Sequence Composition and Mismatch Effects on Triple Helix Formation

Thesis by

George C. Best

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1995

(Submitted September 2, 1994)

For my parents

ACKNOWLEDGMENTS

Five years, and so many people to thank. As is always the case, this thesis could not have been produced without help from a large number of people. First and foremost of these is my advisor, Peter Dervan. Peter manages to be supportive of his students, interested in their research, but never overbearing and intrusive. He consistently gives people sufficient freedom to follow their ideas where they might lead.

Scott Singleton developed the quantitative affinity cleavage titration used in chapter 2. We had several useful discussions about the complications and limitations of the technique. Scott Priestley and I struggled with many of the same problems when we were conducting experiments. I know that his careful insistence on the importance of the details of the procedure improved the quality of my work.

Ken Graham was one of the senior graduate students in the group when I started. He found the time to take me under his wing and show me many things about research at Caltech. Without him, my time at Caltech would have been much more difficult.

Milan Mrksich is one of the best chemists I know. More importantly, he is a good friend, drinking partner, and a golfing buddy. The time I spent with him away from the lab in the company of Shannon, Louie, Sharon, Lance and Sherrie helped keep me sane, and was more important than any of them could know.

The people who inhabited 316 and 312 Church during much of the time I've been here, Natalia Colocci, Michelle Parks, and Jim Kelly, have been great colleagues and have become friends. I could not have asked for a better bunch of

people with whom to spend long stretches of time, and Jim, whenever you want to attend a turf grass disease management seminar, just let me know.

My family has been supportive of me throughout my time at Caltech, even though I'm sure there were times when they wondered what I was doing here. I know that there certainly were times when I wondered what I was doing here. It was at those time when some of the things that they taught me bubbled to the surface and pulled me through.

Finally, I need to thank the people who paid for all of this. The Fannie and John Hertz Foundation gave me a predoctoral fellowship that put bread on the table and beer in the fridge. One group that is all too frequently forgotten in all of this is the American taxpayer who funds most of the research that occurs. If scientists were more conscientious about informing people about the degree to which taxpayer supported research underlies most of the technical advances that have improved life in the last 50 years, there might be more support for such spending

ABSTRACT

The association constants for the formation of sixteen triple helical complexes which vary at a single common position (Z•XY where Z = A, G, ^{me}C, T and XY = AT, GC, CG, TA) were determined by quantitative affinity cleavage titration. The association constants ranged from 9.5 x 10^4 M^{-1} (A•CG) to 5.6 x 10^7 M^{-1} (T•AT) at 22 °C, pH 7.0, 10 mM bis-tris, 100 mM NaCl, 250 μM spermine. For the sequences studied, T•AT is more stable than T•GC, T•CG, or T•TA by ≥ 2.3 kcal mol^{-1}; ^{me}C+GC is more stable than ^{me}C•AT, ^{me}C•CG, or ^{me}C•TA by ≥ 1.4 kcal mol^{-1}; G•TA is more stable than G•AT, G•CG, or G•GC by ≥ 0.8 kcal mol^{-1}. Under the conditions used in these experiments, none of the natural bases form triplets which offer means of recognizing the Watson-Crick CG base pair in a clearly energetically preferred manner.

The association constants for the formation of nine triple helical complexes of varying length and sequence composition were determined by quantitative DNase footprint titration. The association constants ranged from 2.0 x 10^6 M^{-1} to 3.1 x 10^8 M^{-1} at 22 °C, pH 7.0, 10 mM bis-tris, 100 mM NaCl, 250 μM spermine. The free energy of formation of the various triple helices was expressed as the linear combination of the free energy contributions of the various dinucleotide components of the oligonucleotide, and the value of each energetic contribution was determined. The 5′-TT-3′ dinucleotide makes the greatest contribution to the stability of the triple helix, -1.2 kcal•mol^{-1} dinucleotide^{-1}, while the inclusion of a 5′-^{me}C^{me}C-3′ dinucleotide requires the payment of an energetic penalty of 0.5 kcal•mol^{-1} dinucleotide^{-1}.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

Chapter 1.

Figures.

Tables.

Chapter 2.

Figures.

Tables.

Chapter 3.

Figures.

Tables.

CHAPTER ONE

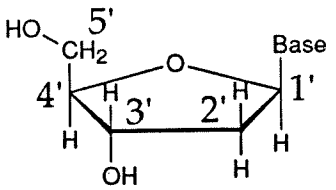
Elements of Molecular Recognition of DNA

Introduction

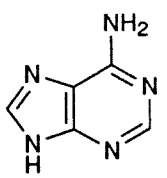
Deoxyribonucleic acid is the central molecule of life. Serving as the genetic repository, the DNA molecules in cells carry the information which is needed for the continuation of the cellular processes which comprise life. The regulation of processes involving DNA is critical to the proper functioning of the cell. This regulation is accomplished via the recognition of specific sequences of DNA. Since these sequence specific recognition events are essential for proper control of cellular differentiation, gene expression, cell division and other cellular processes, they represent potentially attractive targets for therapeutic intervention. As understanding of the forces which control these events increases, the design of sequence specific binding agents has become an important part of bioorganic chemistry. In this chapter, the structure of DNA, classes of DNA binding agents and some of the basic techniques used to study DNA binding agents are reviewed.

DNA Structure

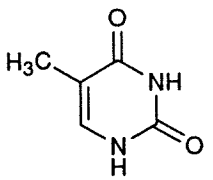
The DNA molecule is a polymer. The basic building block of this polymer is the nucleoside which is composed of a 2′-deoxyribose which is linked via a β glycosidic linkage to an aromatic heterocycle. In DNA there are four



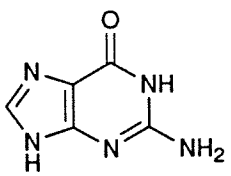
β-D-ribose



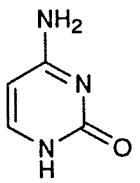
Adenine



Thymine



Guanine



Cytosine

Figure 1.1: Structure of a nucleotide. The structure of the D-ribose (Top). The structure of the four aromatic bases (Bottom).

heterocycles which are commonly found. Two of these heterocycles, thymine and cytosine, are substituted pyrimidine derivatives which are connected to the deoxyribose via N-3. The other two bases, adenine and guanine, are purine derivatives which are connected to the sugar via the N-9. These monomeric units are connected via ester linkages to the 3′ and 5′ oxygen atoms. Since the third acidic oxygen of the phosphate group is normally ionized at physiological pH, the resulting single strand of DNA is a polyanion.

The well known double helix structure of the DNA molecule is the result of the interaction of two of these single strands. In this structure, the antiparallel strands of nucleic acid form a right-handed helix. The helix axis is roughly perpendicular to the plane of the aromatic bases. Under conditions believed to be those found in the nucleus, there are normally 10 monomer units per strand in the helical repeat. The average rise per residue is ≈ 3.4 Å with a rotation of 36° around the helical axis.

The close association of two negatively charged molecules in the double helix is stabilized by hydrogen bonding interactions between the planar aromatic bases of opposite strands and stacking interactions between bases in the same strand. In this structure, the polyanionic backbones are located on the perimeter of the cylindrical cross section of the molecule. As a result of this gross arrangement of the molecule, the phosphodiester backbone of the DNA defines a pair of grooves.

The larger of these grooves, called the major groove is 11.7Å wide and 8.8Å deep. The other groove, the minor groove, is 5.7Å wide and 7.5Å deep.[1] The sides of the grooves are lined with the hydrophobic CH groups of the deoxyribose rings, while the floor of the groove is composed of the edges of the aromatic bases.

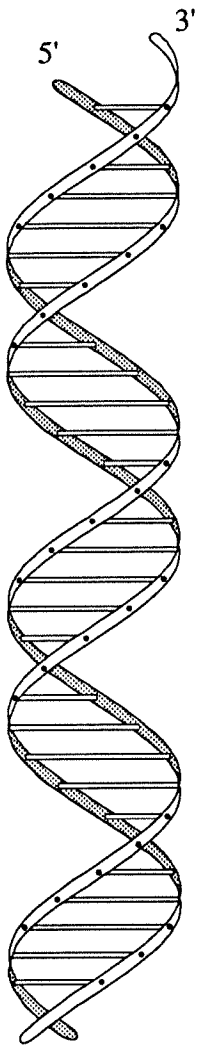
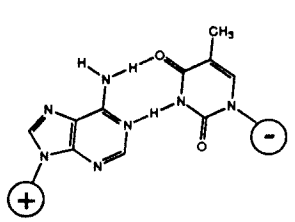


Figure 1.2: Structure of the double helix. A ribbon diagram representation.

