

cannot be explained by the 'mismatched' DNA being a single-strand because it has a melting curve and shows its T_m is about 39.7°C.

The two bands in lane 3 of Figure 22(*b*) are of almost equal intensities indicating that no hexamer duplexes are formed. By comparing lane 3 with lanes 4 and 5, it is concluded that the faster migrating strand is d(CCACCG) while the slower migrating strand is d(CGGTGG). That almost no difference between lanes 1 and 2 indicates that the 'mismatched' DNA molecules adopt the same conformations under these two conditions. The migrating speed of the 'mismatched' DNA is close to that of the slower migrating strand of the hexamer, d(CGGTGG). A striking result is that and the migration difference between the 'mismatched' DNA (lane 1) and the slow strand of the hexamer, d(CGGTGG) is less than that observed between the two hexamer strands (lane 3).

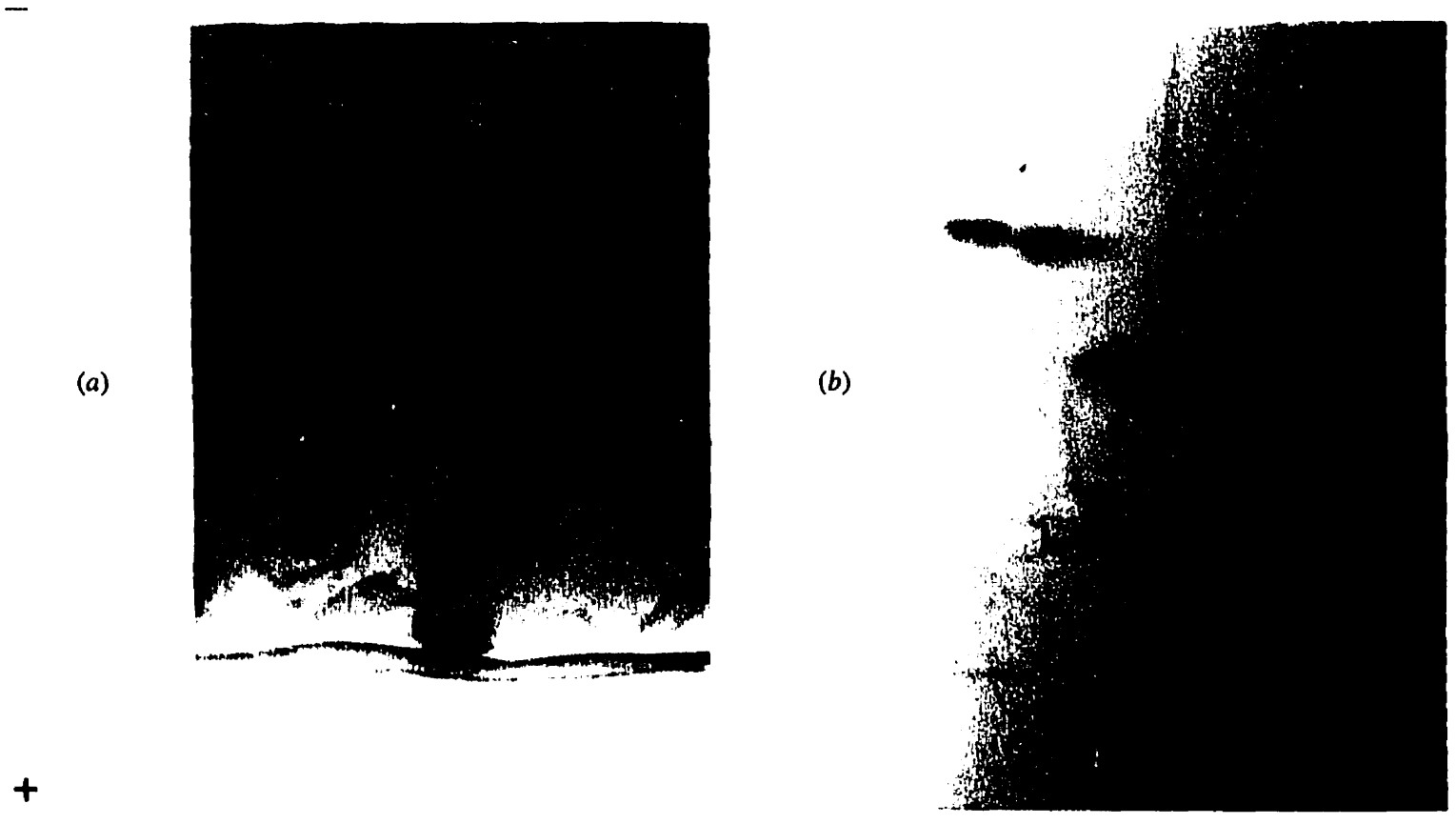


Figure 22. Electrophoretic mobilities in (*a*) nondenaturing polyacrylamide gel. 0.05% bromophenol blue (fast blue) and 0.05% xylene cyanol (slow blue) (lane 3); 1:1 mixture (moles) of d(CGGTGG) and d(CCACCG) (lane 5); 'mismatched' dodecamer d(CGCTGGCCACCG) (lane 6); self-complementary dodecamer d(CGGTGGCCACCG) (lane 7), and (*b*) denaturing gel. 'mismatched' dodecamer d(CGCTGGCCACCG) + formamide (lane 1); 'mismatched' dodecamer d(CGCTGGCCACCG) + sucrose (lane 2); 1:1 mixture (moles) of d(CGGTGG) and d(CCACCG) + formamide (lane 3); single-stranded hexamer d(CGGTGG) + formamide (lane 4); single-strand hexamer d(CCACCG) + formamide (lane 5); 0.05% bromophenol blue (lane 6).

2.5 Discussion

A striking difference in enthalpies between the 'mismatched' and self-complementary DNA in the solution of $MgCl_2$, NaCac, spermine, and MPD indicates a large difference in their conformations. All the CD spectra for these two molecules show two bands. The negative and positive bands are almost centrosymmetric about 256 nm, which means the two DNA dodecamers adopt B-type conformations[80, 81]. If the 'mismatched' DNA were duplexed, the only difference between such a duplex and the self-complementary one would be the C-C mismatches at positions 3 and 10. Even if the fewer hydrogen bonds and geometry change on neighboring bases are considered, the effects are still not significant enough to produce such a large variation in ΔH. One explanation for the large ΔH difference is that 'mismatched' DNA is hairpinned, while the self-complementary DNA is duplexed.

This explanation is consistent with the results of the gel electrophoresis experiments. Comparison of the mobilities of the duplexed self-complementary DNA, the 'mismatched' DNA and the duplexed hexamer clearly suggests that the 'mismatched' DNA is hairpinned under the nondenaturing conditions. The length of a hairpinned 'mismatched' dodecamer is about half that of a duplexed self-complementary dodecamer and is closer to the length of a duplexed hexamer. The slower mobility of the hairpin structure compared to that of the duplexed hexamer as shown in Figure 22(*a*) is due to the bulky loop that the hairpin structure possesses leading to a lower charge density. The migrating difference between the 'mismatched' DNA and the slow single strand of the hexamer, d(CGGTGG) in Figure 22(*b*) is also due, mostly, to charge density differences, the former being less dense and consequently moving slower. The mobility difference between the two single

strands of hexamer can be explained by similar arguments; d(CGGTGG) is bulkier due to a high proportion of purine bases than in d(CCACCG). The fact that the migrating difference between the 'mismatched' DNA and the single strand d(CGGTGG) is less than that observed between the two hexamer single strands in Figure 22(*b*), suggests that even under the denaturing condition, the 'mismatched' DNA is still hairpinned.

How do individual crystallizing agents affect the 'mismatched' DNA conformations in aqueous solutions? Water surrounding DNA interacts strongly with the molecule and hydrophobic force is mainly responsible for the stabilization of secondary and tertiary structures. The high dielectric constant of water and hydrated counterions diminish the phosphate-phosphate electrostatic repulsion. The bases in DNA can stack on top of each other to form a helical structure largely due to hydrophobic forces. The degree of hydration of DNA plays an important role in its conformation. If salts or organic solvents are added into an aqueous DNA solution, the activity of the water is changed and water molecules are withdrawn from the outer hydration shell. High relative humidity favors B-form DNA. Reduced humidity or increased ionic strength leads to a transition from the B-form to other forms, such as A- and Z-forms [23, 80, 81]. A schematic presentation of the CD spectra for A- and B-forms of DNA is shown in Figure 23.

DNA double helices are stabilized mainly by base stacking which is a consequence of the hydrophobic interactions between the bases. While hydrogen bonds are required for the fidelity of replication, they contribute little to the double helix stability[2, 23]. Adding a relatively nonpolar alcohol into a DNA solution strengthens hydrogen bonds between the complementary bases, while disrupting the hydrophobic effects of the base stacking[2]. Many studies concerning the alcohol effects on DNA conformations[82-85] have shown that

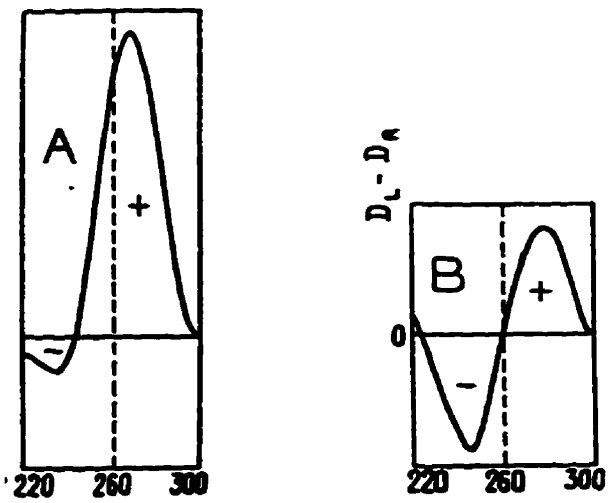


Figure 23. Schematic presentation of the CD spectra for the A- and B-form DNAs. The dotted lines are drawn through the absorption maximum. From Ref. 81.

at low alcohol concentrations DNA molecules change conformation due to the corresponding destruction of the binding water. At high alcohol concentrations, changes in secondary structure were observed, such as the transition from B-DNA to A-DNA. When the alcohol concentrations were very high, the condensation of DNA was observed which may not be accompanied by any changes in secondary structure.

The procedure of DNA 'melting' in water is the unstacking of its component residues. A broad 'melting' range is due to the lack of cooperativity in the single-strand helix-to-coil transition (not an all-or-none transition). There exist other conformations between the helix and coil states, *i.e.* many possible intermediate conformations accompanying the denaturation process. The enthalpy change during such a denaturation equals the total stacking energy.

Curves 1 and 6 in Figure 16 demonstrate that the DNA molecules in water are single-stranded. The hydrogen bonds between the complementary bases of the two strands no

longer exist. The adjacent bases on the single strands are partially stacked upon one another and the whole molecule is arranged in a helical fashion, which was designated "single-strand helix" by Pörschke[75] and Appleby *et al.*[76]. The single-strand helix is relatively stable at low temperature. As the temperature increases, the helical chain gradually becomes more and more flexible and finally the whole strand becomes a random coil (maximum unstacked state).

The relatively sharper increase in absorbance, higher hypochromic effect, and a melting point around 40°C in curve 2 of Figure 16 indicate that addition of $MgCl_2$ and NaCac into the solution stabilizes the helix and promotes the formation of a double strand. Mg^{2+} ions bind to the negatively charged phosphate group in DNA and shield the negative charge and thus attenuate the charge repulsion. Many studies reveal that the increase in melting temperature depends on the logarithm of the Mg^{2+} and Na^+ ion concentrations. Mg^{2+} ions stabilize helical structures very effectively at concentrations several orders of magnitude lower than the effective concentration of Na^+ ions[86]. In the corresponding CD spectra (Figure 17), the ellipticity change between 37°C and 41°C implies some change in secondary structure, which is consistent with the melting point of 39.7°C. The large difference between the CD spectrum at 37°C and at 55°C can be explained by DNA being hairpinned at 37°C and while at 55°C it is single-stranded. The decrease in ellipticity from 55°C to 70°C is due to the unstacking of the single strands. The very weak but nonzero ellipticity at 70°C supports the previous result from NMR studies[87, 88] that even at very high temperature some base stacking still remains.

The UV and CD results verify the significant charge effects of spermine on DNA.

The steeper melting profile and higher T_m in curve 3 of Figure 16 indicate that a more rigid helix is formed on the introduction of spermine. Spermine has been shown to cause an increase in the melting temperature of DNA, which is dependent approximately upon the logarithm of the spermine concentration but independent of DNA concentration[86]. The most likely explanation for this is the effective charge neutralization of the phosphate group by forming a DNA-spermine complex, which results in an increase in the van der Waals' and hydrogen-bonding interaction. The effectiveness even at low spermine concentration signifies the high affinity of spermine for DNA, although there are relatively few quantitative measurements of this affinity. It is also possible that spermine might stabilize DNA by cross-linkage, but the lack of dependence of DNA concentration on T_m makes this less likely. It can be seen from the corresponding CD (Figure 18) spectrum that the large change in ellipticity occurs at higher temperature between 41°C to 55°C compared to Figure 17, which is also consistent with the UV result indicating that DNA melts at 47.9°C in this solution. Furthermore, the DNA helices become more rigid and more resistant to thermal denaturation as smaller elliptical changes are noticeable before and after the melting temperature. In another words, the denaturation is more cooperative. Therefore, the CD experiment provides the same conclusion that spermine stabilizes the DNA by increasing its melting point.

MPD is a di-alcohol with a very weak polarity. The UV melting curve (Figure 16, curve 4) demonstrates that even 2% MPD could strongly destabilize this 'mismatched' DNA as shown by the lower melting temperature and the broad melting range. This is due to MPD molecules displacing water molecules and diminishing the hydrophobic force and hence, weakening the base stacking. The higher the temperature, the easier the replace-

ment, and the more flexible the DNA molecules. The cytosine bases in the 'mismatched' region are not able to form normal Watson-Crick hydrogen bonds. The distortion induced by the C-C mismatch may affect the hydrogen bonding and stacking of the neighboring normal Watson-Crick base-pairs in the double-stranded hairpin stem. Consequently, the whole stem is not a rigid double helix. The existence of a C-C mismatch lessens the interactions between the two strands and the bases may readjust their geometry more easily as the temperature increases.

The fact that the UV melting curve for the 'mismatched' DNA in solution with $MgCl_2$, NaCac, and MPD (curve 4 in Figure 16) is similar to the curve for DNA in water (curve 1 in Figure 16) indicates that the hydrogen bonds between the DNA hairpin stem is very weak. The corresponding CD (Figure 19) results can be interpreted to mean that as the temperature increases the DNA undergoes a change from B- to an A-type conformation in the presence of MPD. The increasing temperature promotes the displacement of water. This gradual dehydration of DNA molecules is similar in effect to reducing the humidity by increasing the alcohol concentration. The increase in the positive band and decrease in the negative band as the temperature increased in the CD spectra clearly indicate the transition from B- to A-form. The spectrum at 70°C is quite similar to that of A-form DNA (Figure 23). The decrease in the positive band at 41°C and the abnormality in the negative band at 37°C are very likely a result of the breakage of the hairpin duplexed stem ($T_m = 40.4$°C).

Figure 20 is the CD spectra of 'mismatched' DNA in $MgCl_2$, NaCac, spermine, and MPD solution. Both the stabilizing effect of spermine and destabilizing effect of MPD on DNA are revealed in this figure. Simply stated, the stabilizing effect of spermine is

revealed by the elliptical change in negative band, whereas, the destabilizing effect of MPD is revealed by the elliptical change in positive band. The positive band still shows the same trend as that the ellipticity increases as the temperature increases below the melting point with a sharp decrease during melting and increases again as the temperature is further increased. The large change in spectral shape between 41°C and 55°C is due to the transition from hairpin to single-strand which is also predictable from the UV melting point (T_m = 44.9°C).

Hairpin conformations have been studied in a number of non-palindromic oligonucleotides where duplex formation is very difficult[25, 89-92]. Detailed structural information was obtained in some cases since these hairpins can be formed in large concentration. DNA hairpin loops of four to five nucleotides were shown to have higher stability than smaller or larger loops. In B-form DNA, the minimum phosphate-phosphate distance occurs across the minor groove after two to three nucleotides are stacked in helical fashion on the 5' side of the stem. Thus, loops of four to five nucleotides in DNA could maximize stacking interactions while satisfying steric constraints. Bolmmer *et al.*[93] have pointed out that the hairpin with a four bases loop (L1 - L4) may be stabilized by formation of an extra base-pairing between L1 and L4. This was claimed to be possible in loops consisting of a PyL2L3Pu sequence, but not in a PuL2L3Py sequence as is the case of the 'mismatched' DNA. Thus, the most likely structure for this 'mismatched' dodecamer d(CGCTGGCCACCG) is a hairpin consisting of a B-DNA stem of three Watson-Crick base-pairs, one C-C mispair and a loop formed by the four remaining central bases which are partially stacked (Figure 24).

As three hydrogen bonds are formed between a G-C base-pair while only two

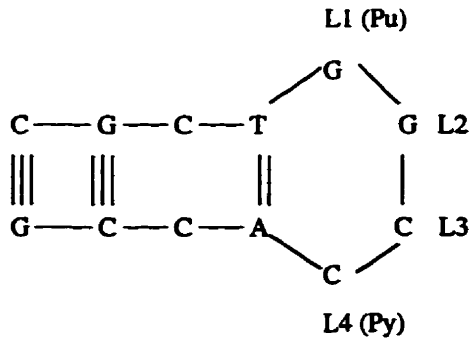


Figure 24. Possible hairpin structure of the 'mismatched' DNA dodecamer d(CGCTGGC-CACCG).

between a A-T base-pair, it is likely that the melting of the hairpin will start with the opening of the A-T base-pair followed by the opening of C-G base-pairs at the end of the stem.