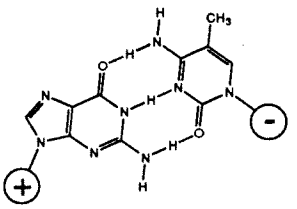
There are, formally, two distinct means of accomplishing the sequence specific recognition of a particular segment of the double helix. The first of these is indirect readout of the sequence via the sequence dependent conformational changes of the phosphodiester backbone of the DNA. While these factors may play a role in recognition of the double helix, especially by proteins, sequence specific recognition is primarily accomplished by the formation of hydrogen bonds with the edges of the bases along the floor of the groove.

The specific hydrogen bonding pattern present on the floor of the groove is a function of the base pair identity.[2] In the major groove, there are four different contacts which can be made. The AT base pair offers the lone pairs of the adenine N7 and thymine O6 as H bond acceptors, the hydrogen of the exocyclic amino group of the adenine is an H bond donor, and the 5-methyl group on the thymidine offers a surface for forming a van der Waals contact. The GC base pair has two H bond acceptors, the N7 and O6 of the guanine, one H bond donor, the exocyclic amine of the cytosine, and a differently positioned surface for van der Waals contacts, the C5 hydrogen of the cytosine.

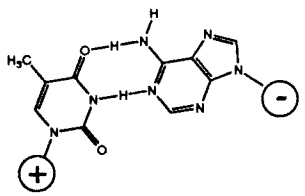
In the minor groove, the hydrogen bonding contacts are not as varied. The AT base pair has two hydrogen bond acceptors, the adenine N3 and the thymine O2 lone pairs. The GC base pair has the corresponding lone pairs, which can serve as H bond acceptors, and an H bond donor in the exocyclic amine of the guanine residue. In contrast to the major groove, it may not be possible to completely distinguish the four base pairs via specific contacts to the bases in the minor groove. In particular, discrimination between the AT and TA base pairs may be difficult.



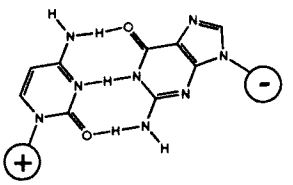
AT



GC



TA



CG

Figure 1.3: Structures of the four naturally occurring base pairs of DNA.

Classes of DNA Binding Molecules

Organic molecules which recognize and bind to DNA in a sequence specific manner can be divided into four classes according to the portion of the DNA molecule that they recognize. The first of these classes is composed of a number of polycyclic aromatic cations which bind to DNA via insertion of the aromatic ring into the vertical stack of the aromatic bases. This insertion is accomplished by the unwinding of the DNA and an elongation of the DNA along the helical axis. As a class, intercalators display a slight preference for intercalation at dinucleotide sites containing a GC base pair. This preference is thought to be the result of the slightly greater intrinsic dipole moment of GC base pairs, which is better able to induce polarization in the aromatic ring of the intercalating molecule.

The second class of molecules which bind to DNA in a sequence specific manner is a group of small molecules which recognize DNA via the formation of specific hydrogen bonds to the edges of the bases in the minor groove of the DNA. Perhaps the best understood members of this class of molecules is distamycin and its synthetic derivatives. Under solution conditions, these molecules form a 2:1 complex with DNA, binding in the minor groove of AT rich tracts of DNA. While specificity is achieved by the formation of hydrogen bonds with the adenine N3 and thymine O2 lone pairs, the driving force for the initial association is likely a combination of electrostatics and the hydrophobic effect. Molecules in this class are typically composed of a series of linked aromatic rings with a positively charged tail. The aromatic rings form favorable stacked complexes with each other and good stearic contacts with the sugar protons which line the sides of the minor groove. Recent work has extended the ability of

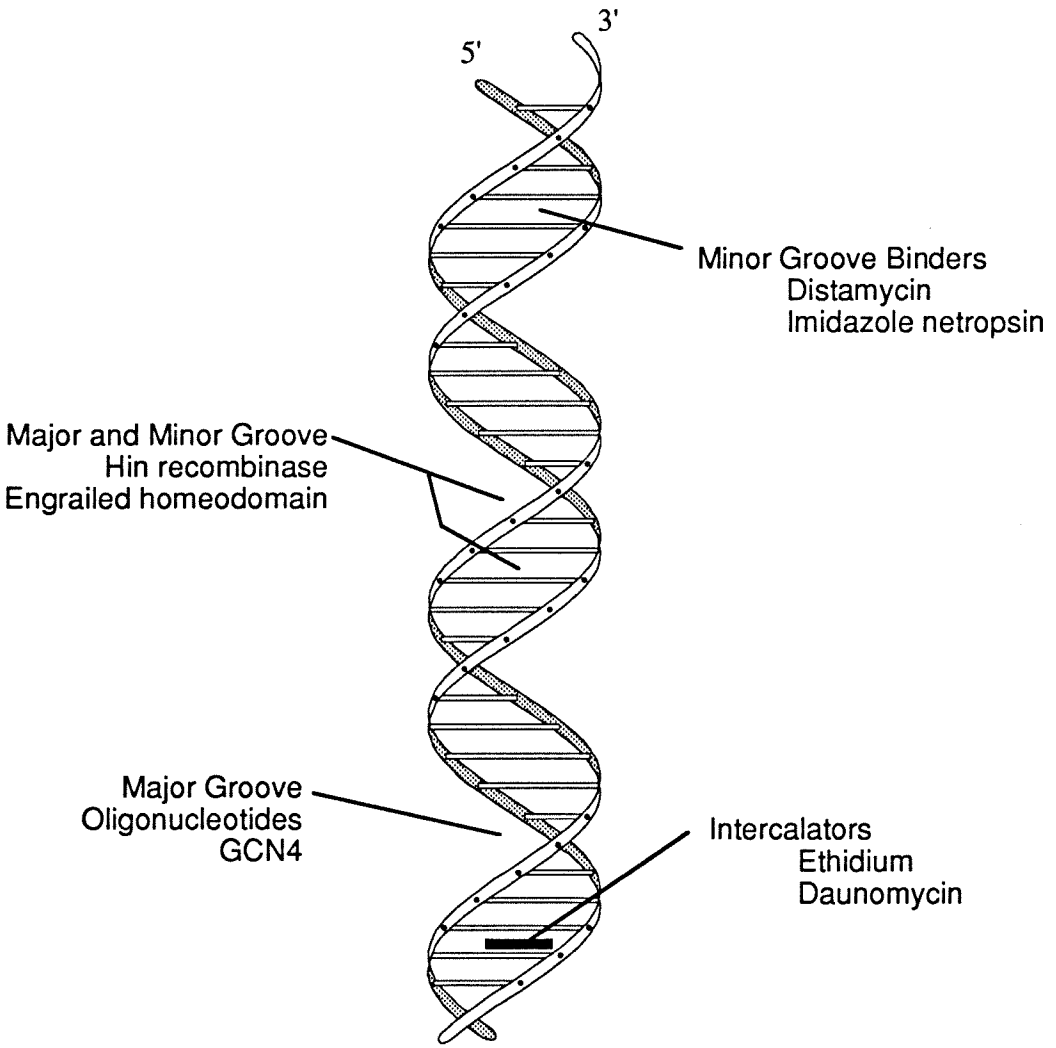
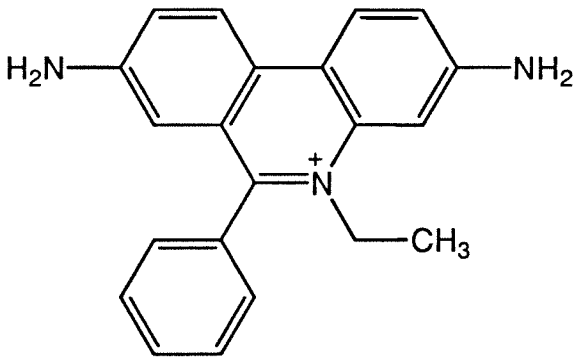
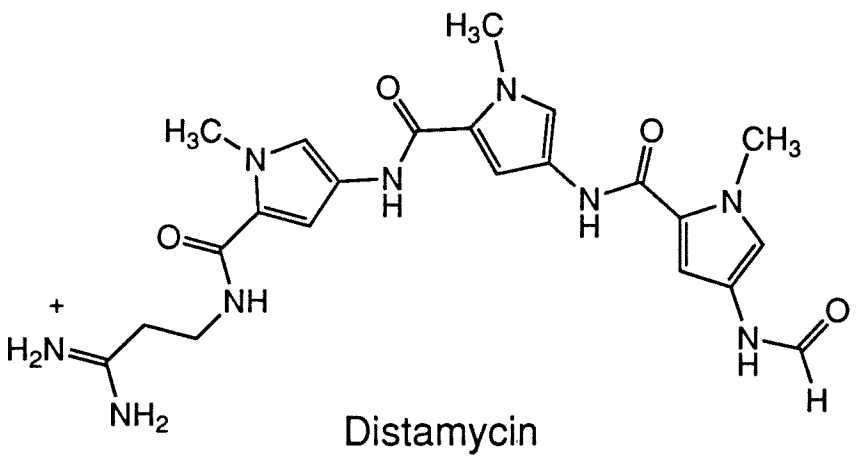


Figure 1.4: Four classes of DNA binding ligands.



Ethidium



Distamycin

Figure 1.5: Structures of an intercalator and a groove binding drug. Ethidium (Top). Distamycin (Bottom).

molecules in this class to recognize DNA sequences which contain GC base pairs.[3-8]

The third class of molecules which bind DNA in a sequence specific manner are other short pieces of nucleic acid. These molecules bind in a sequence specific manner by the formation of hydrogen bonds with the edges of the bases on the floor of the minor groove. More will be said about the use of oligonucleotides in the recognition of DNA in a following section.

The final class of agents for the sequence specific recognition of DNA is comprised of the DNA binding proteins which are found in nature. DNA binding proteins recognize their target sequences through a combination of shape selection, in which the conformation of the phosphodiester backbone of the DNA is recognized, and formation of specific hydrogen bonds to bases in the major, or both the major and minor grooves. In order to accomplish this recognition, the protein folds into one of family of conserved motifs which enable the protein to insert a set of residues, generally along one face of an α helix or a β sheet, into the major groove. The side chains of these residues make the sequence specific contacts to the bases.

Experimental Techniques

There are two complementary techniques which are used to study the interactions of DNA binding molecules with their target sequences in the absence of high resolution crystal or NMR structures of the complex. The first of these is footprinting. In a footprinting experiment, a sequence neutral DNA cleaving agent is used to generate random cleavage on a radiolabeled piece of DNA which contains one or more ligand binding sites. When the products of such a cleavage reaction are separated on a polyacrylamide sequencing gel, a ladder of cleavage

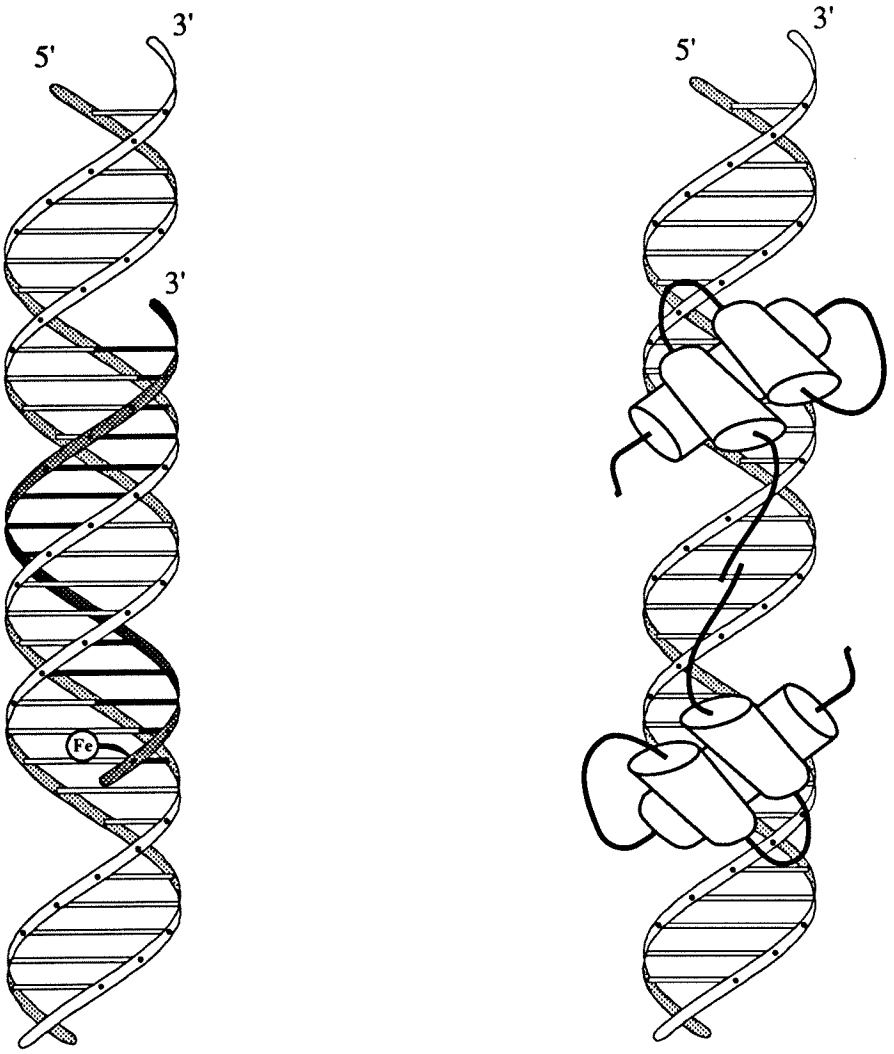


Figure 1.6: Schematic diagrams of a sequence specific DNA binding oligonucleotide and a DNA binding protein.