The question remains, why does the self-complementary DNA favor the duplex to the hairpin structure under the same solution conditions? There is no clear answer at this point. The UV melting curves of some nucleotides exhibit biphasic behavior in contrast to the monophasic melting at low salt concentrations[35, 68-70], *i.e.* there are two transitions during a whole melting procedure. The transition at the low temperature (transition I) which corresponds to duplex to hairpin is sharp, while the transition at high temperature (transition II) corresponding to hairpin to coil is relatively broad (Figure 25). It can

The predominant structural form of DNA molecules in solution will depend partly on both the DNA and salt concentrations. When the DNA concentration is constant, the melting curve shows a biphasic to monophasic transition with increasing salt concentrations (Figure 25, from curve 2 to 3). For 'mismatched' DNA molecules, such a transition happens at a much higher salt concentration than that for self-complementary DNA molecules. It is possible that the melting curves for the 'mismatched' DNA are actually bipha-

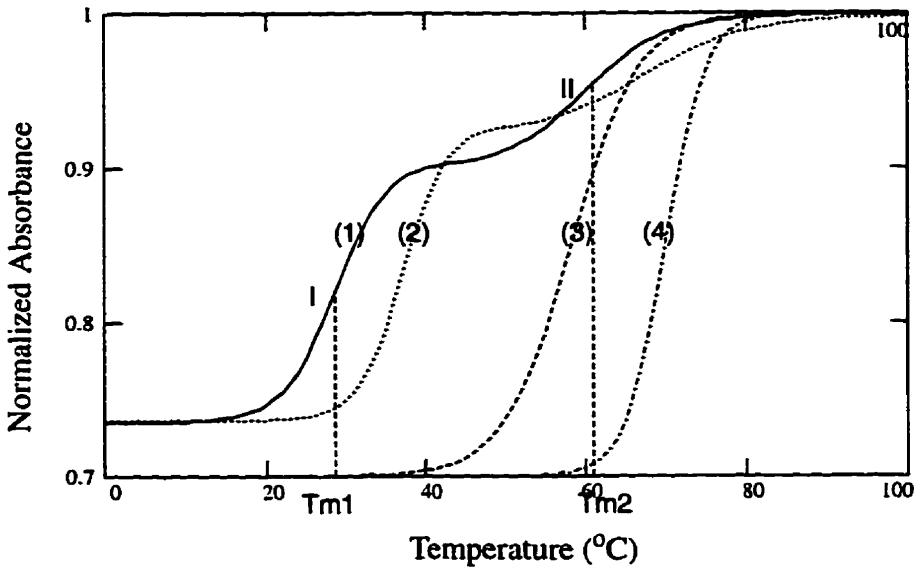


Figure 25. A schematic representation of normalized melting curves as a function of NaCl concentrations when the DNA concentration is constant. (1) 0.001 M (2) 0.01 M (3) 0.1 M (4) 1 M

sic as the curve 1 in Figure 25; however, the curves obtained are only the transition II section (transition I is not achievable). Under the same solution condition, the melting behavior of self-complementary DNA d(CGGTGGCCACCG) is monophasic as depicted by curve 3 or 4 in Figure 25 where only two species are available—duplexes at low temperature and single strands at high temperature.

2.6 Conclusion

The CD, UV and gel electrophoresis studies of d(CGCTGGCCACCG) and d(CGGTGGCCACCG) were undertaken to elucidate the structure of C-C 'mismatched' DNA in crystallizing solutions. The results suggest that under the conditions of the crystallizing solutions the 'mismatched' DNA forms a hairpin structure, possibly, with the first and the last four bases in the stem in B-DNA helical form and the central four bases partially stacked in the loop. Under the same crystallizing conditions, the self-complementary DNA is duplexed in B-form.

2.7 Suggested Further Studies

Since hairpin crystals are difficult to obtain, efforts should be made on setting up crystallizing experiments under conditions in which duplex DNA will be the predominant species. From Figure 25, it is known that DNA duplexes can be obtained either by lowering the experimental temperature if the melting is biphasic as seen in curves 1 and 2, or adjusting the components in the crystallizing solution to obtain monophasic melting as seen in curve 3 and 4 where there are only two species, duplexes and single strands in solution. Inspection of the previous results reveals some important trends which will favor the formation of duplex relative to hairpin:

(*1*) High DNA concentration

T_{m1} (melting point of transition I) increases linearly with DNA concentration as the DNA concentration is increased, while T_{m2} (melting point of transition II) is independent of DNA concentration. If the DNA concentration is high enough, only the duplex-to-single-strand transition (curve 3 or 4 in Figure 25) is observed.

(*2*) High salt concentrations

T_{m1} increases linearly with the logarithm of the cation concentration. This is true for both Mg^{2+} and Na^{+} ions. Mg^{2+} ions have an effect at concentrations of several orders of magnitude lower than the effective concentration of Na^{+} ions[86]. However, very high salt concentrations will condense DNA.

(*3*) High spermine or spermidine concentrations

T_{m1} increases linearly with the logarithm of the spermine and spermidine concentration. Spermine is about ten times effective than spermidine. Still,

very high polyamine concentrations will cause DNA condensation.

(*4*) Low temperature

Transition I occurs at lower temperature than transition II. Therefore, the chance for obtaining duplexes is higher at low temperature.

Based on the results of the studies above and those from the previous reports, it is proposed the 'mismatched' DNA dodecamer be crystallized according to the following principles:

(*1*) High DNA concentration in the droplet

Practically it may be in the range from 0.5 mM to 2.5 mM.

(*2*) High salt concentrations, especially when the sequence contains mismatches

A DNA segment which contains C-C mismatches can form a duplex at pH values less than 6.2. Aboul-ela *et al.*[93] have reported that mismatches containing cytosine destabilize the duplex more than any other mismatches. At neutral pH, a C-C mispair may form only one hydrogen bond. This type of mismatch may significantly distort the phosphate backbone of the duplex. The stabilization of such distorted backbone would require high salt concentrations (~300 mM or greater).

(*3*) Relatively high polyamine concentrations

The practical concentrations should be tested with care.

(*4*) Low MPD concentration due to its destabilizing effect

(*5*) Initial pH of 6.0 to 5.0

Even though low pH will aid in forming one more hydrogen bond between

one protonated cytosine and one unprotonated cytosine[12], it is unclear whether lowering pH could induce some detrimental effects on the crystallization.

(6) Low temperature

Many DNA crystals can only be obtained at 4°C because the melting points are very low, *e.g.* the hexamer crystal d(CGGTGG)/d(CCACCG) obtained before in this lab.

Chapter 3

Self-Complementary DNA Dodecamer d(CGGTGGCCACCG): Crystallization and Determination of Cell Parameters

"The pursuit of a structure is rather like hunting: it requires some skill, a knowledge of the victim's habits, and a certain amount of low cunning."[94]

3.1 Introduction

There are numerous physical chemical approaches that yield information regarding macromolecular structures. Some of these methods, such as CD, NMR and molecular dynamics, are becoming increasingly valuable in defining detailed protein structure, particularly for low molecular mass proteins. However, few techniques offer more basic structural information about biological macromolecules than the x-ray single crystal diffraction analysis method which provides a detailed and precise description of the structures of macromolecules. X-ray diffraction analysis is unique in its ability to determine atomic positions in a crystal lattice and hence define molecular conformation which enables one to explain biological processes at molecular level. The insights derived from the x-ray studies have had a profound impact on the advances in molecular biology.

The crystallographic results of DNA parallel those obtained in solution and from fibers, but the higher resolution of crystallography shows an impressive conformational diversity at the level of individual nucleotide units within A-, B-, and Z-form DNA families. The conformations observed in crystals result from interactions both within the duplex and between duplexes and other species in the extended crystal structure.

Computers of extraordinary speed and capacity as well as advanced computer graphic systems are common tools in crystallographic studies. The question is then, what prevents the full utilization and exploitation of this enormously powerful approach? The answer is that for application of the method to a particular macromolecule, the molecule must first be crystallized. Not only must crystals be grown, but they must be diffraction-quality crystals, *i.e.* crystals suitable for a high-resolution x-ray diffraction analysis.

3.1.1 Fundamentals of X-ray Diffraction Analysis

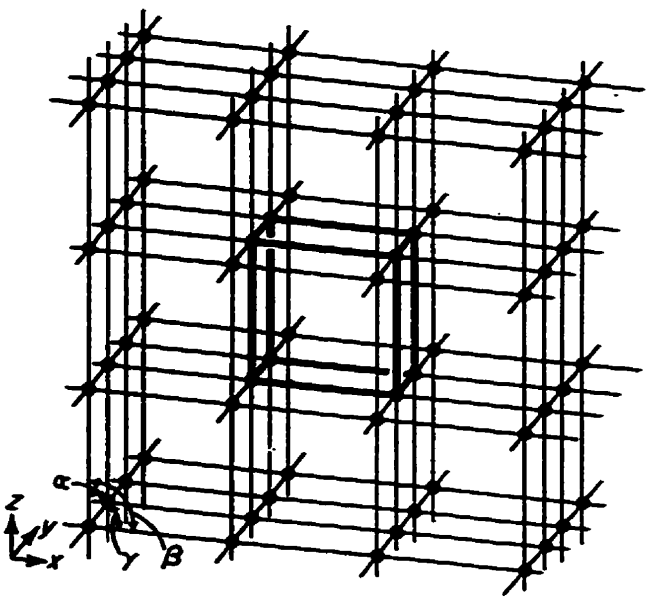
The fundamental principles of x-ray diffraction are well established. X-rays are electromagnetic radiation of short wavelength (0.1 - 100 Å) and are generated by the sudden deceleration of high speed electrons at a target material. X-rays in the wavelength range 0.5 - 2.5 Å are used in crystallographic study because the distance between adjacent atoms in crystalline solids are on the same scale. X-rays impinging on crystalline solids will give rise to observable diffraction patterns as the x-rays are scattered coherently from the electrons of the constituent atoms of the solid substances.

A crystal is a substance with a very high degree of internal order in three dimensions. Atoms, molecules, ions, or a complex assembly of molecules of which a crystal is composed are arranged in a regular way such that it may be regarded as being built up by continuing three-dimensional translational repetition of some basic structural pattern. If the detailed structure of the comprising objects is ignored, each repeated unit can be simply represented by a point, called a lattice point. All lattice points have the same environment and each is indistinguishable from any other. A crystal lattice is an infinite set of lattice points that may be generated from a single starting point by infinite repetition of a set of fundamental translations that characterize the lattice (Figure 26(*a*)). A unit cell is the basic parallelepiped repeating unit in a crystal. It contains a complete representation of the contents of the repeating unit of the crystal and is always chosen as the ensemble of lattice points comprising the smallest box with the highest symmetry. Thus, any crystal may be regarded as being built up of regularly stacked unit cells in three dimensions (Figure 26(*b*)). As each unit cell has identical contents, the structure determination of a crystal is reduced to the determination of the spatial arrangement of atoms within a single unit

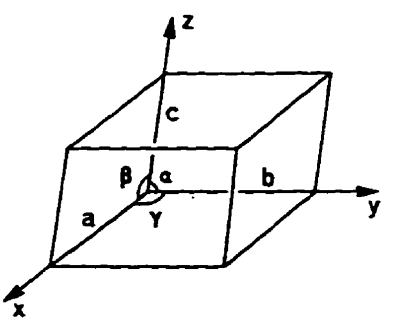
cell. If the unit cell has some internal symmetry, *i.e.* only a portion of it is unique, the work is further reduced to the determination of this unique portion called an asymmetric unit. The parameters that define a unit cell are axial lengths a, b, c, and interaxial angles α, β, γ. Once the crystallographic axes are chosen, the set of planes that have intercepts a/h, b/k, c/l, where h, k, l are integers, along the x, y, z axes respectively, can be expressed in Miller indices as (hkl) (Figure 26(*c*)).

In addition to the crystal lattice in real space, there is a second, related to the first, called the reciprocal lattice. The reciprocal lattice is constructed from the same origin as the real lattice by taking normals to all possible direct lattice planes (hkl) and terminating each normal at a point at a distance $1/d_{hkl}$, where d_{hkl} is the perpendicular distance between planes of the set (hkl). The set of points so determined constitutes the reciprocal lattice (Figure 27). The larger a real cell is, the smaller its reciprocal cell will be. Therefore, large molecules, like DNA, have correspondingly small reciprocal cells. An x-ray diffraction pattern recorded by photographs visualizes the general geometry and symmetry of the reciprocal lattice.

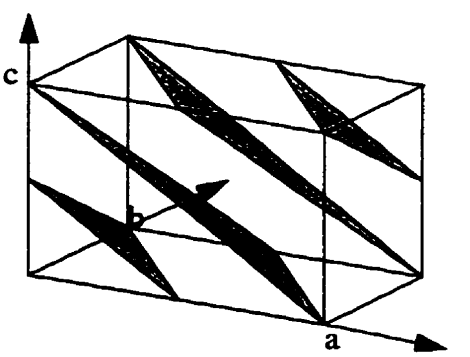
The interaction of x-rays with a crystal is a complex process. Although strictly speaking it is a combined scattering and interference effect, this process is often described as a diffraction phenomenon. Bragg's treatment gives a clear and accurate picture of a diffraction pattern, even though it is an oversimplification of the complete process. Bragg's law states that an x-ray beam incident on a pair of parallel planes with interplanar spacing d at an angle θ will show maximal constructive interference if the path difference $\delta = 2d\sin\theta$ is equal to an integral number of wavelengths, *i.e.* when $2d\sin\theta = n\lambda$ is



(*a*)



(*b*)



(2, 1, 2) or $(\bar{2}, \bar{1}, \bar{2})$ planes

(*c*)

Figure 26. (*a*). Three-dimensional lattice, showing unit cell (heavy line). From Ref. 95. (*b*) Unit-cell nomenclature. The reference axes x, y, z are right-handed. The length of the unit-cell edge parallel to each reference axis is respectively a, b, c, and the interaxial angles are denoted α, β, γ. From Ref. 96. (*c*) The planes corresponding to Miller indices (2,1,2) are a series of lattice planes which cut the unit cell h times along a, k times along b, and l times along c.

satisfied (Figure 28). If Bragg's law is considered in the form of $\sin\theta = \frac{n\lambda}{2}\left(\frac{1}{d}\right)$, it is seen that $\sin\theta$ is inversely proportional to d. Since θ is half the angular deviation of the diffracted beam from the direct beam, the structures with large d will exhibit compressed diffraction patterns, which is just the case of macromolecules.

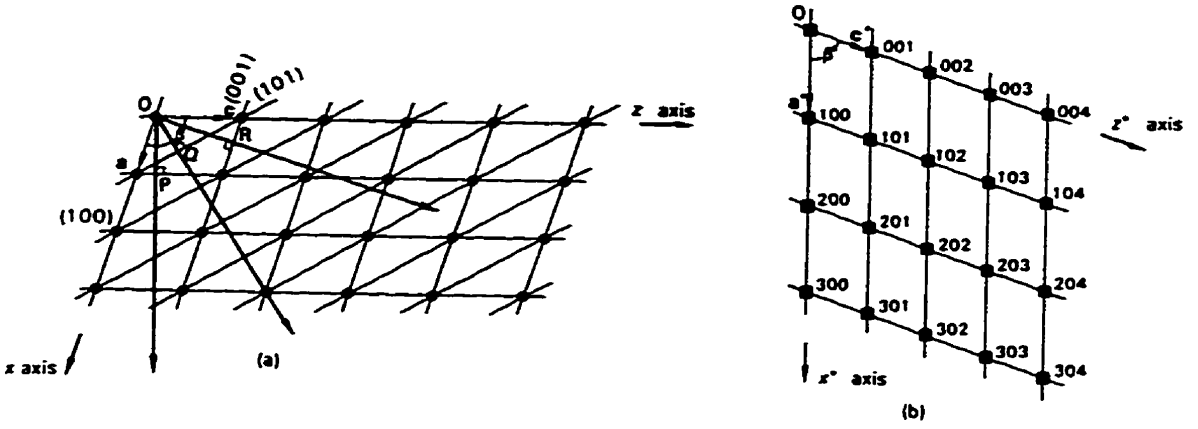


Figure 27. Direct and reciprocal lattices: (*a*) monoclinic P, as seen in projection along b, showing three families of planes. (*b*) corresponding reciprocal lattice. From Ref. 97.

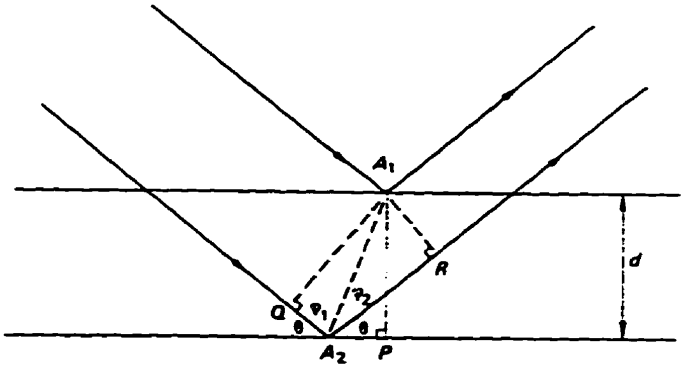


Figure 28. Detailed geometry of x-ray reflection. The path difference between the two typical rays reflected from successive planes is ($QA_2 + A_2R$), and is equal to $2d\ sin\theta$. From Ref. 97.

The scattering power of a given atom from a given reflection is known as its scattering factor (f_{hkl}). When an incident x-ray impinges on the atoms in a unit cell, all the scattered wavelets will add together and interfere with one another, and thus the total scattering from an object is sensitive to its internal structure. The structure factor F(hkl) = $|F(hkl)|\exp(i\varphi)$ expresses the combined scattering of all the atoms in a unit cell compared to that of a single electron. Since only the intensity of scattering which is proportional to the square of the structure factor, $I(hkl) \propto |F(hkl)|^2$, can be measured, the phase φ which is required for structure determination is lost in x-ray experiments. This is the phase problem.