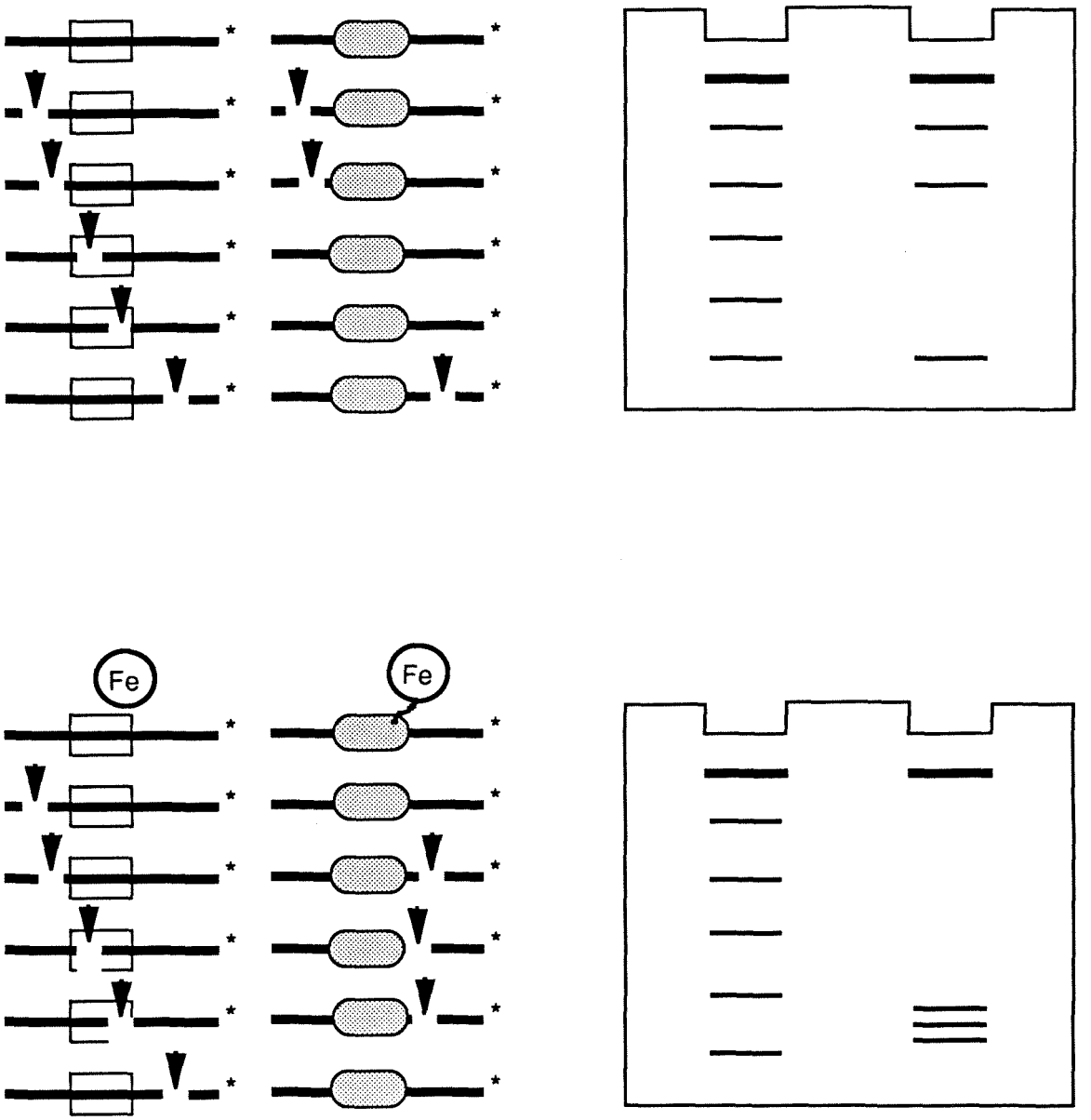


Figure 1.7: Schematic diagrams of footprinting and affinity cleavage reactions.

products is produced on the resulting autoradiogram. If a DNA binding ligand is present, it serves as a stearic blockade to the action of the cleaving agent, producing a diminution of bands relative to those seen in the ligand free control. Comparison of the footprint location with the products of Maxam-Gilbert reactions allows determination of the sequence of the DNA to which the ligand binds. A number of DNA cleaving agents, including the enzyme deoxyribnuclease I (DNase) and the small molecules methidium propyl EDTA•Fe(II) (MPE), and dimethyl sulfate (DMS), have been used in footprinting studies.

The complementary experiment to the footprinting experiment is affinity cleavage.[9] In an affinity cleavage experiment, a sequence neutral DNA cleaving moiety, commonly EDTA•Fe, is attached to a sequence specific DNA binding molecule. After equilibration of the ligand with the radiolabeled target, the cleavage reaction is initiated. When the products of the reaction are separated on a polyacrylamide sequencing gel, comparison of the reaction products with the products of Maxam-Gilbert sequencing reactions allows determination of the location of the EDTA•Fe moiety, and hence, the portion of the ligand to which it is attached, relative to the binding site.

The use of EDTA•Fe as the cleaving agent is advantageous in that, through the labeling of both strands of the DNA target, the identity of the groove in which cleaving moiety is located can also be obtained. The diffusable oxidant, presumably hydroxyl radical, which is responsible for the EDTA•Fe cleavage generated cleavage of the phosphodiester backbone, combines with the right-handed nature of double helical DNA to produce cleavage patterns which are diagnostic of the groove in which the EDTA•Fe is located. If the EDTA•Fe is located in the minor groove, the resulting cleavage pattern is roughly equal and

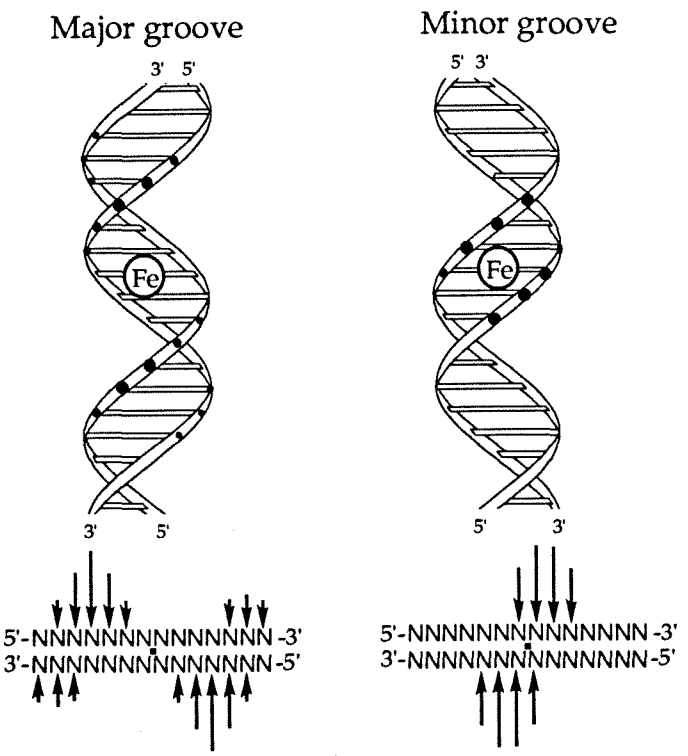


Figure 1.8: Relationship between groove location and cleavage pattern.

shifted in the 3′ direction on each strand. A major groove location results in a pair of these 3′ shifted cleavage patterns in which the cleavage intensity is unequally distributed. This is due to the diffusion of the oxidant into both of the adjacent minor grooves, which results greater cleavage at those locations on the phosphodiester backbone which are proximal to the EDTA•Fe.

Both the footprinting and affinity cleavage methodology can be adapted to provide quantitative binding data.[10-14] Since these techniques make use of high resolution sequencing gels, they can be used to determine the individual binding constant of the ligand at each of multiple sites on the target DNA. This is in contrast to many of the other techniques, such as the gel mobility shift assay, which only yield average properties when used with target DNA which contains multiple binding sites.

In order to use footprinting or affinity cleavage to determine the association constant for the binding of the ligand to the DNA, a series of equilibrium mixtures of radiolabeled DNA with varying concentrations of ligand are subjected to cleavage conditions. After the reaction products have been separated on a sequencing gel, an autoradiogram is obtained and used for the quantitative analysis. In the case of a footprinting experiment, the intensity of the bands at the binding site relative to those seen in a ligand absent control lane is inversely proportional to the degree of occupation of the site. After normalization of the data, an isotherm can be constructed and the association constant determined by a least squares curve fitting procedure.

In the case of the affinity cleavage titration, the intensity of cleavage proximal to the binding site is directly proportional to the occupancy of the site. After correction of the data for unspecific cleavage, an isotherm can be

constructed and the association constant determined by fitting the data to a binding isotherm.

Oligonucleotide Directed Triple Helix Formation

Given the central importance of DNA recognition events in the life cycle of a cell, the design and synthesis of artificial DNA binding ligands is a potentially powerful strategy for the generation of new therapeutics. However, the same features which make DNA ligand interactions such an attractive target for intervention argue that the side effects of such treatments could be devastating. For this reason it is desirable that novel DNA binding ligands be as specific as possible. Ligands which bind to sites composed of 15-18 base pairs (bp) are statistically likely to have only one binding site in the human genome of 3×10^9 bp. (See Table 1.1.) Of the four classes of DNA binding ligands described earlier, the use of oligomers of nucleic acid offers the easiest means to create molecules which are capable of recognizing binding sites of this size.

The earliest observation of a three stranded nucleic acid complex came in 1957, when Felsenfeld, Davies and Rich observed the formation of a 2:1 complex of poly(rU) to poly(rA) in optical mixing experiments.[15] It was proposed that the complex was composed of a normal Watson-Crick duplex to which the second poly(rU) strand was bound by the formation of hydrogen bonds to the edge of the adenine to form a U•AU base triplet.