There are two different problems involved in the collection of diffraction data. The first is to locate the position of various reflections, from which the size, shape and symmetry of the reciprocal and therefore the real lattices may be calculated. The second is the determination of the intensity at every point in the reciprocal lattice, *i.e.* to every reflecting plane in the real lattice. The intensity is ultimately related to the distribution of electron density in the unit cell, *i.e.* the molecular structure.

Bragg's law states that each value of 2θ at which a diffraction maximum is observed is a function only of the cell dimensions and the wavelength of the x-rays. Thus, the unit cell dimensions can be determined from the 2θ angles with the aid of a four-circle diffractometer (Figure 29). In a diffractometer, the crystal is centered in the incident beam, and the detector can move in a plane parallel to the instrument base (referred to as the equatorial plane) which contains the crystal and the incident and diffracted x-ray beams. The value of 2θ may be then measured from the position of the incident beam and the position of the maximum of the diffracted beam. Crystals can be rotated about three axes

(χ, ϕ, and ω) independently, and a fourth angle 2θ which is concentric with ω.

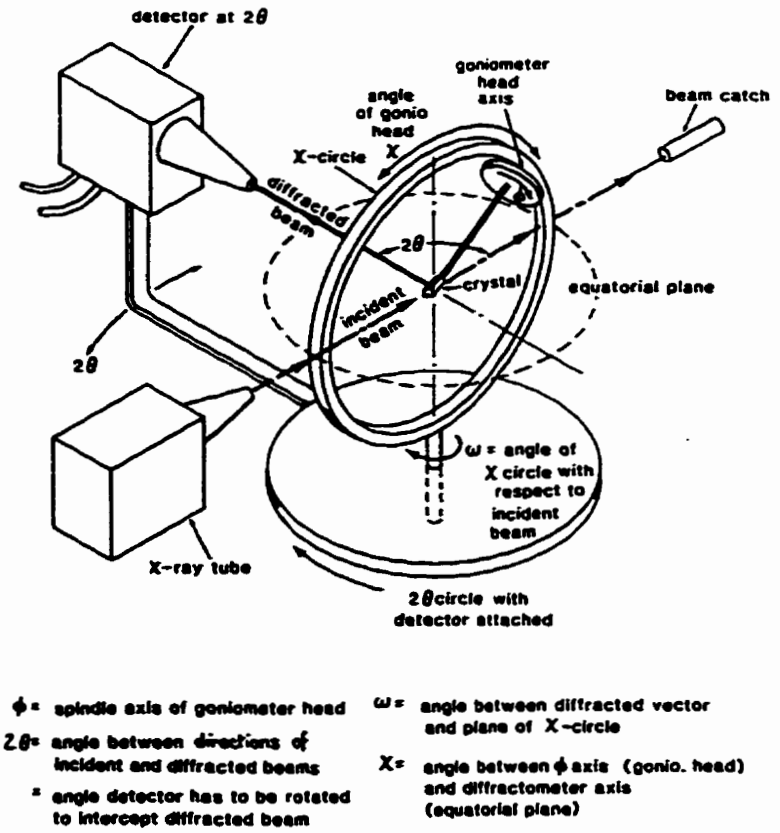


Figure 29. Diffractometry. From Ref. 98.

The crystals to be measured on a diffractometer are selected under a microscope because the crystal quality is critical to the success of a data collection. Normally, the external forms of crystals manifest their internal integrity. Therefore the crystals with regular faces, sharp edges and shiny surfaces are favored. Furthermore, crystals must extinguish polarized light properly.

Once a crystal has been chosen, it is mounted in a sealed capillary. An obvious difference between macromolecule crystals and small molecule crystals is the high solvent content in the former. Maintaining the solvent content during data collection is very

important. Dehydrated crystals will not diffract properly, and in some cases not at all. Therefore, macromolecule crystals used in diffraction experiments are usually mounted in sealed capillaries containing a drop of mother liquor from which the crystals were grown. Because DNA crystals contain a lot of solvent, they are very fragile and must be handled with extreme care[38, 99].

A polaroid photograph of the diffraction pattern taken after centering the crystal records the positions of some strong reflections as white spots. Most reflections of macromolecular crystals detectable on photographs have small 2θ angles, *i.e.* the diffraction pattern is concentrated around the center of the photograph. Once the coordinates of the reflection spots are available, the cell parameters can be determined by the *index* routine of the diffractometer control software[100]. An alternative for obtaining initial reflections is the *search* routine. The *search* routine locates and centers the reflections by a search of reciprocal lattice. Before starting data collection, a better cell based on high angle reflections is essential because high angle cell parameters have smaller standard deviations and therefore are more precise.

The *Laue* routine is generally performed after obtaining the high angle cell to ensure that the crystal system which is the basis for the following data collection is correct. This routine is designed for checking the intensity symmetry in the reciprocal lattice. The check is carried out by generating the indices of potentially equivalent reflections, measuring the corresponding intensities, and statistically evaluating the goodness of fit.

Values of the four corresponding angles (χ, ϕ, ω, 2θ) for all possible reflections can be calculated based on the best unit cell. The angles for a particular reflection are set by a computer, the intensity of the reflection is measured with the detector and recorded,

together with the measurements of the background intensity near the reflection. A whole data set collection is normally processed by advancing incrementally through the Miller indices. In this way a systematic scan of all desired reflections is done completely and automatically.

The intensity data collected are only the raw intensities. In most cases they represent all the information that will be obtained from physical measurements on the crystal, and the subsequent solution of the structure will depend on the skillful extraction of the information contained within the intensities. The procedure for converting the measured intensity data to a form suitable for further calculation is called data reduction.

3.1.2 Macromolecular Crystallography

The analysis of any crystal by x-ray diffraction requires every unit cell be identical. Any reduction in the extent of order will result in a decrease of information in the diffraction pattern. Macromolecule crystals are strikingly different from small molecule crystals although they are optically indistinguishable. Macromolecule crystals suffer many sources of statistical disorder in addition to their weak reflection magnitude. Disorder may arise from different sources. (*i*) Some molecules may occupy positions that vary slightly from lattice points. The displacement is enhanced by the relatively large spaces between the adjacent molecules. (*ii*) The number of bonds in a macromolecule are much greater than a conventional small molecule. Since it is intermolecular as well as intramolecular interactions (salt, hydrogen, hydrophobic, *etc.*) that maintain the integrity of the crystal, the enormous conformational possibilities are responsible in large part for the difference in ordered properties between small and macromolecule crystals. (*iii*) Macromolecule crystals normally contain a very high solvent content, from twenty to seventy percent by

volume. The solvent usually occupies large channels in the crystals. Generally, those solvent molecules weaken the intermolecular interactions and exert a destabilizing influence on the crystal lattice. By comparison with small molecule crystals, macromolecules are very soft, crush easily, disintegrate if allowed to dehydrate, and diffract x-rays poorly.

Conventional diffraction measurements encounter extreme challenges to technological limits when they are applied to macromolecular crystals. A more powerful x-ray beam is required for macromolecule diffraction because of the weak intensity of reflection. Furthermore, a longer data collecting period, typically from a couple of days to even a week, is essential because there are as many as several tens of thousands of reflections to be collected. Thus, macromolecular crystals undergo serious damage due to prolonged exposure to x-ray radiation. Usually a number of crystals are needed to complete a structural analysis and the levels of detail are no greater than 1.5 Å and frequently no better than 2.0 Å resolution[101].

Crystallographic structure determination of a macromolecule entails a sequence of experimental and analytical operations. The route to solution of a macromolecular crystal structure is: (*i*) crystallization of the macromolecule; (*ii*) x-ray diffraction data collection; (*iii*) phase evaluation; (*iv*) synthesis of an electron density map; (*v*) establishment of an atomic model; (*vi*) refinement of the model to an optimal match with the observations[105].

3.2 Crystallization

3.2.1 Initial Trials

The most critical step in determining a structure is acquiring diffraction-quality crystals. A great deal of labor, patience and ingenuity are essential to carry out a systematic search of crystallization conditions, and at least a little luck. Both the size of the crystal and the regularity of the molecular packing are essential to the success of analysis.

Obtaining diffraction-quality crystals usually proceeds in two steps. The first is the screening of the crystallization conditions to obtain the initial crystals; the second is the optimization of these conditions to improve crystal size and quality. The discovery of initial crystals can be obtained from an incomplete factorial method (see Chapter 2) and grid screening. An incomplete factorial method was first tried in the crystallization experiments of the self-complementary DNA d(CGGTGGCCACCG). The pH was kept at 6.9 and therefore was not considered as a factor in matrix design. The factors involved were DNA, $MgCl_2$, spermine, and MPD concentrations. The experimental matrix is shown on Table 4.

Some practical considerations when setting up the self-complementary dodecamer were:

(1) [DNA] = 0.1 - 1.0 mM (double stranded concentration) in the droplet.

(2) $[Mg^{2+}]$ / [DNA] = 0.5 - 32

(3) [spermine] / [DNA] = 1 - 6

(4) [NaCac] = 10 - 40 mM (it has been observed that changing the [NaCac] in this range does not make a noticeable difference in crystallization.)

(5) pH = 6.9

Table 4: Incomplete factorial method experimental matrix for the self-complementary DNA d(CGGTGGCCACCG) crystallization screen

Trial No.	DNA	$MgCl_2$	spermine	MPD
1	+	+	+	+
2	+	+	+	-
3	+	+	-	+
4	+	+	-	-
5	+	-	+	+
6	+	-	+	-
7	+	-	-	+
8	+	-	-	-
9	-	+	+	+
10	-	+	+	-
11	-	+	-	+
12	-	+	-	-
13	-	-	+	+
14	-	-	+	-
15	-	-	-	+
16	-	-	-	-

Figure 30 is a schematic phase diagram for a macromolecule showing its solubility as a function of precipitating agent concentration. The most important region on the diagram is the supersaturated region which supports the formation of stable nuclei as well as the crystal growth. The metastable region supports the crystal growth, but not the formation of stable nuclei. Diffraction-quality crystals can only grow from these two regions. In the labile region, the probability of both nuclei formation and the speed of growth are

high. However, the probability of spontaneous and uncontrolled nucleation is also enhanced.

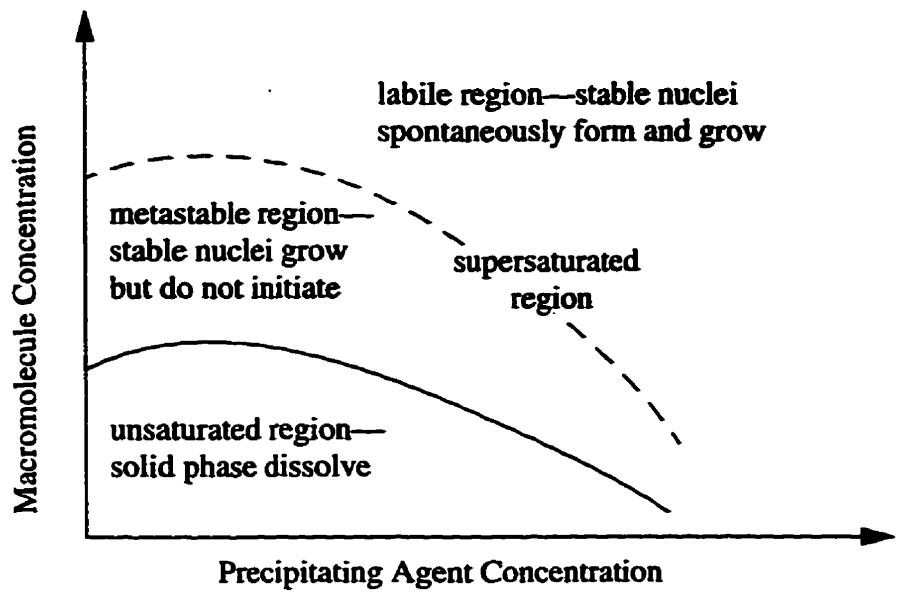


Figure 30. A phase diagram of a macromolecule describing its solubility as a function of precipitating agent concentration.

Usually, precipitate appeared in some droplets within a day or so after the droplets were setup. Based on the knowledge of the phase diagram, it was deduced that the droplets were at the far side of the labile region. Showers of microcrystals may have resulted from this condition due to the high growing speed. However, if a lot of precipitate was present, the concentration of macromolecule was decreased to half of the original trial amount, not only to save sample, but also to bring the droplets closer to the supersaturated region. Decreasing the precipitating agent concentration also worked well.

3.2.2 Optimization of Crystallizing Conditions

Grid screening

If the crystals obtained are not large enough for x-ray diffraction analysis, the first option to consider is varying the MPD concentration and/or the DNA concentration in the droplet to shift the droplet from the labile region to the supersaturated region as suggested by the phase diagram. A more refined screen can then be set up and searched in the neighborhood of the initial promising condition, a process referred to as grid screening. The initial stage is to design a screen bracketing the reagent conditions which produced the crystals (Figure 31). In the crystallizing experiments of the self-complementary dodecamer, no crystals appeared from the droplets set up according to the incomplete factorial design. However, grid screening was performed on some of the promising conditions and microcrystals appeared after one week by only increasing the MPD concentration from 10% to 22%.

Some 'crystalline' particles were first observed as shown in Figure 32(*a*) (Table 5, condition *a*). The shape of the particles was irregular and no sharp external edges were observed. The crystals did not extinguish under the polarized light. The crystalline material actually was at best composed of irregularly packed short-ordered DNA aggregates. They are known as spherulites. The presence of such aggregates indicated that it may be possible to grow larger crystals by adjusting the crystallizing conditions.

The first available crystal which extinguished polarized light properly was from the droplet using PEG400 as the precipitating agent (Table 5, condition *b* and Figure 32(*b*)). However, it was clear that the quality of the crystal was not good simply judged by the external irregular surfaces.

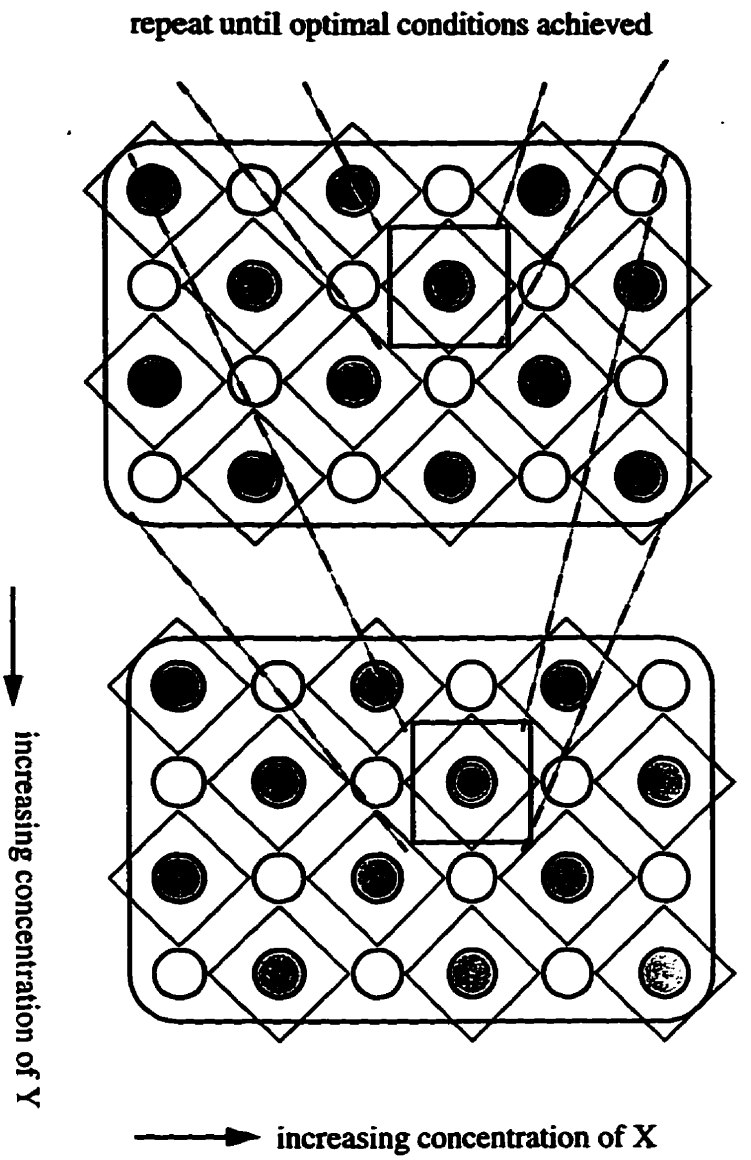


Figure 31. Schematic for successive narrowing of optimal conditions for crystallization.

A second kind of crystal was obtained from the droplet with MPD as the precipitating agent (Table 5, condition *c* and Figure 32(*c*)). Unfortunately, this condition was not reproducible, probably because the condition was close to the of boundary of the metastable region. The difficulty in nuclei formation was where the problem existed, but once nuclei were formed they could grow steadily.

The protocol leading to the first diffraction-quality crystal was modified from the 24-condition matrix for the crystallization of nucleic acid fragments reported by Berger *et al.*[41]. The first 18 out of 24 conditions were tried on the dodecamer and no crystals appeared in the first two and a half months. The diffraction-quality crystals were obtained by adjusting the crystallizing factors based on all those 18 conditions. During the optimization of the crystallization conditions, it was found that $BaCl_2$ worked as well as $MgCl_2$, and so it was the case with KCl and NaCl (at least the crystals were indistinguishable under the microscope).

Initially, attention was focused on optimizing the DNA and MPD concentrations as suggested by the phase diagram. If the DNA concentration alone was increased from 0.17 mM to 1.67 mM, a lot of small but very good quality crystals appeared in a day due to the fast nucleation and growing speed (Table 5, condition *d* and Figure 32(*d*)) which meant the droplet was in the labile region. If the MPD concentration alone was increased from 6.67% to 23.3%, very good quality crystals appeared the next day and usually only limited nuclei were formed (Table 5, condition *e* and Figure 32(*e*)) which suggested that the droplet was in the supersaturated region. Even though the initial growing speed of such crystals was fast, it obviously slowed down after a month. The following tuning of the crystallization conditions did not improve the crystal size significantly. As yet, these crys-

tals are still not large enough for data collection on a sealed tube diffractometer. However, some preliminary results have been obtained from a couple of these small crystals on the Brookhaven National Laboratory's synchrotron.

If the MPD concentration was further increased to about 30%, crystal growth directions spiked out from the original crystal surfaces due to the high growing speed (Table 5, condition *f* and Figure 32(*f*)) which indicated the droplet was at the far labile region. Sometimes, the same multidirectional growth occurred in crystals crystallized by condition *e* some weeks after the droplet was set up as the MPD concentration in the droplet became high enough. Crystals of different external shape were also obtained from a relatively high salt solution (Table 5, condition *g* and Figure 32(*g*)). Unfortunately, those crystals are still too small to determine if they are different types of DNA or simply only the same DNA form with different lattices.

The crystal which yielded a complete data set was obtained by using $ZnCl_2$ in place of $MgCl_2$ and simultaneously adjusting the other crystallizing factors. **There is no published work reporting the crystallization of an oligonucleotides with Zn^{2+} as a counterion. Therefore, this appears to be the first example of such a crystallization.** The crystal had sharp edges and extinguished the polarized light perfectly at every 90 degrees. Each crystal was prismatic with an equilateral triangular cross-section. The crystals appeared in 24 hours and stopped growing after a period of an additional 60 hours (Table 5, condition *h* and Figure 32(*h*)). The crystal in Figure 32(*i*) (Table 5, condition *i*) is a side view (parallel to a 3-fold axis) of the crystal which yielded the whole data set collected locally and Figure 32(*j*) is the end view (down a 3-fold axis) of a different crystal obtained from the same crystallizing condition as (*i*). The impressive external 3-fold sym-

metry of Figure 32(*j*) just reflected its internal 3-fold symmetry which was later found by x-ray analysis.

Table 5. Crystallizing Conditions for Self-complementary DNA Crystals

Condition	DNA (mM)	salt (mM)	spermine (mM)	NaCac (mM)	precipitant (v/v%)	pH	Temperature	Reservoir (v/v%)
a	0.1	0.40 $MgCl_2$	0.4	25.0	5% MPD	6.9	4°C	20% MPD
b	0.2	3.0 $MgCl_2$	2.4	30.0	15% PEG	6.9	room	35% MPD
c	0.2	4.0 $MgCl_2$	0.8	30.0	2% MPD	6.9	room	30% MPD
d	1.67	7.4 $MgCl_2$, 29.6 KCl	4.3	14.3	21% MPD	6.9	room	35% MPD
e	0.17	6.67 $BaCl_2$, 26.7 NaCl	4.0	13.3	23% MPD	6.9	room	35% MPD
f	0.19	11.1 $MgCl_2$, 21.2 NaCl	3.1	9.1	30% MPD	6.5	room	35% MPD
g	0.1	88 $MgCl_2$	0.1	25.0	10% MPD	6.9	room	35% MPD
h	0.47	13.8 $ZnCl_2$, 55.0 NaCl	8.25	27.5	5.5% MPD	6.9	room	35% MPD
i	0.87	8.5 $ZnCl_2$, 34 NaCl	5.1	17	1.7% MPD	7.45	room	10% MPD
j	0.87	8.5 $ZnCl_2$, 34 NaCl	5.1	17	1.7% MPD	7.45	room	10% MPD

(*a*)

(*b*)

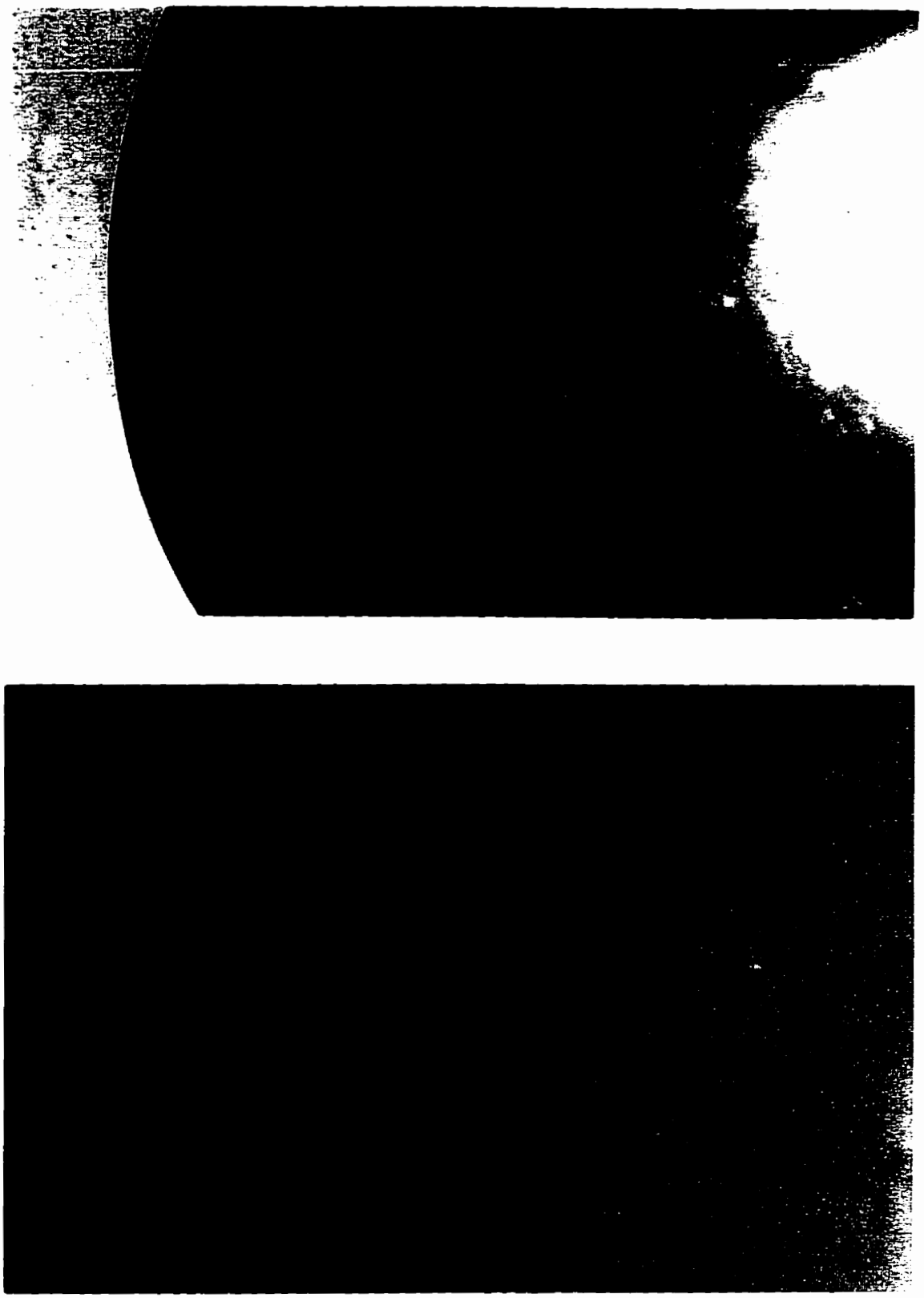
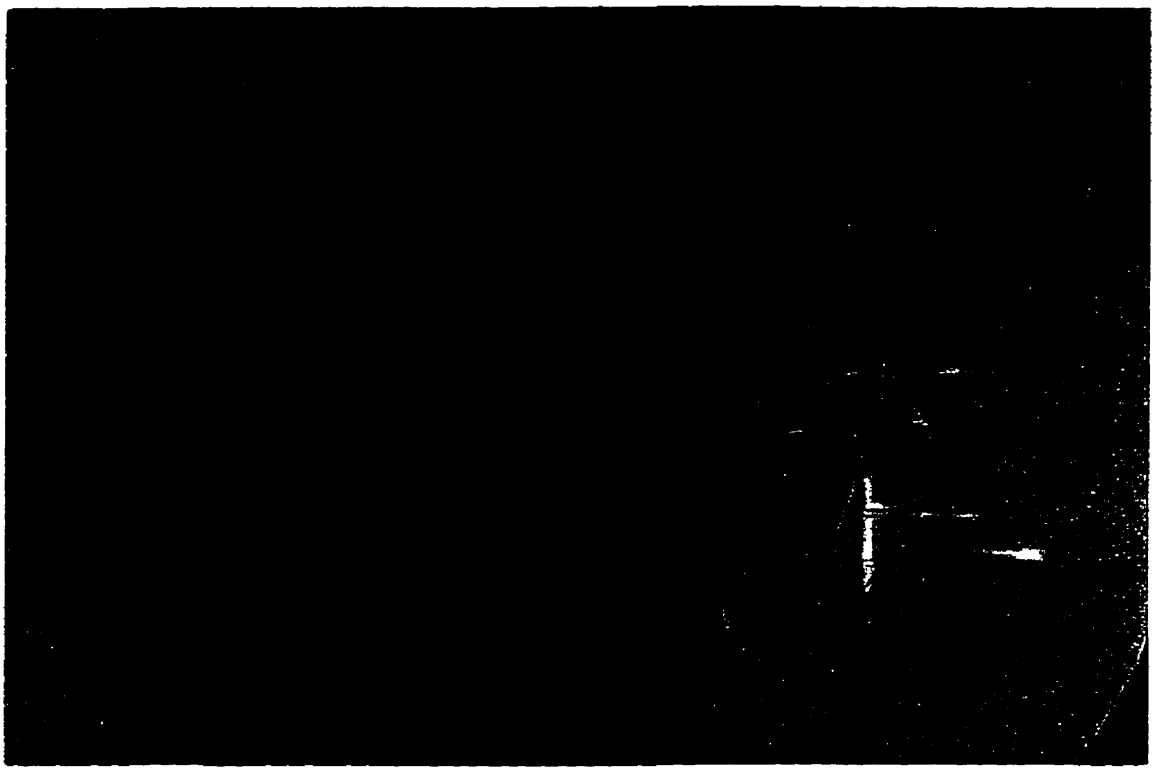


Figure 32. Crystals obtained under the conditions of (*a*) and (*b*) in Table 5.

(*c*)



(*d*)

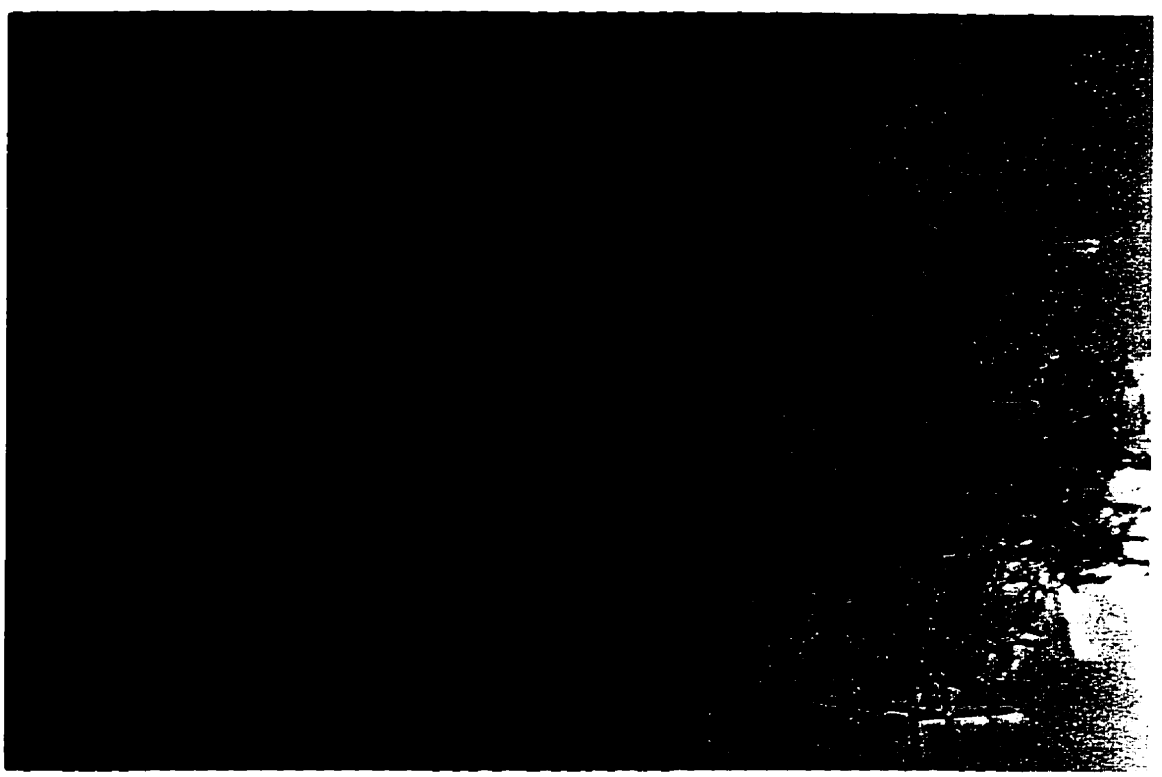
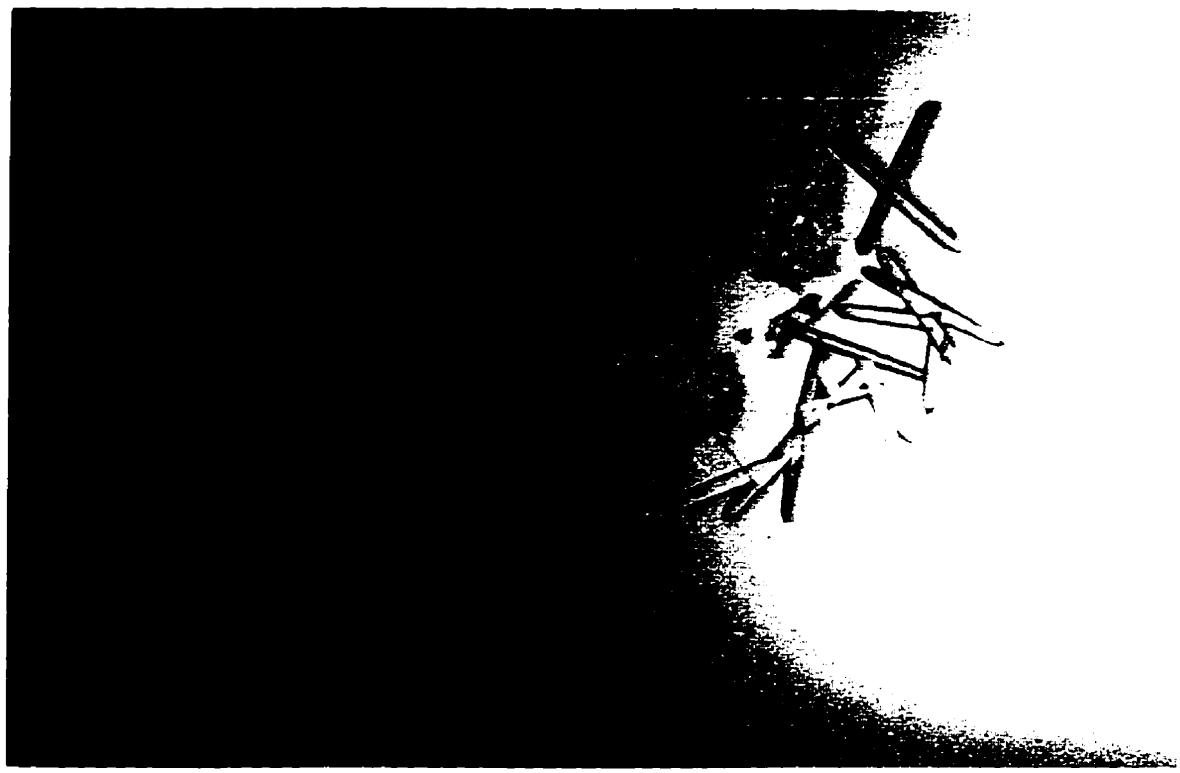


Figure 32. Crystals obtained under the conditions of (*c*) and (*d*) in Table 5.

(*e*)



(*f*)

Figure 32. Crystals obtained under the conditions of (*e*) and (*f*) in Table 5.