Shortly thereafter, the formation of a similar complex between poly(rC) and oligomers of riboguanosine was observed.[16, 17] In these complexes, the molar ratio of C to G was 2:1. Since the complex was stable only at low pH, it was thought that the complex was composed of C+GC base triplets, which was isomorphous with the U•AU triplet. The observed necessity of acidic pH was

Table 1.1. Relationship between duplex binding site size and the number of unique sequences.

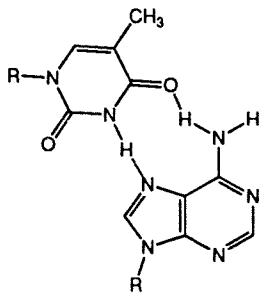
Binding Site Size[a]	Unique Sequences[b]	Binding Site Size[a]	Unique Sequences[b]
4	136	12	8,390,656
6	2080	15	536,870,912
8	32,896	18	68,719,607,808
10	524,800	21	2,199,023,255,552

[a] The binding site size is given in number of base pairs. [b] The number of unique sequences is determined from $4^n/2$ when n (the binding site size) is odd, and $(4^n + 4^{n/2})/2$ when n is even.[18, 19]

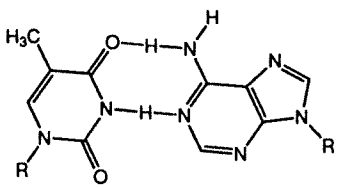
explained by the postulated protonation of the third strand cytosine in order to form two hydrogen bonds to the edge of the guanosine in the Watson-Crick duplex. Furthermore, the ability of poly[r(U-C)] to form a stable complex with a poly[d(G-A)]•poly[d(C-T)] duplex suggested that any homopurine strand could form a stable 1:2 complex with the appropriate homopyrimidine strand or strands at low pH.[20, 21]

Since hydrogen bonds of the type proposed in the models of the three stranded structures proposed above had been observed by Hoogsteen in the crystal structure of the 1-methylthymine and 9-methyladenine, these hydrogen bonds are often referred to as Hoogsteen hydrogen bonds and the third strand of the triple helix is referred to as the Hoogsteen strand. Subsequent fiber diffraction studies provided a rough model for the structure of the triple helix. In this model the third strand binds in the major groove of a Watson-Crick duplex in a parallel orientation to the purine strand.[22-24] The stabilization of this three stranded structure by the formation of Hoogsteen hydrogen bonds has been confirmed by recent NMR structural studies.[25-34]

These results languished in the literature until the advent of automated synthesis of short DNA oligomers made it possible to test the generality of triple helix formation using short pieces of DNA rather than polymers as the Hoogsteen strand. In 1987, Moser and Dervan provided direct evidence that a 15 nucleotide (nt) long oligomer consisting of 10 T residues and 5 C residues could form a stable triple helix at a single site on a 4.5 kbp linearized plasmid.[35] It was also discovered that changing one of the residues to create a base triplet mismatch dramatically reduced the apparent stability of the triple helix. Subsequent studies have shown that the triple helix may be used to recognize a single site against a background of several million base pairs of DNA, sufficient

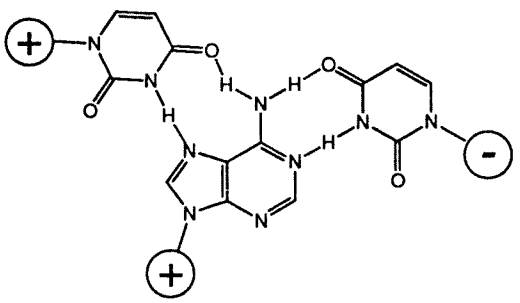


Hoogsteen

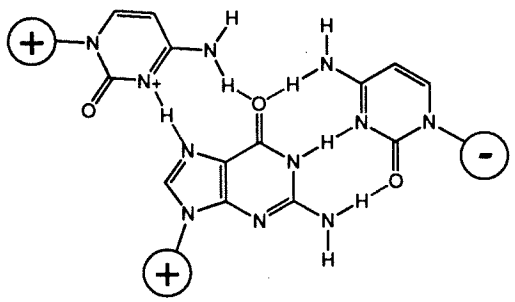


Watson-Crick

Figure 1.9: Hoogsteen and Watson-Crick hydrogen bonding schemes.



U•AU



C+GC

Figure 1.10: Structures of the U•AU and C+GC base triplets.

specificity to bind a single site in the yeast genome, or a human chromosome.[36-38]

As a methodology, oligonucleotide-directed triple helix formation has several advantages over other methods for the sequence specific recognition of DNA. The principles for designing an oligonucleotide ligand for given subset of target sites are well understood. In order to bind an AT base pair, a T is included in the third strand, while a C is included for binding to a GC base pair. This is in contrast to the other methods for accomplishing the sequence specific recognition of DNA, for which the rules for designing new ligands for specific sites are not as certain. Furthermore, the lack of a well defined, highly stable secondary structure for the oligonucleotide in solution means that there is no equivalent of the protein folding problem for oligonucleotides.

A number of recent studies have served to clarify the factors which influence the stability of triple helical complexes. The effect of pH on the stability of triple helices, presumably due to the necessity of protonating the C residues in the third strand, had been noted in the initial reports on triple helix formation with polymers.[20, 21] This sensitivity has also been noted in studies using shorter oligonucleotides to bind discrete sites on larger pieces of DNA.[35, 39-42]

Attempts to reduce this sensitivity have included the modification of the cytosine nucleoside,[40, 43] and the synthesis of novel nucleosides which do not require protonation in order to bind to GC base pairs.[44-46] Recently, Singleton and Dervan reported the results of a quantitative study of the effects of pH on triple helix formation.[47] In this work, they found that increasing the pH from 5.8 to 7.6 resulted in a tenfold decrease in the measured association constant for an oligonucleotide which contained thymine and cytosine residues. The

substitution of 5-methylcytosine for the cytosine resulted in a 0.1 - 0.4 kcal•mol^{-1} stabilization of the triple helical complex.

The stability of the triple helix is also sensitive to other solution conditions, particularly the concentration and valency of the cations present and the temperature of the medium.[14, 42, 48-56] As might be expected, since the triple helix is formed of three strands of polyanion, the stability of the complex is enhanced by increasing concentrations of cations. However, this simple formulation fails to capture the subtleties of triple helix behavior in the presence of complex mixtures of cations. In particular, Singleton and Dervan have presented an analysis of the effects of variation of the concentration of cations of varying valency in a complex mixture which is believed to represent the conditions found inside a cell (140 mM potassium, 1 mM magnesium, and 1 mM spermine).[48] Under these conditions, decreasing the concentration of the monovalent cation potassium to 90 mM resulted in a 20-fold increase in the stability of the complex. Similarly, decreasing the concentration of magnesium also resulted in an increase in the stability of the complex, albeit only slightly. However, the stability of the complex was found to be highly sensitive to the concentration of the tetravalent cation spermine. An increase in the spermine concentration from 0.40 mM to 1.0 mM resulted in an approximate 30-fold increase in the stability of the complex.

These results are explained by reference to the counterion condensation model of Manning and Record.[57, 58] In this model, the various cationic species may be thought of as being in competition for a limiting number of cation binding site on the triple helix. Increasing the concentration of potassium results in a greater fraction of these sites being occupied by potassium relative to spermine. The net condensation of several potassium ions per spermine released