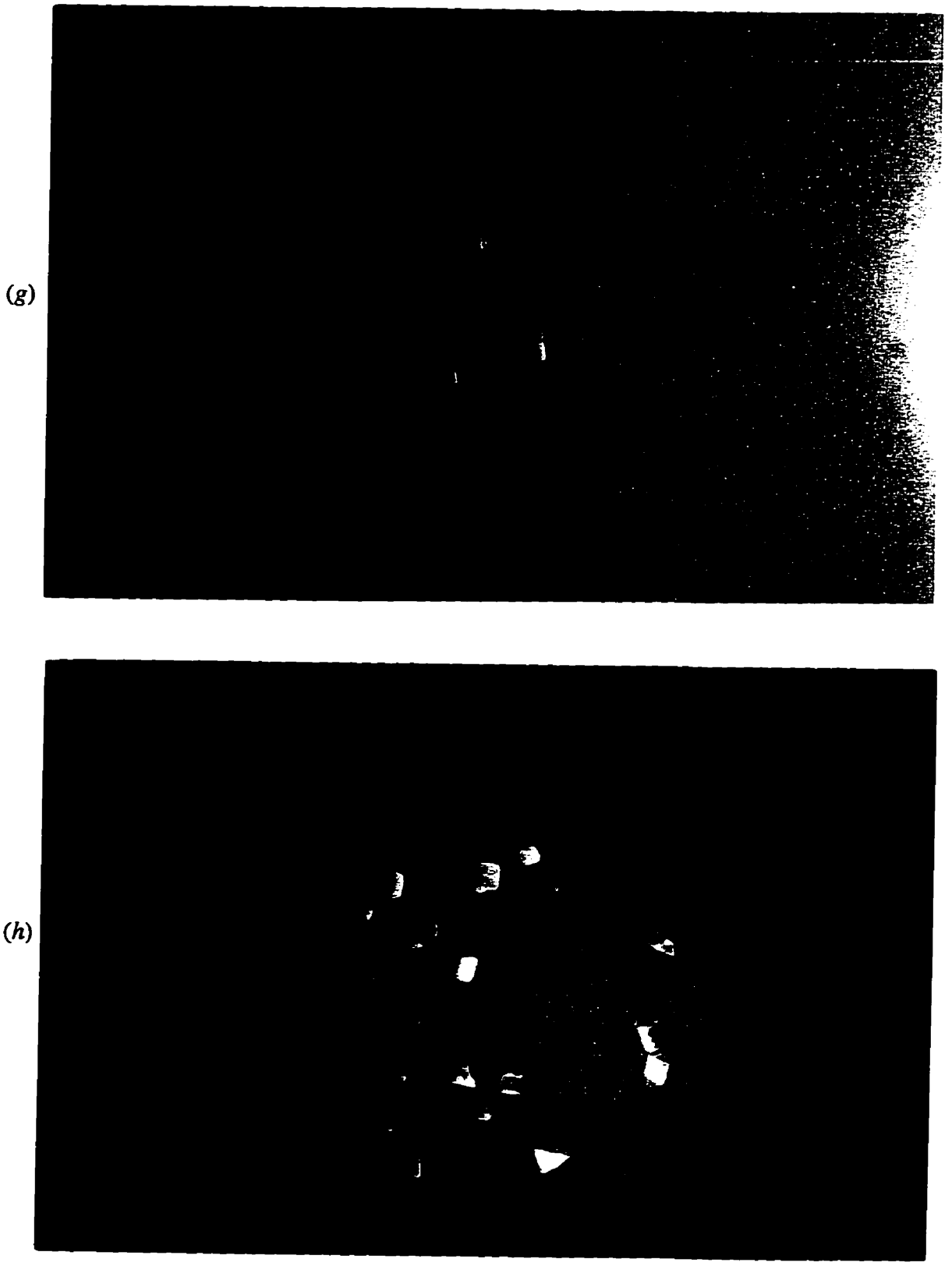


Figure 32. Crystals obtained under the conditions of (*g*) and (*h*) in Table 5.

(*i*)

(*j*)

Figure 32. Crystals obtained under the conditions of (*i*) and (*j*) in Table 5.

3.2.3. Special Techniques

The following special techniques were consistently tried in crystallizing experiments to improve the crystal quality.

Changing the concentration of the reservoir

If crystals are small and stop growing because the equilibrium between the droplet and reservoir has been reached, the concentration of the precipitating agent in the reservoir may be increased in order to establish a new equilibrium. This increase can be accomplished by gently rotating the cover slip, lifting it from one side, removing the reservoir contents entirely using a pipet, and replacing it with a more concentrated solution. The grease rim may be mended if required and the cover slip resealed to the reservoir. This procedure must be completed as fast as possible to prevent the droplet from drying out, especially when the volume of the droplet is small. The technique has been shown to be effective in increasing the crystal size.

Streak seeding

One of the most frustrating aspects in crystallization is that often one can obtain small crystals, but can not increase the size of the crystals by simply changing the concentrations of the components. The streak seeding method may be employed under such circumstances.

A segment of clean, thick human hair about 5 cm long is prepared for use as a simple streak seeding tool. Both the cover slips of the seeding droplet and the droplet in which there exist many microcrystals are gently uncovered. The hair is drawn along the surface

of the droplet containing microcrystals, dipped into the reservoir to remove excess crystals, and then streaked along the middle of the seeding droplet. Both droplets are then sealed against the reservoirs as soon as possible.

Some difficulties in performing streak seeding are:

(*i*) determination of the degree of supersaturation for seeding is not easy.

The solution to be seeded should be only slightly undersaturated so that controlled, slow growth will occur. This is usually done by simultaneously reducing the MPD and DNA concentrations in the droplet, and MPD concentration in the reservoir to ensure that the desired state of supersaturation will be reached more slowly. Normally, the seeding solutions are allowed to equilibrate for 3 - 5 days against the reservoir before the subsequent seeding. However, if the concentration is still too low, the seed crystals dissolve; if the concentration is too high, numerous additional small crystals appear.

(*ii*) The number of nuclei induced is difficult to control. Seeding solutions containing too many seeds yield showers of microcrystals.

Feeding[39]

When crystallization has reached equilibrium, a small aliquot of DNA solution is fed to the droplet. The concentration of DNA used for feeding is the same as that used for setup. A volume of 1 - 2 μL works well.

Feeding is done by pipetting 1 - 2 μL "DNA stock 2" and uncovering the cover slip of the droplet to be fed. The feeding DNA solution is gently expelled from the pipet and is kept on the outside of the pipet tip. The tip is then lowered close to the droplet to be

fed and the feeding solution is allowed to slide down the tip into the droplet. This must be done very carefully never allowing the pipet tip to touch the droplet being fed, otherwise the droplet may experience a mechanical shock and this perturbation may adversely affect crystal growth or even worse existing crystals may be destroyed. Under most circumstances, crystals grow larger by feeding. The feeding technique has been shown to be more successful than the other special techniques mentioned above.

3.3 X-ray Diffraction Analysis

The crystal to be tested on the diffractometer was mounted into a capillary by the following procedure:

(*1*) Thin-walled capillaries of 0.2 to 1.0 mm (Charles Supper Company), epoxy resin (MC LePage Limitee), paper wicks (Hampton Research, CAT No. 100-9271), a sharp razor blade, a goniometer head, a brass pin, some clean glass slides and some sheets of Kimwipes (Kimberly-Clark) were assembled near the microscope.

(*2*) The size of the crystal was assessed under the microscope and a capillary about 0.2 mm wider in diameter was selected. The sealed end of the capillary was removed with a sharp razor blade. The jagged end was then trimmed to be relatively even to prevent damage to the crystal when drawing it into the capillary. The capillary was connected to a tygon tubing that is terminated with a large plastic pipet tip as a mouthpiece.

(*3*) The cover slip of the droplet was removed and 5 - 15μL of its reservoir was added to the droplet to prevent the crystal from dehydrating during the manipulation.

(*4*) The cover slip was placed under the microscope. The position of the open end of the capillary was adjusted to approach the crystal and a gentle mouth suction was given to pick the crystal up into the capillary.

(*5*) The open end of the capillary was pointed against a glass slide and at the same time the excess solution in the capillary was gently blown out by mouth drop by drop. If necessary, a paper wick was stuck into the capillary end to

suck out the excess solution.

It required some practice to get the knack of the removing excess solution operation. The liquid tended to stick at first because of surface tension, and then it might come out in a rush, requiring a little back pressure. This frequently required blowing liquid gently back and forth. Instead of mouth pressure, a syringe could be used to blow out the liquid. However, it was harder to control the syringe. When using a syringe, a paper wick was always put at the exit end of capillary because the wick can not only absorb the excess liquid but also prevent the crystal from being blown out the end.

(*6*) A clean and dry paper wick was dipped into the reservoir and then taken out. The soaked tip was squeeze-dried with a Kimwipe to reduce its absorption capability. The damp wick was inserted into the capillary to suck out a small amount of solution. The wick was pulled out and squeeze-dried again with a Kimwipe. These squeezing and wicking procedures were repeated several times and the amount of liquid left between the crystal and the capillary was adjusted. If there insufficient solution present, the crystal may be dehydrated during data collection; if it is too much, the crystal may move. One drop of solution was left far from the crystal to keep the vapor equilibrium in the capillary.

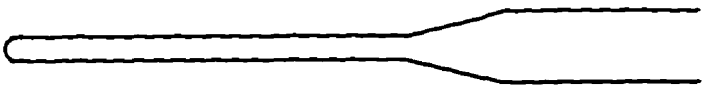
(*7*) The capillary was then cut at the other end within 1 cm of the crystal.

(*8*) Both ends of the capillary were sealed with epoxy resin.

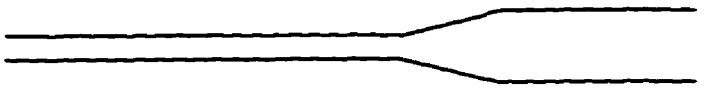
(*9*) The sealed capillary was mounted onto a pin and fixed with epoxy resin. Its height was adjusted on a goniometer head before the epoxy resin was com-

pletely hardened.

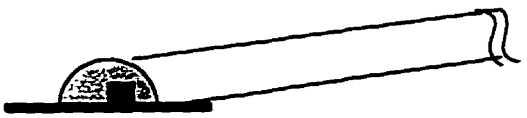
The whole mounting procedure is described in Figure 33.



a capillary of proper diameter



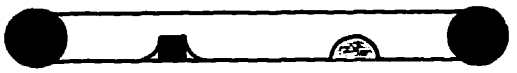
cut end of the capillary



the open end of the capillary approaching the crystal

the crystal with excess liquid in the capillary

wicking away excess mother liquor



one drop of mother liquor in the capillary and ends sealed with epoxy resin

Figure 33. Crystal mounting and drying stages illustrated.

3.3.1 Tetragonal and Monoclinic Crystal Cell Parameters

Two needle-like crystals of dimensions of $0.3 \times 0.04 \times 0.04$ mm^3 crystallized from the condition *e* of Table 5 were tested on the synchrotron at the Brookhaven National Laboratory in New York. Under the intense radiation of the synchrotron, the decay of the crystals was significant. The effective data collection time was about 600 seconds. Therefore only limited information was obtained.

The cell dimensions of the two crystals are (*i*) a = b = 77 Å, c = 40 Å, $\alpha = \beta = \gamma =$ 90°, for the tetragonal cell, and (*ii*) a = 93.56 Å, b = 16.82 Å, c = 48.15 Å, α= 90°, β = 113°, γ = 90°, for the monoclinic cell.

For the tetragonal cell a = b = 77 Å, c = 40 Å, $\alpha = \beta = \gamma =$ 90°, and a volume per base-pair of about 1,225 Å^3 (average value obtained from some orthorhombic cells)[102], a DNA dodecamer duplex will have a volume of $1,225 \times 12 = 14,700$ Å^3. Using these values, the number of molecules in a unit cell is $\frac{77 \times 77 \times 40}{14,700} = 16$. Since most of the chiral tetragonal space groups have eight symmetry equivalent positions, the sixteen duplex molecules are likely to be accommodated in the cell as pairs of duplexes per asymmetric unit.

3.3.2 Trigonal Crystal

3.3.2.1 Data Collection

Two crystals A (crystallized from 0.73 mM DNA, 7.1 mM $ZnCl_2$, 28.4 mM NaCl, 4.26 mM spermine, 14.2 mM NaCac, 2.8% MPD, pH 7.0, and reservoir 17.5% MPD) and B (crystallized from 0.87 mM DNA, 8.5 mM $ZnCl_2$, 34 mM NaCl, 5.1 mM spermine, 17 mM NaCac, 1.7% MPD, pH 7.45, and reservoir 10% MPD) were tested. A lot of preliminary crystallographic information was obtained from crystal A and crystal B yielded the whole data set. The the cell dimensions for crystal A were calculated from 25 reflections including some Friedel mates and with 2θ well distributed from 6.8° to 14.1°. The dimensions are a = b = 46.167 (38) Å, c = 101.064 (42) Å, $\alpha = \beta = 90.0^{\circ}$, $\gamma = 120.0^{\circ}$, and volume = 186,547 (311) $Å^3$.

X-ray data collection was performed on a RIGAKU AFC6S four-circle diffractometer at room temperature using the crystal in Figure 32(*i*) which had dimensions of 0.7 × 0.38 × 0.38 mm^3. The crystal was mounted and centered manually with the aid of an optical microscope. The power settings for collection were 2 kW (50 kV and 40 mA) with copper as the target material (λ = 1.54 Å).

From a photograph taken with an exposure time of 1 hour, the coordinates of the reflection spots were measured manually and indices (hkl) were then calculated by the computer. The cell dimensions were determined by using three shortest non-coplanar reciprocal vectors and the orientation matrix (the orientation of the reciprocal cell relative to the diffractometer axial system) was calculated.

A random "zig-zag" *search* routine was then executed automatically in order to obtain additional reflections with higher 2θ angles. The search scan type was an ω scan

with 2θ between 10° and 21°. Figure 34 is an ω scan profile of the reflection which was later selected as the standard reflection for monitoring crystal decay during data collection. The quality of the crystal was very good as demonstrated by the narrow peak. Because the initial cell obtained at the low 2θ angles was not very precise and therefore some peaks found by the search were poorly centered, frequently the refinement of those peaks failed in the *refine* routine. Therefore, the *menu center* routine was used to successfully refine the angular settings of the peaks. As additional high 2θ angle reflections became available, a better cell was calculated using fifteen reflections with 2θ between 10.4° and 20.8°. These included some Friedel mates and were well distributed in the reciprocal space. The standard deviations of the cell dimensions were minimized by least-squares refinement of the crystal orientation matrix and the lattice constants with hexagonal symmetry constrains. The unit cell defined was hexagonal, with the dimensions of a = b = 47.313 (49) Å, c = 102.098 (63) Å, $\alpha = \beta = 90.0°$, $\gamma = 120.0°$, and volume = 197,925 (418) $Å^3$. The orientation matrix that yielded this cell was used to locate reflections during data collection.

During the *search* routine, a serious peak overlap problem was found when the reflections perpendicular the c axis, such as (006), (007), and *etc.* were scanned. The (006) peak was well centered and properly collected. However, the scan range for (007) overlapped with peak (006) because the two peak centers were so close due to the very small reciprocal cell dimensions of this hexagonal system. Experimentally, the scan of (007) actually included half the profile of (006) and was flagged as asymmetric even though theoretically the intensity of (007) was zero. If the cell were hexagonal, this would be an inherently unsolvable problem. Before starting any data collection, this problem had to be solved. Otherwise, the data collection would be useless for solving the structure if the

```
Profile for an input hkl

Profile is for reflection index   h =   -2     k =   -2     l =    9

2theta =    10.797    omega =     5.399    chi =    56.814     phi =   -89.021
Scan type is omega
Step time (sec.) is   0.5

Integrated intensity is        4618.   (    68.)
Full width at half height =   0.229  degrees   omega
Peak offset =   0.057  degrees   omega
Peak width  =   0.480  degrees   omega
Full scale for the plot =     379.  counts/step  or        758.  cps
```

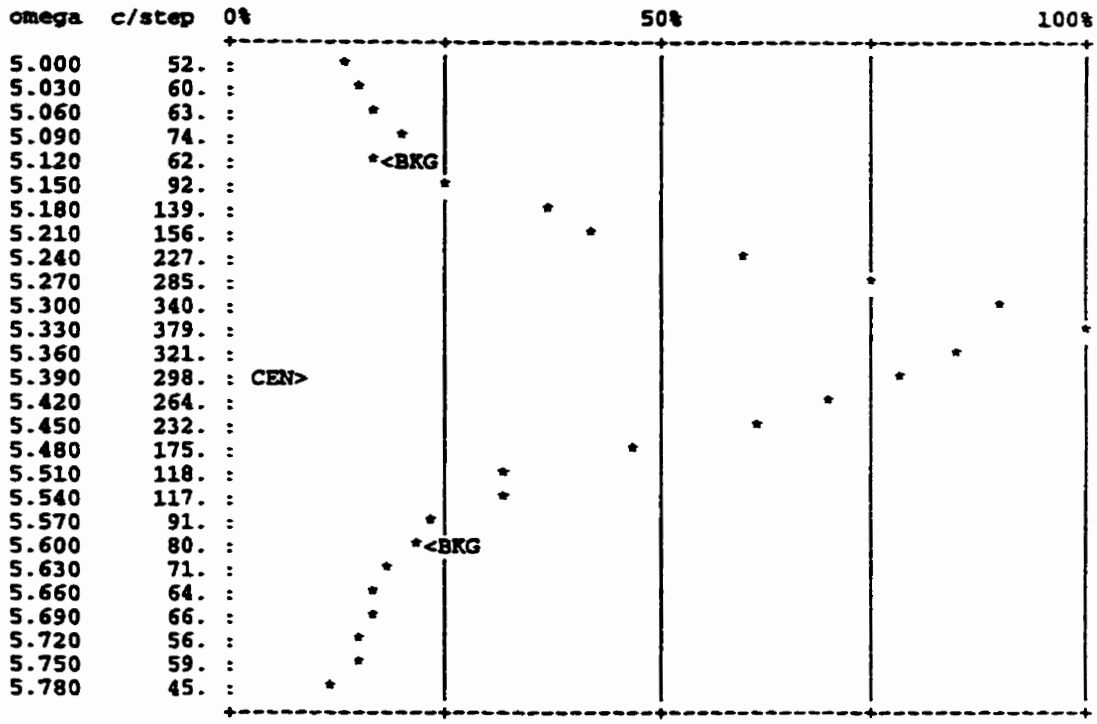


Figure 34. ω-scan of the standard reflection $(\bar{2}\,\bar{2}\,9)$.

detector could not separate those kinds of peaks. In another words, it must be very clear which was which.

The *Laue* routine was done to check the intensity symmetry in reciprocal lattice. The 6-fold rotation axis parallel to c which is a Laue symmetry element for hexagonal system was checked and the result showed that the only a 3-fold, not a 6-fold symmetry existed, indicating the crystal system should not be described as hexagonal.

In order to derive the correct cell system from the intensity symmetries, ten strong reflections, together with the corresponding equivalent reflections were scanned by inputting reflection indices to check the 3-fold inversion axis (expressed as $\bar{3}$-axis) and a mirror plane which are the Laue symmetry elements for the trigonal system, the only other possibility consistent with the cell dimensions. For each (hkl) reflection, there were six equivalent reflections related by the $\bar{3}$-axis, and six related by the mirror plane.

$\bar{3}$-axis	(h, k, l), ($\bar{h}$, $\bar{k}$, $\bar{l}$), (k, $\bar{h}$+$\bar{k}$, l), ($\bar{k}$, h+k, $\bar{l}$), ($\bar{h}$+$\bar{k}$, h, l), (h+k, $\bar{h}$, $\bar{l}$)
mirror plane	(k, h, $\bar{l}$), ($\bar{k}$, $\bar{h}$, l), (h, $\bar{h}$+$\bar{k}$, $\bar{l}$), ($\bar{h}$, h+k, $\bar{l}$), ($\bar{h}$+$\bar{k}$, k, $\bar{l}$), (h+k, $\bar{k}$, l)

By comparing the integrated intensities, it was concluded that both the $\bar{3}$-axis and the mirror plane existed. Therefore the Laue symmetry was concluded to be R$\bar{3}$m and the crystal belonged to the trigonal class. Once the unit cell was transformed from the hexagonal to the trigonal, the overlap problem did not exist any more. The (006) in the hexagonal cell was indexed (40$\bar{2}$) and the (007) in the hexagonal cell was not a peak position in the trigonal cell.

. In order to minimize the data collection time, data were collected from a unique portion of the reciprocal sphere (an quadrant) with the limits $0 \leq h$, $0 \leq k$, and no restriction on l, from 5^{o} to 30^{o} on 2θ. Data were collected in shells starting from high to low res-

olution shells. The collection was carried out with ω-scans and scan widths of 0.63 + (0.14) tanθ degrees. The base scan width of 0.63° was calculated based on the peak widths found from the initial search reflections, and the (0.14)tanθ term compensated for peak broadening as a function of θ. The scan speed was 4 degrees/minute. The F/ σ(F) ratio was 10.0 and the weak reflections could be rescanned until the F/ σ(F) ratio was obtained or up to five times in order to obtain better counting statistics, which was analogous to scanning the weak reflections at a slow speed. One standard ($\bar{2}$, $\bar{2}$, 9) was measured after every 200th reflection throughout the whole data collection procedure to monitor the crystal decay. It showed that crystal B was amazingly robust compared to most macromolecular crystals. The entire data collection was completed on one crystal over a period of eight days.

3.3.2.2 Data Reduction

teXsan (95)[103] was used to process the data. Some important statistical data derived from the intensities are:

		Percentage out of total refl.
Number of general reflections (excluding standards)	2514	100%
Number of reflections between 0 and 1 σ(I)	713	28%
Number of reflections between 1 and 2 σ(I)	186	7.3%
Number of reflections between 2 and 3 σ(I)	127	5.1%
Number of reflections greater than 3 σ(I)	464	19%
Number of reflections with negative intensity	1024	41%

The Laue class is $R\bar{3}m$. The crystal system is trigonal. The unit cell is R-centered. The suggested space group number is #155 and space group symbol is R32. There are 18

symmetry equivalent positions in the unit cell. Therefore the possible number of molecules in one unit cell is either 9 or 18. The approximate volume per base-pair[102] in trigonal cell is about 1,650 $Å^3$ and a dodecamer has a volume of $1,650 \times 12 = 19,800Å^3$. The volume of a unit cell is known to be 186,547 $Å^3$. The maximum number of molecules in a unit cell is $\frac{186,547}{19,800} = 9$, *i.e.* there are 9 molecules in a unit cell and one strand, or alternatively six base-pairs in an asymmetric unit. One molecule occupies two adjacent asymmetric units related by 2-fold symmetry. The average change in intensity for the standard reflection (decay) is -22.3%. No absorption and Lorentz polarization corrections were applied.

3.4 Discussion

The reflection overlap problem in the hexagonal cell arose from the fact that a trigonal system is generally described in terms of the same unit cell as that for a hexagonal system. The trigonal system is characterized by the presence of an inverse triad which is parallel to the z axis of the unit cell and the x and y axes are 120° apart in the plane perpendicular to the z axis. While a trigonal system can be expressed in the more common hexagonal axial system, a trigonal system has its own unit cell description. This is the unit cell illustrated in Figure 35(*a*); it has its x, y, and z axes equally inclined to the triad with a = b = c, and $\alpha = \beta = \gamma < 120°$. This kind of cell is called rhombohedral. This is the only conventional unit cell in crystallography in which all the symmetry axes are necessarily non-parallel to its reference axes. Because of its inconvenient geometry, the rhombohedral unit cell is rarely used for the description of trigonal crystals and the hexagonal cell is generally preferred. The relationship between a hexagonal unit cell and a rhombohedral unit cell is described in Figure 35(*b*).

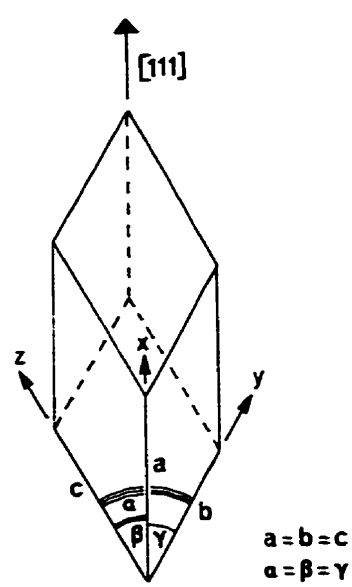


Figure 35. (*a*) A rhombohedral unit cell: a = b = c, $\alpha = \beta = \gamma \neq 90°, 120°$. From Ref. 99.

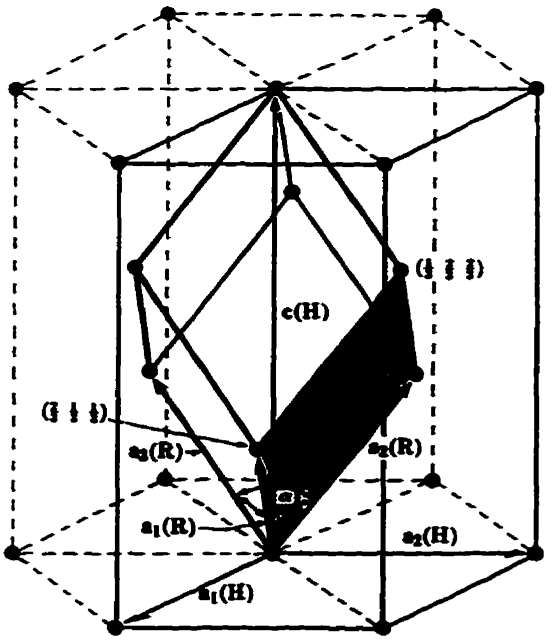


Figure 35. (*b*) Rhombohedral and hexagonal unit cells in a rhombohedral lattice. From Ref. 104.

The monoclinic cell with a = 93.56 Å, b = 16.82 Å, c = 48.15 Å, $\alpha = 90^{\circ}$, $\beta = 113^{\circ}$, $\gamma = 90^{\circ}$, is less reliable due to the fact that the shortest dimension of this cell is only 16.82 Å which is shorter than the shortest dimension of a dodecamer molecule, *viz.* 20 Å. This can not be a duplex DNA dodecamer cell because a duplexes molecule will not fit. Furthermore, little confidence is placed in this cell as the diffraction from the crystal was extremely weak, even under synchrotron collection conditions. The likelihood of error in determining the cell dimensions from the limited number of reflection is very high. No further consideration was given to this cell.

DNA crystals are weaker scatterers than most other kinds of macromolecular crystals and B-form DNA is the weakest among all forms of DNA. Furthermore, some B-form DNA crystals do not diffract at all, even though these crystals often have uniform external shapes. The intensity statistics from the data collection show that the present DNA crystal

is a very weak scatterer, a characteristic of B-DNA. Based on these results and the crystallization condition, it is deduced that the DNA dodecamer d(CGGTGGCCACCG) crystal is B-form DNA.

Zinc is one of the most important heavy metals in biological systems. The total concentration of zinc in the body is 3 g/70 kg. Tightly bound zinc ions play important structural and functional roles in a number of enzymes of nucleotide and nucleic acid metabolism[23]. This DNA dodecamer d(CGGTGGCCACCG) appears to be the first oligonucleotide crystallized using $ZnCl_2$ as salt. The mechanism of Zn^{2+} facilitating DNA crystallization, *i.e.* stabilizing DNA helices, is unclear at this stage. Mg^{2+} ions can stabilize the DNA helical structure by cross linking phosphate groups. Zn^{2+} ions do not behave in such a clear-cut manner; they selectively stabilize some structures, while destabilizing others, depending on the amount of electrolyte present and the sequence of DNA[105].

Metal ions exhibit different effects because they are able to bind to both phosphates and heterocyclic bases. In solution studies of the series Mg^{2+}, Co^{2+}, Ni^{2+}, Mn^{2+}, Zn^{2+}, Cd^{2+}, Cu^{2+}, the affinity for base complexation relative to phosphate binding increases from the left to right due to each metal ion possessing a different relative affinity for these two sites[105].

Even though there are only two crystal structures for zinc-nucleic acid complexes available, Zn(5'-IMP) and Zn(5'-CMP), they show a clear view of the binding sites of zinc atoms. The x-ray investigation of Zn(5'-IMP)[106] has shown that the crystal contains polymeric chains of composition $[Zn(5'\text{-IMP})]_n$ in which each zinc atom is bonded to the N(7) position on the base and to phosphate oxygens from three other 5'-IMP groups (Fig-

ure 36). This structure is unusual in that there are no water molecules associated with the metal ion.

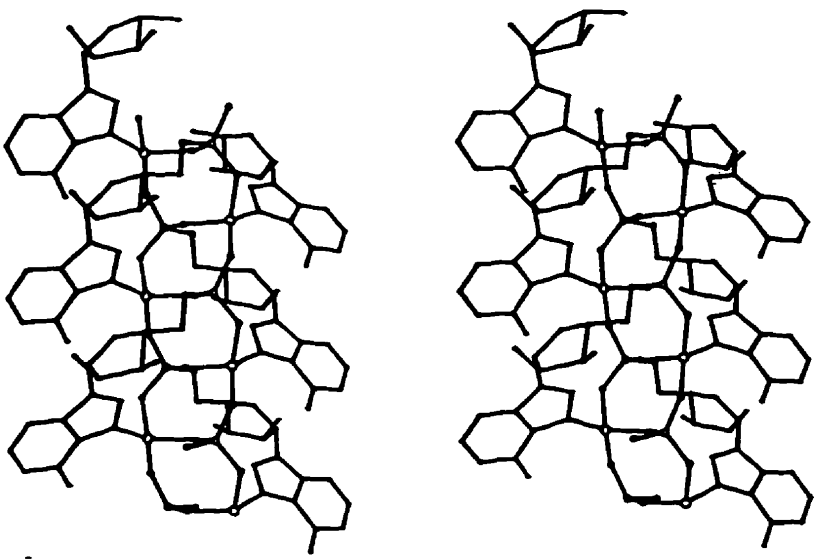


Figure 36. Stereoscopic pair of drawing a fragment of the chain structure of $[Zn(5'\text{-IMP})]_n$ viewed perpendicular to the screw diad axis running centrally along the chain. (The largest circles represent zinc atoms). From Ref. 106.

The complex Zn(5'-CMP) is a two-dimensional polymer. The zinc atoms are tetrahedrally coordinated to N(3) of the pyrimidine, to two phosphate oxygen atoms and to one water molecule. The zinc atoms also form weak intramolecular interactions with O(2) of the pyrimidine (Figure 37)[107]. Both results show that Zn^{2+} ions possess fairly strong affinities for both the nucleotide bases and the phosphates.

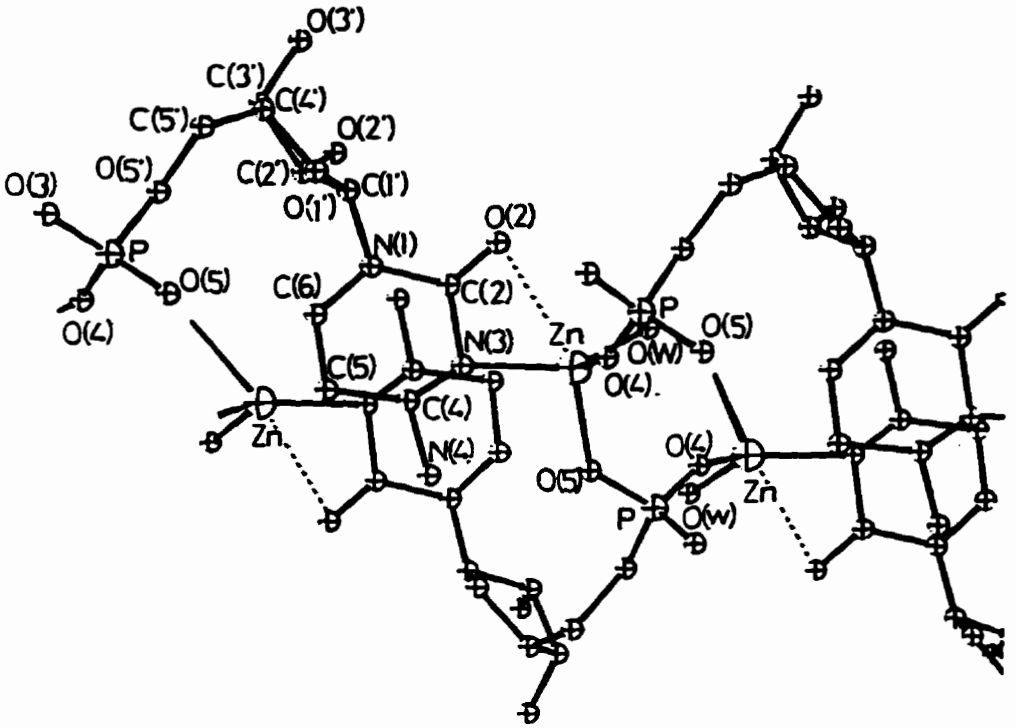


Figure 37. Principal features of the structure of the Zn(5'-CMP) complex, indicating the zinc atom coordination and partial base stacking. Broken lines represent weak interaction between the zinc atom and O(2) of the pyrimidine. From Ref. 107.

3.5 Conclusion

The self-complementary DNA dodecamer d(CGGTGGCCACCG) was crystallized using $MgCl_2$, $BaCl_2$ and $ZnCl_2$ as salts. The cell dimensions crystallized with $BaCl_2$ are a = b = 77 Å, c = 40 Å, $\alpha = \beta = \gamma = 90^o$. The unit cell is tetragonal with sixteen molecules in each cell.

The crystal which finally yielded the complete data set was crystallized using $ZnCl_2$. There is no published work reporting the crystallization of an oligonucleotide with Zn^{2+} as a counterion. Therefore, this appears to be the first example of such a crystallization. The crystal system is trigonal. The unit cell is R-centered and expressed as a hexagonal cell, with the dimensions of a = b = 47.313 (49) Å, c = 102.098 (63) Å, $\alpha = \beta = 90.0^o$, $\gamma = 120.0^o$, and volume = 197,925 (418) $Å^3$. The Laue class is $R\bar{3}m$. The suggested space group number is #155 and space group symbol is R32. There are 9 molecules in the unit cell. The solution of the structure is currently being attempted in this laboratory.

Appendix A:

Processing of UV Experimental Data with Mathcad 6.0

1. Processing of UV Experimental Data

Units: j ≡ joule mol ≡ mole $kj \equiv 10^3 \cdot j$ kcal ≡ 4.18·kj

degC ≡ 1

Constants Gas constant $R \equiv 1.987 \cdot 10^{-3} \cdot kcal \cdot mol^{-1} \cdot K^{-1}$

A matrix **AA** is constructed with the first column being the temperature readings and the second column being the corresponding absorbance readings.

$$AA = \begin{bmatrix} 17.5 & 1.1573 \\ 29.5 & 1.1744 \\ 38 & 1.1971 \\ 45.5 & 1.238 \\ 51.0 & 1.2629 \\ 57.5 & 1.2931 \\ 66 & 1.3073 \\ 72.5 & 1.3166 \\ 80.5 & 1.318 \end{bmatrix}$$

Define a range variable **i**

$i = 0, 1 .. 8$

This definition indicates that i takes on the values 0, 1, 2 .. 8.

Extract each column from the matrix **AA** as **T1** and **A1**, respectively.

$A1 = AA^{\langle 1 \rangle}$ $T1 = AA^{\langle 0 \rangle}$

The **genfit** function solves a nonlinear problem by iteration. The **genfit** function needs a good choice of the function to be fitted, **f (x, u)**, and a vector of guess values, **vg**, as an argument. **F (x, u)** is a function that returns an n + 1 element vector containing f and its partial derivatives with respect to its n parameters. The function to be fitted is chosen as tanh(x) due to the sigmoidal feature of the melting curves. Four parameters u_0, u_1, u_2, and u_3, respectively, are defined to make the best fit.

$$F(x,u) \equiv \begin{bmatrix} u_3 \cdot \tanh(u_1 \cdot x + u_0) + u_2 \\ u_3 \cdot \operatorname{sech}(u_1 \cdot x + u_0) \cdot \operatorname{sech}(u_1 \cdot x + u_0) + u_2 \\ u_3 \cdot x \cdot \operatorname{sech}(u_1 \cdot x + u_0) \cdot \operatorname{sech}(u_1 \cdot x + u_0) + u_2 \\ 1 \\ \tanh(u_1 \cdot x + u_0) + u_2 \end{bmatrix}$$

<----The first element contains the function to be fitted. There are a total of four unknown parameters, u_0 to u_3.

<-----The others contain the partial derivatives.