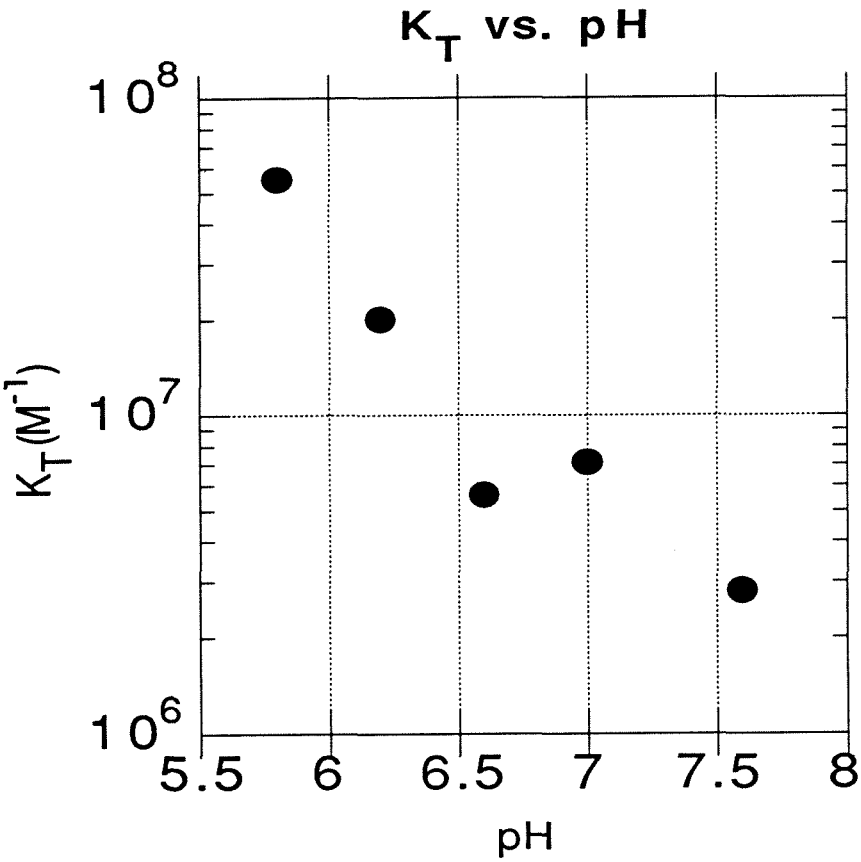


Figure 1.11: The effect of pH on the association constant of an oligonucleotide. Data taken from reference 14.

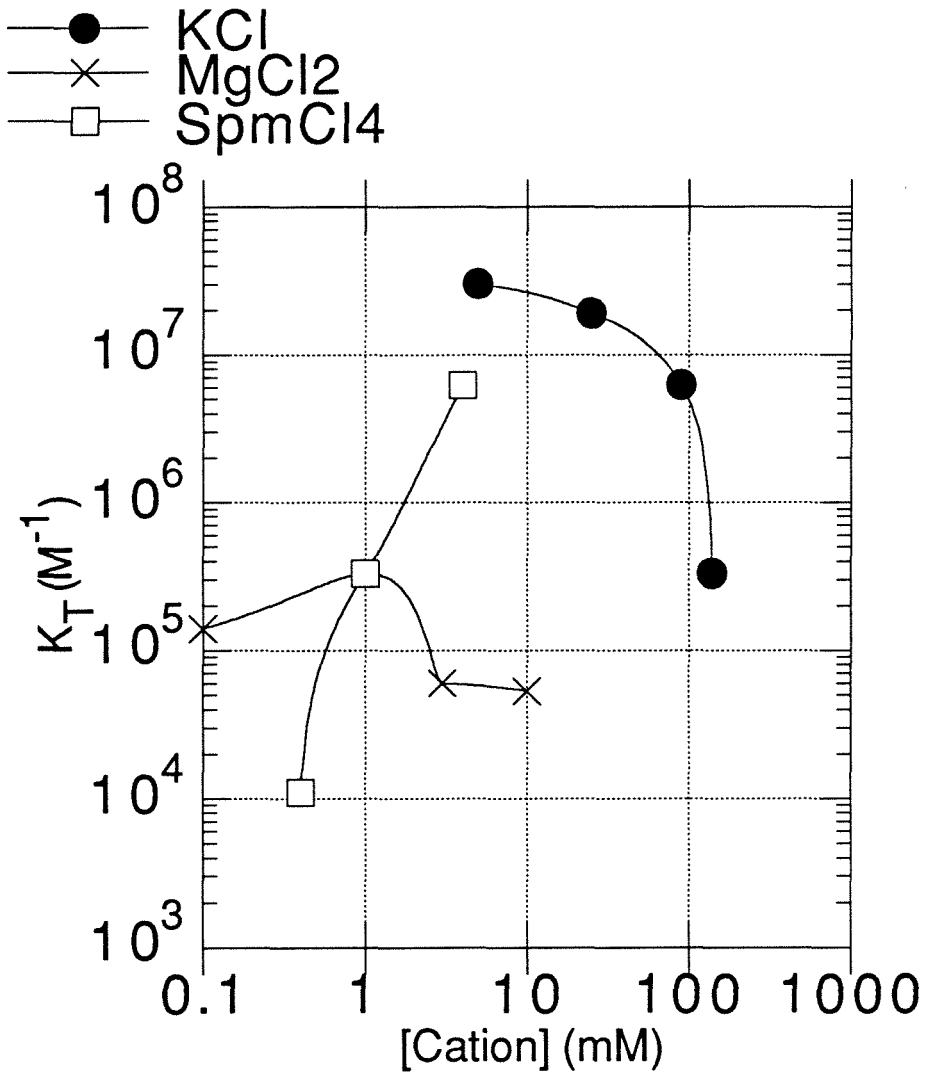


Figure 1.12: The effect of cation concentration and valency on the association constant for an oligonucleotide. The cation whose concentration is varied is indicated in the legend. Unvaried concentrations are 140 mM KCl, 1 mM $MgCl_2$, and 1 mM $SpmCl_4$. Data taken from reference 14.

results in a net decrease in the entropy of the system and hence affects the binding constant in a deleterious manner. Increasing the concentration of spermine acts in the opposite manner, resulting in an increase in the fraction of sites occupied by spermine with the concomitant release of four potassium ions per spermine bound. This release is entropically favorable and increases the stability of the complex.

The stability of triple helical complexes is also sensitive to solution temperature. As might be expected for an associative reaction, triple helix formation is disfavored by increasing temperatures. As a result, optical melting experiments have been used to study relative triple helix stabilities. In more quantitative work, Singleton and Dervan reported that the association constant for the formation of a 15 nt triple helix decreased approximately 100-fold as the temperature was raised from 8°C to 37°C.[14, 59] Although the actual magnitudes of the enthalpic and entropic contributions to triple helix formation on a per nucleotide triplet basis are somewhat uncertain, it can safely be stated that the ethalpy of triple helix formation is negative, as is the entropy of triple helix formation. The magnitude of these terms, particularly the enthalpy, seems to be dependent on the measurement technique and sequence composition of the particular triple helix used in a given study.

While the effects of various solution conditions on the stability of triple helical complexes has been studied extensively, the effects of the composition of the oligonucleotide on the stability of the resulting complex has not been as thoroughly studied. The work presented in this thesis represents the first quantitative examination of the effects of base triplet mismatches and oligonucelotide sequence composition on the stability of the resulting triple helical complexes.