Define the range variable **r** for the fitting function. Choose 10°C above and below the max. and min. experimental points respectively. 10°C is the first value taken by the range variable **r**. The step size is two degrees and 90°C is the last value in the range. By defining the range, **r**, the experimental points are to be fitted from 10°C to 90°C by the fitting function.

$$r := 10, 12 .. 90$$

Make a vector of guess values of u_0 to u_3. u_0 is associated with the melting temperature (T_m); the higher the T_m, the lower the u_0, and u_0 should be always negative as T_m is always positive; u_1 is associated with the sharpness of the melting curve, the sharper the curve, the lower the u_1; u_2 is approximately the corresponding absorbance value at the T_m; and u_3 is associated with the total percent increase in absorbance; the higher the percentage increase, the larger the u_3.

$$vg \equiv \begin{bmatrix} -3 \\ 0.1 \\ 1.2 \\ 0.1 \end{bmatrix} \quad \begin{matrix} \text{<--- } u_0 \\ \text{<--- } u_1 \\ \text{<--- } u_2 \\ \text{<--- } u_3 \end{matrix}$$

Define temperature and absorbance readings with global definitions as **T2** and **A2**, respectively. Global definitions are exactly like local definitions except that they are evaluated before any local definitions.

$$T2 \equiv \begin{bmatrix} 17.5 \\ 29.5 \\ 38 \\ 45.5 \\ 51.0 \\ 57.5 \\ 66 \\ 72.5 \\ 80.5 \end{bmatrix} \qquad A2 \equiv \begin{bmatrix} 1.1573 \\ 1.1744 \\ 1.1971 \\ 1.238 \\ 1.2629 \\ 1.2931 \\ 1.3073 \\ 1.3166 \\ 1.318 \end{bmatrix}$$

Define a vector **P** containing the parameters that make the function of x, *i.e.* **F**(x,u) and the four parameters u_0, u_1, u_2, and u_3 best approximate the data in **T2** an **A2**.

$$P \equiv genfit(T2, A2, vg, F) \qquad g(r) \equiv F(r, P)_0$$

Display the calculated vector **P**.

$$P = \begin{bmatrix} -2.938 \\ 0.064 \\ 1.237 \\ 0.083 \end{bmatrix}$$

Define the absorbance values at temperatures of 0°C and 100°C (labeled t and s below) as **f1** and **f2**, respectively. **f1** and **f2** are used as the lower and upper limits to calculate the hairpin fraction and to normalize the absorbances.

$$t := 0$$

$$f1 := 1.237 + 0.083 \cdot \tanh(t \cdot 0.064 - 2.938) \qquad f1 = 1.154$$

$$s := 100$$

$$f2 := 1.237 + 0.083 \cdot \tanh(s \cdot 0.064 - 2.938) \qquad f2 = 1.32$$

Calculate the equilibrium constant **K**. Define **u** as the inverse of the temperature (Kelvin), **p** the logarithm of **K**, **S(r)** the normalized fitting function, and **A** the normalized absorbance. Then, display **A**, **u** , and **p** values.

$$K_i := \frac{A1_i - f1}{f2 - A1_i} \qquad u_i := \frac{1}{T1_i + 273} \qquad p_i := \ln(K_i)$$

$$S(r) := \frac{g(r)}{f2} \qquad A_i := \frac{A1_i}{f2}$$

A_i
0.877
0.89
0.907
0.938
0.957
0.98
0.991
0.998
0.999

u_i
$3.442 \cdot 10^{-3}$
$3.306 \cdot 10^{-3}$
$3.215 \cdot 10^{-3}$
$3.14 \cdot 10^{-3}$
$3.086 \cdot 10^{-3}$
$3.026 \cdot 10^{-3}$
$2.95 \cdot 10^{-3}$
$2.894 \cdot 10^{-3}$
$2.829 \cdot 10^{-3}$

p_i
-4.049
-1.987
-1.057
0.021
0.644
1.646
2.501
3.914
4.489

Define a least-squares line to fit the points on a van't Hoff's plot. Calculate the slope and intercept of the line.

$p_i := \text{slope}(u, p) \cdot u_i + \text{intercept}(u, p)$

$a := \text{slope}(u, p) \qquad a = -1.411 \cdot 10^4$

$b := \text{intercept}(u, p) \qquad b = 44.41$

Calculte the ΔH and ΔS by using the slope and intercept values above.

$$\Delta H := a \cdot (-1) \cdot R \qquad \Delta S := b \cdot R$$

Calculate the melting point Tm (Kelvin)and convert it into T_m (degrees Celsius).

$$Tm := \frac{\Delta H}{\Delta S} \qquad T_m := (Tm - 273.15) \cdot \text{degC}$$

Display the ΔH, ΔS, and T_m values.

$\Delta H = 28.035 \cdot \text{kcal} \cdot \text{mol}^{-1}$ $\qquad \Delta S = 88.099 \cdot \text{K}^{-1} \cdot \text{cal} \cdot \text{mol}^{-1}$ $\qquad T_m = 44.55 \cdot \text{degC}$

2. Plotting the melting curve and van't Hoff plot.

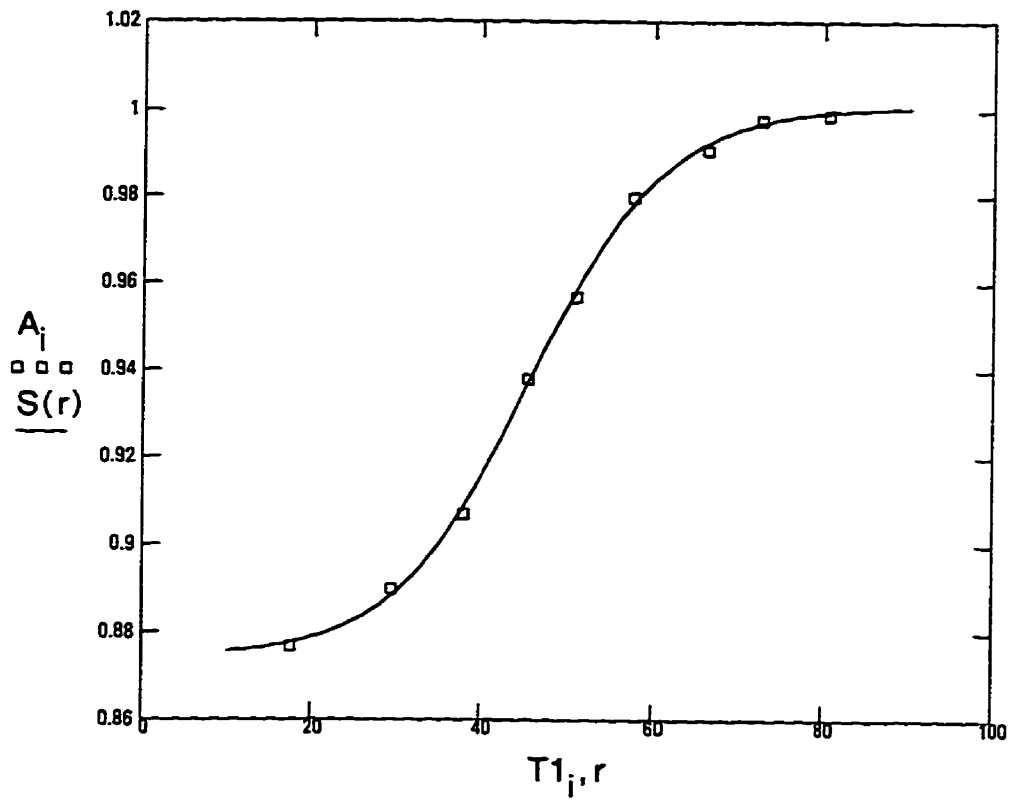


Figure A1. Melting curve of 'mismatched' DNA in $MgCl_2$, NaCac, MPD, and spermine solution.

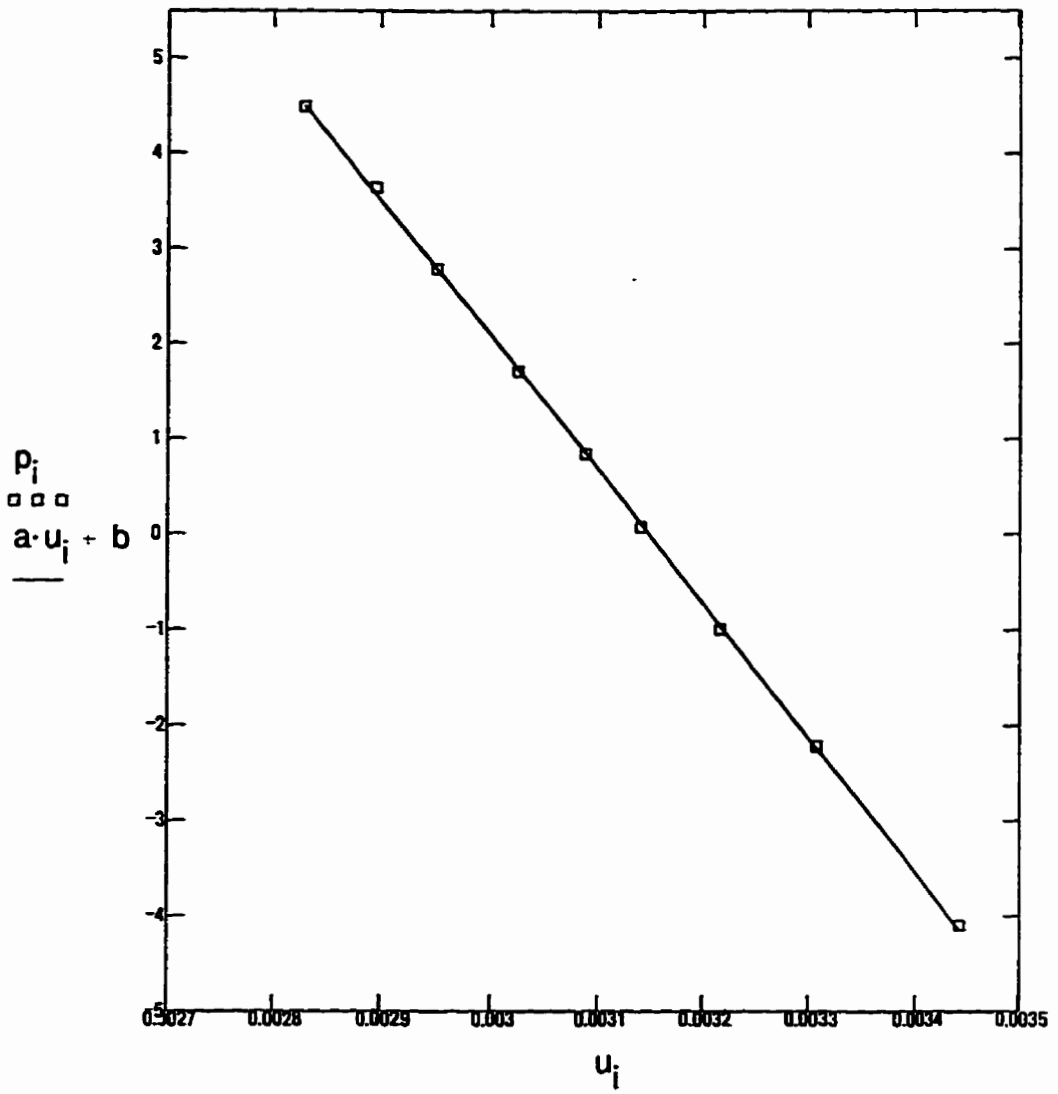


Figure A2. Van't Hoff plot for 'mismatched' DNA in $MgCl_2$, NaCac, MPD, and spermine solution.

Appendix B

Procedure for Processing CD Data and Producing Plots

1. Processing data

Example: 'mismatched' DNA in $MgCl_2$, NaCac, Spermine, and MPD solution—
reference actual Figure 20 in text.

Data from the CD experiment are obtained in terms of voltages that must be converted in order to be expressed as molar ellipticity values. The conversion factor is calculated from the equation $[\Theta] = 100 * \Theta/lC$.

$[\Theta] = 100 * \Theta/ lC$ = 100*voltage reading*41.24*sensitivity*$10^{-3}/lC$ (degree M/cm)

With the equipment sensitivity of 1, the pathlength 0.099 cm, and the molar concentration of the DNA sample 0.05 mM, the conversion factor is calculated as

$$\frac{100 \cdot 41.24 \cdot 1 \cdot 10^{-3}}{0.05 \cdot 10^{-3} \cdot 0.099} = 8.331 \cdot 10^{5}$$

The factor of 83.31 is used in subsequent equations to convert to molar ellipticity in units of 10^4 degree M/cm.

The final CD curve is to be plotted from 300 - 220 nm. There are 321 voltage readings in this range. The file **buf1** is the data file associated with the first buffer 'blank' scan. **buf1** is read into a matrix **BUFF1**. This is a 421x2 matrix with the original experimental data ranging from 340 - 220 nm. Since CD plots will only include the range 300 - 220 nm, a subset of **BUFF1** is required. **Buf1** is a 321x2 submatrix of **BUFF1** including data in the range 300 - 220 nm.

j is the number of points to be plotted; **x** is the wavelength in increments of 1/4 nm.

$$j := 0, 1 .. 320 \qquad x_j := 300 - \frac{j}{4}$$

The data file **buf1** is read into the matrix **BUFF1**, with the first column of the **BUFF1** being the voltage readings and the second being the corresponding wavelength readings.

BUFF1 := READPRN(buf1)

The voltage readings now in **BUFF1** are converted to molar ellipticity by the factor 83.31 and then input into matrix **BUF1** .

BUF1 := BUFF1 <1> 83.31

A submatrix **Buf1** containing the ellipticity values from 300 - 220 nm is obtained by taking the readings from the 160th to the 480th rows of column 1.

Buf1 := submatrix(BUF1, 160, 480, 0, 0)

Process the second buffer 'blank' scan in the same manner as above.

BUFF2 := READPRN(buf2)

BUF2 := BUFF2$^{<1>}$ 83.31

Buf2 := submatrix(BUF2, 160, 480, 0, 0)

It was found that the spectropolarimeter had different nonzero readings for each scan. This nonzero 'spectrum' was obtained by choosing 3 readings from 340 nm to 337 nm where it should show zero ellipticity calculated by taking the average of three ellipticity readings in that range. **con** is defined as the nonzero value for the first buffer scan.

$$con := \frac{BUFF1_{0,1} + BUFF1_{5,1} + BUFF1_{10,1}}{3} 83.31$$

The buffer baseline was found to drift during the experiment. **drift** is defined as the total change during the experiment and is assumed to be linearly dependent on time.

drift := BUF2 - BUF1

A1 is the data file associated with the spectrum obtained at 20°C. **READPRN(A1)** reads a structured data file **a1** and returns a matrix **A**.

A := READPRN(A1)

The data in **A** are reduced by the nonzero reading for this scan and then the data are converted into molar ellipticity by multiplying the conversion factor of 83.31. The buffer scan and buffer nonzero reading are subtracted.

$$aa := \left(83.31 \cdot A^{<1>} \cdot \frac{A_{0,1} + A_{5,1} + A_{10,1}}{3} 83.31\right) - (BUF1 - con)$$

Each scan was carried out from 340 - 220 nm. For plotting the data from 300 - 200 nm, a submatrix **a1** was defined which contains the data from 300 - 220 nm.

a1 := submatrix(aa, 160, 480, 0, 0)

There are a total of six scans (data **A1** through **F1**, corresponding to 20°C, 30°C, 38°C, 41°C, 55°C, and 70°C, respectively) and thus five scan intervals. At each interval, the baseline is considered to drift one-fifth of the total drift. Thus the following spectrum should subtracte from it an additional amount of drift/5 more than the previous scan.

B := READPRN(B1)

$$bb := \left(83.31 \cdot B^{<1>} \cdot \frac{B_{0,1} + B_{5,1} + B_{10,1}}{3} 83.31\right) - (BUF1 - con) - \frac{drift}{5}$$

b1 := submatrix(bb, 160, 480, 0, 0)

C := READPRN(C1)

$$cc := \left(83.31 \cdot C^{\langle 1 \rangle} \cdot BUF1 - \frac{C_{0,1} + C_{10,1} + C_{5,1}}{3} 83.31\right) + con - \frac{2 \cdot drift}{5}$$

c1 := submatrix(cc, 160, 480, 0, 0)

D := READPRN(D1)

$$dd := \left(34.71 \cdot D^{\langle 1 \rangle} \cdot BUF1 - \frac{D_{0,1} + D_{5,1} + D_{10,1}}{3} 83.31\right) + con - \frac{3 \cdot drift}{5}$$

d1 := submatrix(dd, 160, 480, 0, 0)

E := READPRN(E1)

$$ee := \left[83.31 \cdot E^{\langle 1 \rangle} \cdot BUF1 - \left(\frac{E_{0,1} + E_{5,1} + E_{10,1}}{3}\right) \cdot 83.31\right] + con - \frac{4 \cdot drift}{5}$$

e1 := submatrix(ee, 160, 480, 0, 0)

F := READPRN(F1)

$$ff := \left[83.31 \cdot F^{\langle 1 \rangle} \cdot BUF1 - \left(\frac{F_{0,1} + F_{5,1} + F_{10,1}}{3}\right) \cdot 83.31\right] + con - drift$$

f1 := submatrix(ff, 160, 480, 0, 0)

A range variable **k** is defined for the smoothing function. The smoothing function **supsmooth** is used to smooth the experimental spectra.