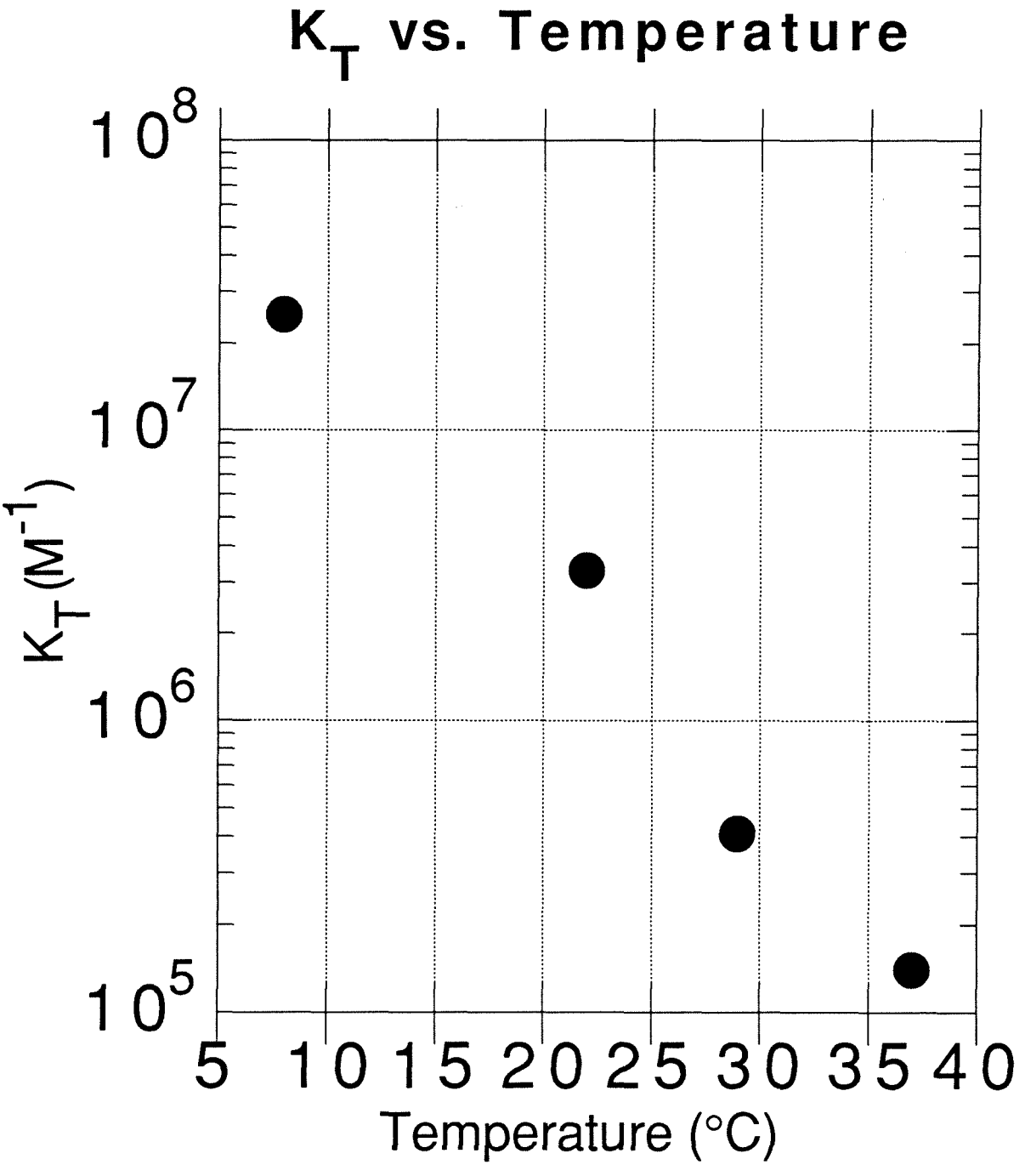
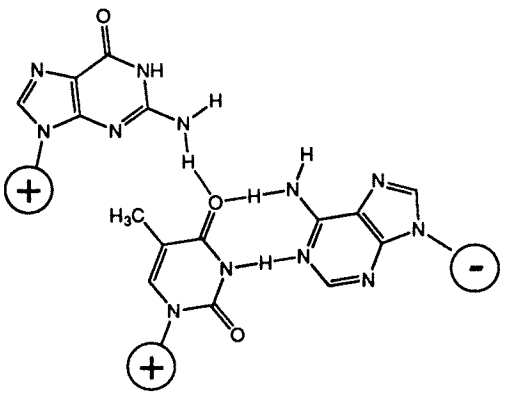


Figure 1.13: The effect of temperature on the association constant for a given oligonucleotide. Data taken from reference 14.

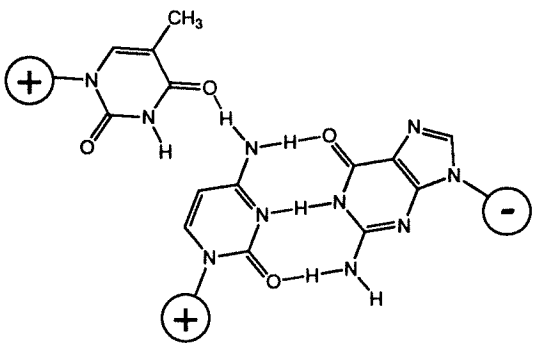
Other Directions in the Study of Triple Helix Formation

Efforts to extend the versatility of triple helix mediated recognition of DNA are the subject of much ongoing research. The primary goal of these efforts is to overcome the current limitations of triple helix formation. As has been noted above, the synthesis of novel nucleosides aimed at reducing the sensitivity of the triple helical complex to the pH of the medium has already made some progress.[40, 43-46] The other major limitation of triple helix mediated DNA recognition is the requirement that the binding site be entirely composed of AT and GC base pairs. The earliest approach to alleviating this limitation was the identification of other natural base triplets which are compatible with the pyrimidine•purine•pyrimidine (Y•RY) triple helix. These studies have suggested that the G•TA and T•CG base triplets might be sufficiently stable to accomplish recognition of all four base pairs. At the least, these might be interactions which are energetically tolerable, and allow the inclusion of a few pyrimidine bases in an otherwise homopurine run.[60, 61]

Another approach towards the expansion of the number of sequences which can be recognized by triple helix formation is the synthesis of novel nucleosides which are capable of recognizing the TA and CG base pairs.[62, 63] While these efforts have yielded some intriguing leads, the synthesis of novel bases requires the investment of a considerable amount of time and effort. Alternate strand triple helix formation allows recognition of sites of the types 5′-(purine)$_m$(pyrimidine)$_n$-3′ and 5′-(pyrimidine)$_m$(purine)$_n$-3′ in a much more synthetically accessible manner.[64, 65] It is likely, however, that the size of the "half-sites" will have to exceed some minimum length (m, n $\geq$ x) if the structure is to be stable.

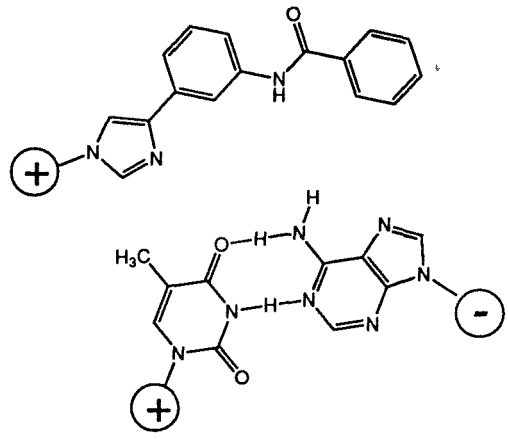


G•TA

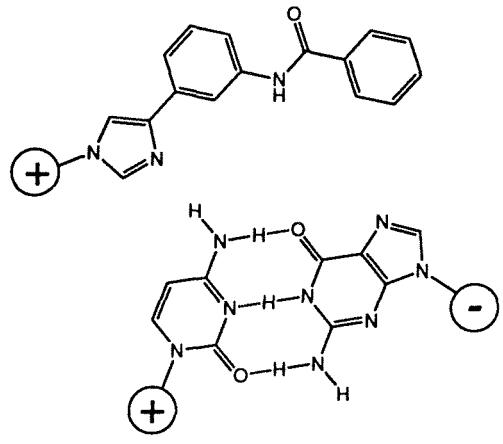


T•CG

Figure 1.14: Structure of the G•TA and T•CG base triplets.



D_3•TA



D_3•CG

Figure 1.15: Structure of the D_3•TA and D_3•CG base triplets.