$k_j := 300 - x_j$

a := supsmooth(k, a1)

b := supsmooth(k, b1)

c := supsmooth(k, c1)

d := supsmooth(k, d1)

e := supsmooth(k, e1)

f := supsmooth(k, f1)

2. Plotting CD spectra

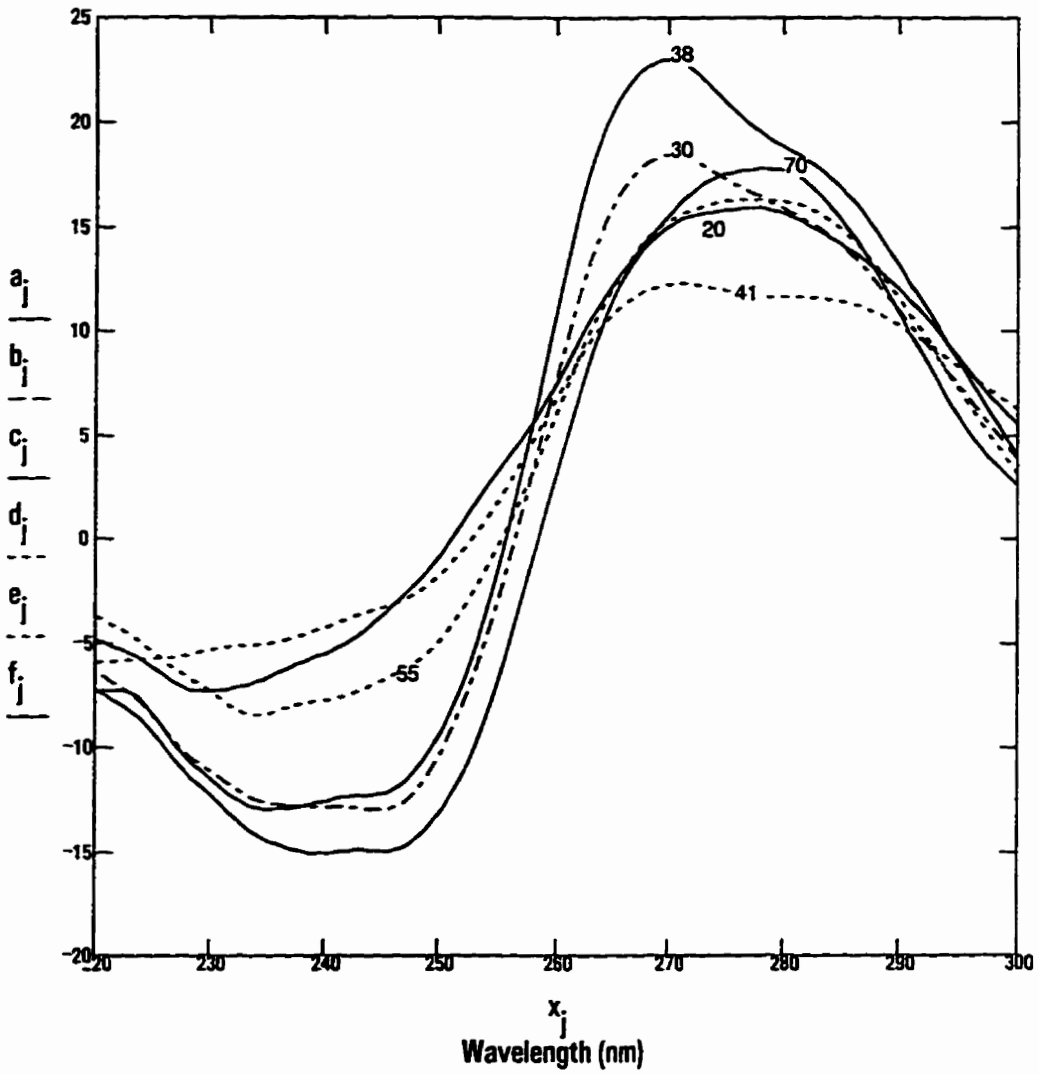


Figure B1. CD spectra of 'mismatched' DNA in $MgCl_2$, NaCac, spermine, and MPD solution. a: 20°C b: 30°C c: 38°C d: 41°C e: 55°C f: 70°C

References

1. Brown, T., Hunter, W. N., and Leonard, G. A. (1993) *Chem. Brit.*, June, 484-488.
2. Voet, D. and Voet, J. G. (1990) *Biochemistry*, John Wiley & Sons.
3. Chen, X., Santhana Mariappan, S. V., Catasti, P., Patliff, R., Moyzis, R. K., Laayoun, A., Smith, S. S., Morton Bradbury, E., and Gupta, G. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 5199-5203.
4. Sutherland, G. R. (1977) *Science* **197**, 265-266.
5. Nadel, Y., Weisman-Shomer, P., and Fry, M. (1995) *J. Biol. Chem.* **270**, 28970-28977.
6. Ashley, C. T. Jr. (1995) *Annu. Rev. Genet.* **29**, 703-728.
7. Modrich, P. (1987) *Annu. Rev. Biochem.* **56**, 435-466.
8. Neidle, S. (1994) *DNA Structure and Recognition*, 1st edition, Oxford Press, England.
9. Gray, D. M., Vaughan, M., Ratliff, R. L., and Hayes, F. N. (1980) *Nucleic Acids Res.* **8**, 3695-3707.
10. Thiele, D., Sarocchi, M.-T., Guschlbauer, W., Lezius, A., and Marck, C. (1978) *Mol. Biol. Rep.* **1**, 155-160.
11. Pilch, D. S. and Shafer, R. H. (1993) *J. Am. Chem. Soc.* **115**, 2565-2571.
12. Kouchakdjian, M., Li, B. F. L., Swann, P. F., and Patel, D. J. (1988) *J. Mol. Biol.* **202**, 139-155.
13. Langridge, R. and Rich, A. (1963) *Nature (London)* **198**, 725-728.
14. Gray, D. M. and Cui, T. (1984) *Nucleic Acids Res.* **12**, 7565-7580.
15. Bhattacharyya, A. and Lilley, D. M. J. (1989) *J. Mol. Biol.* **209**, 583-597.

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

17. Timsit, Y. and Moras, D. (1992) *Methods in Enzymol.* **211**, 409-429.

18. Werntges, H., Steger, G., Riesner, D., and Fritz, H. (1986) *Nucleic Acids Res.* **14**, 3773-3791.

19. Jurnak, F. and McPherson, A. (Eds.) (1985) *Biological Macromolecules and Assemblies*, Vol.2, Wiley, New York.

20. Shakked, Z. and Rabinovich, D. (1986) *Prog. Biophys. Mol. Biol.* **47**, 159-195.

21. Kennard, O. and Salisbury, S. A. (1993) *J. Biol. Chem.* **268**, 10701-10704.

22. Giegé, R., Lorber, B., and Theobald-Dietrich, A. (1994) *Acta Crystallogr.* **D50**, 339-350.

23. Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer-Verlag, New York, pp 220-349.

24. Wells, R. D., Collier, D. A., Hanvey, J. C., Shimizu, M., and Wohlrab, F. (1988) *FASEB J.* **2**, 2939-2949.

25. van de Ven, F. J. M. and Hilbers, C. W. (1988) *Eur. J. Biochem.* **178**, 1-38.

26. Ikuta, S., Chattopadhyaya, R., Ito, H., Dickerson, R. E., and Kearns, D. R. (1986) *Biochemistry* **25**, 4840-4849.

27. Wells, R. D. (1988) *J. Biol. Chem.* **263**, 1095-1098.

28. Hald, M., Pedersen, J. B., Stein, P. C., Kirpekar, F. K., and Jacobsen, J. P. (1995) *Nucleic Acids Res.* **23**, 4576-4582.

29. Tari, L. (1995) *Ph.D. Thesis*, University of Manitoba, Canada.

30. Chattopadhyaya, R., Grzeskowiak, K., and Dickerson, R. E. (1990) *Mol. Biol.* **211**, 189-211.

31. Varani, G. (1995) *Annu. Rev. Biophys. Biomol. Struct.* **24**, 379-404.

32. Mizuuchi, K., Mizuuchi, M., and Gellert, M. (1982) *J. Mol. Biol.* **156**, 229-243.

33. Panayotatos, N. and Wells, R. D. (1981) *Nature (London)* **289**, 466-470.

34. Avizonis, D. Z. and Kearns, D. R. (1995) *Nucleic Acids Res.* **23**, 1260-1268.

35. Marky, L. A., Blumenfeld, K. S., Kozlowski, S., Breslauer, and Breslauer, K. J. (1983) *Biopolymers* **22**, 1247-1251.

36. Rosenberg, M. and Court, D. (1979) *Annu. Rev. Genet.* **13**, 319-351.

37. Wells, R. D., Goodman, T. C., Hillen, W., Horn, G. T., Klein, R. D., Larson, J. E., Muller, U. R., Neundorf, S. K., Panayotatos, N., and Stirdivant, S. M. (1980) *Prog. Nucleic Acids Res. Mol. Biol.* **25**, 167-267.

38. Ducruix, A. and Giege, R. (Eds.) (1992) *Crystallization of Nucleic Acids and Proteins*, Oxford University Press, New York.

39. Wang, A. H. Jr., and Gao, Y. (1990) *Methods: A Companion to Methods in Enzymol.* **1**, 91-99.

40. Cudney, B. and Patel, S. (1994) *Acta Crystallogr.* **D50**, 414-423.

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

42. Cudney, B., Patel, S., and McPherson, A. (1994) *Acta Crystollogr.* **D50**, 479-483.

43. *Crystal Growth 101*, Hampton Research, Riverside, California.

44. Michel, H. (Ed.) (1991) *Crystallization of Membrane Proteins*, CRC Press, Inc., Boston.

45. McPherson, A. (1991) *Macromolecular Crystallization*, American Crystallographic Association Workshop, University of Toledo, Toledo, Ohio, July 21.

46. Carter, C. W. Jr. (1990) *Methods: A Companion to Methods in Enzymol.* **1**, 12-24.

47. Wang, A. H. Jr., Quigley, G. J., Kolpak, F. J., Crawford, J. L., van Boom, J. H., van der Marel, G. A., and Rich, A. (1979) *Nature (London)* **282**, 680-686.

48. Liaw, Y.-C., Gao, Y.-G., Robinson, H., Sheldrick, G. M., Sliedregt, L. A. J. M., van der Marel, G. A., van Boom, J. H., and Wang, A. H. Jr. (1990) *FEBS Lett.* **264**, 223-227.

49. Wang, A. H. Jr., Gessner, R. G., van der Marel, G. A., van Boom, J. H., and Rich, A. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 3611-3615.

50. Brown, T., Hunter, W., Kneale, G., and Kennard, O. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 2403-2406.

51. Brown, T., Leonard, G., Booth, E., and Chambers, J. (1989) *J. Mol. Biol.* **207**, 455-457.

52. Webster, G., Sanderson, M., Skelly, J., Neidle, S., Swann, P., Li, B., and Tickle, J. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 6693-6697.

53. Hunter, W., Brown, T., Kneale, G., Anand, N., Rabinovich, D., and Kennard, O. (1987) *J. Biol. Chem.* **262**, 9962-9970.

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

55. Hunter, W., Brown, T., and Kennard, O. (1987) *Nucleic Acids Res.* **15**, 6589-6606.

56. Hunter, W. N., Kneale, G., Brown, T., Rabinovich, D., and Kennard, O. (1986) *J. Mol. Biol.* **190**, 605-618.

57. Rabinovich, D., Haran, T., Eisenstein, M., and Shakked, Z. (1988) *J. Mol. Biol.* **200**, 151-156.

58. Chevrier, B., Dock, A.-C., Hartmann, B., Leng, M., Moras, D., Thuong, M. Y., and Westhof, E (1986) *J. Mol. Biol.* **188**, 707-719.

59. Williams, L. D., Frederick, C. A., Gessner, R. V., and Rich, A. (1991) *Molecular Conformation and Biological Interactions*, pp 295-309, Indian Academy of Sciences, Bangalore.

60. Thomas, T. J. and Bloomfield, V. A. (1984) *Biopolymers* **23**, 1295-1306.

61. Gessner, R. G., Frederick, C. A., Quigley, G. J., Rich, A., and Wang, A. H. Jr. (1989) *J. Biol. Chem.* **264**, 7921-7935.

62. Lee, J. C. and Lee, L. L. Y. (1981) *J. Biol. Chem.* **256**, 625-631.

63. Sherman, S. E., Gibson, D., Wang, A. H. Jr., and Lippard, S. J. (1985) *Science* **230**, 412-417.

64. Cantor, C. R., Warshaw, M. M., and Tinoco, I. Jr. (1970) *Biopolymers* **9**, 1059-1077.

65. *Handbook of Biochemistry and Molecular Biology, 3rd Edition, Nucleic Acids Vol. I.*, CRC Press, Cleveland, Ohio, pp589.

66. Cantor, C. and Schimmel, P. (1980) *Biophysical Chemistry, Part II. Techniques for the study of Biological Structure and Function*, Freeman, San Francisco, California.

67. Lovett, C. M. Jr., Fitzgibbon, T. N., and Chang, R. (1989) *J. Chem. Edu.* **66**, 526-528.

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

69. Xodo, L. E., Manzini, G., Quadrifoglio, F., van der Marel, G. A., and van Boom, J. H. (1986) *Nucleic Acids Res.* **14**, 5389-5398.

70. Xodo, L. E., Manzini, G., Quadrifoglio, F., Yathindra, N., van der Marel, G. A., and van Boom, J. H. (1989) *J. Mol. Biol.* **205**, 777-781.

71. Seela, F. and Kehne, A. (1987) *Biochemistry* **26**, 2232-2238.

72. Xodo, L. E., Manzini, G., Quadrifoglio, F., van der Marel, G., and van Boom, J. H. (1988) *Biochemistry* **27**, 6321-6326.

73. Ts'o, P. O. P. (Ed.) (1974) *Basic Principles in Nucleic Acid Chemistry*, Academic Press, New York.

74. Uhlenbeck, O. C., Borer, P. N., Dengler, B., and Tinoco, I. Jr. (1973) *J. Mol. Biol.* **73**, 483-496.

75. Pörschke, D. (1978) *Eur. J. Biochem.* **39**, 117-126.

76. Appleby, W. D. and Kallenbach, N. R. (1973) *Biopolymers* **12**, 2093-2120.

77. Stannard, B. S. and Felsenfeld, G. (1975) *Biopolymers* **14**, 299-307.

78. *Mathcad 6.0 User's Guide*, Mathsoft Inc., Cambridge, Massachusetts.

79. Glasel, J. A. and Deutscher, M. P. (1995) *Introduction to Biophysical Methods for Protein and Nucleic Acid Research*, Academic Press, Inc., California.

80. Ivanov, V. I., Minchenkova, L. E., Minyat, E. E., Frank-Kamenetskii, M. D., and Schyolkina, A. K. (1974) *J. Mol. Biol.* **87**, 817-833.

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

82. Girod, J. C., Curtis Johnson, W. Jr., Huntington, S. K., and Maestre, M. F. (1973) *Biochemistry* **12**, 5092-5096.

83. Jeffrey, G. A. and Saenger, W. (1991) *Hydrogen Bonding in Biological Structures*, Springer-Verlag, Berlin Heidelberg.

84. Freier, S. M., Albergo, D. D., and Turner, D. H. (1983) *Biopolymers* **22**, 1107-1131.

85. Frisman, E. V., Veselkov, A. N., Slonitsky, S. V., and Karavaev, L. S. (1974)

Biopolymers **13**, 2169-2178.

86. Tabor, H. (1962) *Biochemistry* **1**, 496-501.

87. Hare, D. R. and Reid, B. R. (1986) *Biochemistry* **25**, 5341-5350.

88. Haasnoot, C. A. G., Hilbers, C. W., van der Marel, G. A., van Boom, J. H., Singh, U. C., Pattabiraman, N., and Kollman, P. A. (1986) *J. Biomol. Struc. Dyn.* **3**, 843-857.

89. Bolmmers, M. J. J., van de Ven, F. J. M., van der Marel, G. A., van Boom, J. H., and Hilbers, C. W. (1991) *Eur. J. Biochem.* **201**, 33-51.

90. Bolmmers, M. J. J., Walters, J. A. L. I., Haasnoot, C. A. G., Aelen, J. M. A., van der Marel, G. A., van Boom, J. H., and Hilbers, C. W. (1989) *Biochemistry* **28**, 7491-7498.

91. Haasnoot, C. A. G., de Bruin, S. H., Berendsen, R. G., Janssen, H. G. J. M., Binnendijk, T. J. J., Hilbers, C. W., van der Marel, G. A., and van Boom, J. H. (1983) *J. Biol. Struct. Dyn.* **1**, 115-129.

92. Ikuta, S., Chattopadhyaya, R., Ito, H., Dickerson, R. E., and Kearns, D. R. (1986) *Biochemistry* **25**, 4840-4849.

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

94. Crick, F. H. C., and Kendrew, J. C. (1957) *Adv. Protein Chem.* **12**, 133-.

95. Stout, G. H. and Jensen, L. H. (1989) *X-ray Structure Determination*, John Wiley and Sons, Inc., New York.

96. Giacovazzo, C., Monaco, H. L., Viterbo, D., Scordari, F., Gilli, G., Zanotti, G., and Catti, M. (1992) *Fundamentals of Crystallography*, Oxford Science Publications, New York.

97. Ladd, M. F. C. and Palmer, R. A. (1985) *Structure Determination by X-ray Crystallography*, Plenum Press, New York and London.

98. Glusker, J. P. and Trueblood, K. N. (1985) *Crystal Structure Analysis*, Oxford, New York.

99. McRee, D. (1993) *Practical Protein Crystallography*, Academic Press, Inc., San Diego.

100. *MSC/AFC Diffractometer Control Software*, Molecular Structure Corporation, Woodlands, Texas.

101. McPherson, A. (1983) *Methods Biochem. Annu.* **23**, 249-345.

102. Dickerson, R. E., Grzeskowiak, K., Grzeskowiak, M., Kopka, M. L., Larsen, T., Lipanov, A., Privé, G. C., Quintana, J., Schultze, P., Yanagi, K., Yuan, H., and Yoon, H. (1991) *Nucleosides & Nucleotides* **10**, 3-24.

103. *teXsan (95)*, Molecular Structure Corporation, Woodlands, Texas.

104. Cullity, B. D. (1978) *Elements of X-ray Diffraction*, Addison-Wesley Publishing Company, Inc..

105. Shin, Y. A. (1973) *Biopolymers* **12**, 2459-2475.

106. De Meester, P., Goodgame, D. M. L., Jones, T. J., and Skapski, A. C. (1974) *Biochim. Biophys. Acta* **353**, 392-394.

107. Aoki, K. (1976) *Biochim. Biophys. Acta* **447**, 379-381