More recently, a second structural family of triple helices has been identified. In this family, the purine strand of a homopurine run is recognized via the formation of G•GC, A•AT and T•AT base triplets.[66-69] The third strand in this structure is oriented in an antiparallel manner to the homopurine tract in the binding site,[68] and the hydrogen bonds between the third strand and the purine base are of the reverse Hoogsteen type, in which thymine bases in the third strand have been flipped relative to those found in the Y•RY triple helix discussed above.[70]

Work on this family of triple helices is proceeding along similar lines to that of the Y•RY triple helix. A search for other natural base triplets has been completed,[71] and work commenced on the synthesis of non-natural nucleosides for the recognition of the CG and TA base pairs.[72]

Since the third strand of the purine•purine•pyrimidine (R•RY) triple helix does not require protonation in order to form two hydrogen bonds to each base within the binding site, this motif is likely to prove less sensitive to pH variations. For this reason, triple helices of the R•RY type may prove most useful for the recognition of target sites which are rich in GC base pairs. Furthermore, the two families of triple helices may be combined and used to recognize binding sites of the 5′-$(purine)_m(pyrimidine)_n$-3′ and 5′-$(pyrimidine)_m(purine)_n$-3′ types.[73-76]

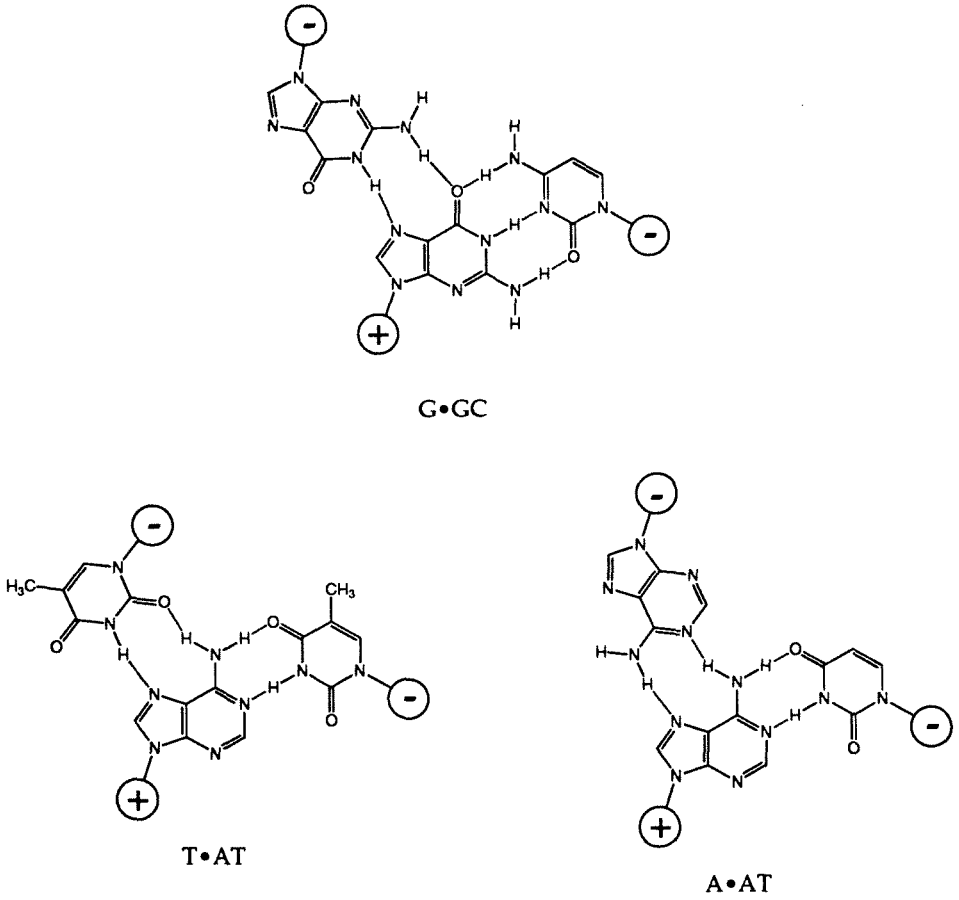


Figure 1.16: Structure of the G•GC, A•AT, and T•AT base triplets in the R•RY triple helix motif.

References

(1) Blackburn, G. M.; Gait, M. J. eds., Nucleic Acids in Chemistry and Biology, New York: Oxford University Press, 1990.

(2) Seeman, N. C.; Rosenberg, J. M.; Rich, A. *Proc. Natl. Acad. Sci. USA,* **1976,** **73,** 803-808.

(3) Mrksich, M.; Wade, W. S.; Dwyer, T. J.; Geierstanger, B. H.; Wemmer, D. E.; Dervan, P. B. *Proc. Natl. Acad. Sci., USA,* **1992,** *89,* 7586-7590.

(4) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1993,** *115,* 2572-2576.

(5) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1993,** *115,* 9892-9899.

(6) Wade, W. S.; Mrksich, M.; Dervan, P. B. *Biochemistry,* **1993,** *32,* 11385-11389.

(7) Geierstanger, B. H.; Jacobsen, J. P.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *Biochemistry,* **1994,** *33,* 3055-3062.

(8) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1994,** *116,* 3663-3664.

(9) Dervan, P. B. *Science,* **1986,** *232,* 464-471.

(10) Brenowitz, M.; Senear, D. F.; Shea, M. A.; Ackers, G. K. *Meth. Enzymol.* **1986,** *130,* 132-181.

(11) Senear, D. F.; Brenowitz, M.; Shea, M. A.; Ackers, G. K. *Biochemistry,* **1986,** *25,* 7344-7354.

(12) Brenowitz, M.; Senear, D. F.; Shea, M. A.; Ackers, G. F. *Proc. Natl. Acad. Sci. USA,* **1986,** *83,* 8462-8466.

(13) Singleton, S. F.; Dervan, P. B. *J. Am. Chem. Soc.* **1992,** *114,* 6957-6965.

(14) Singleton, S. F.: Ph.D. Thesis. California Institute of Technology, 1995.

(15) Felsenfeld, G.; Davies, D. R.; Rich, A. *J. Am. Chem. Soc.,* **1957,** *79,* 2023-2024.

(16) Lipsett, M. N. *J. Biol. Chem.,* **1964,** *239,* 1256, 1260.

(17) Howard, F. B.; Frazier, J.; Lipsett, M. N.; Miles, H. T. *Biochem. Biophys Res. Commun.*, **1964**, *17*, 93-102.

(18) Dervan, P. B. *Biochemistry*, **1987**, *26*, 4171-4171.

(19) Strobel, S. A. Ph.D. Thesis, California Institute of Technology, 1992.

(20) Miller, J. H.; Sobell, H. M. *Proc. Natl. Acad. Sci. USA*, **1966**, *55*, 1201-1205.

(21) Morgan, A. R.; Wells, R. D. *J. Mol. Biol.* **1968**, *37*, 63-80.

(22) Arnott, S.; Bond, P. J. *Nature New Biol.* **1973**, *244*, 99-101.

(23) Arnott, S.; Bond, P. J.; Selsing, E.; Smith, P. J. C. *Nucleic Acids Res.*, **1976**, *3*, 2459-2470.

(24) Arnott, S.; Selsing, E. J. Mol. Biol., *J. Mol. Biol.*, **1974**, *88*, 509-521.

(25) Gorennstein, D. G.; Luxon, B. A.; Goldfield, E. M.; Lai, K.; Vegeais, D. *Biochemistry*, **1982**, *21*, 580-589.

(26) de los Santos, C.; Rosen, M.; Patel, D. *Biochemistry*, **1989**, *28*, 7282, 7289.

(27) Rajagapol, P.; Feigon, J. *Nature*, **1989**, *339*, 637-640.

(28) Rajagapol, P.; Feigon, J. *Biochemistry*, **1989**, *28*, 7859-7870.

(29) Pilch, D. S.; Levenson, C.; Shafer, R. H. *Proc. Natl. Acad. Sci. USA*, **1990**, *87*, 1942-1946.

(30) Mooren, M. M. W.; Pulleyblank, D. E.; Wijmenga, S. S.; Blommers, M. J. J.; Hilbers, C. W. *Nucleic Acids Res.*, **1990**, *18*, 6523-6529.

(31) Umemoto, K.; Sarma, M. H.; Gupta, G.; Luo, J.; Sarma, R. H. *J. Am. Chem. Soc.*, **1990**, *112*, 4539-4545.

(32) Live, D. H.; Radhakrishnan, I.; Misra, V.; Patel, D. J. *J. Am. Chem. Soc.*, **1991**, *113*, 4687-4688.

(33) Radhakrishnan, I.; Gao, X.; de los Santos, C.; Live, D.; Patel, D. J. *Biochemistry*, **1991**, *30*, 9022-9030.

(34) Koshlap, K. M.; Gillespie, P.; Dervan, P. B.; Feigon, J. *J. Am. Chem. Soc.*, **1993**, *115*, 7908-7909.

(35) Moser, H. E.; Dervan, P. B. *Science*, **1987**, *238*, 645-650.

(36) Strobel, S. A.; Moser, H. E.; Dervan, P. B. *J. Am. Chem. Soc.*, **1988**, *110*, 7927-7929.

(37) Strobel, S. A.; Dervan, P. B. *Science*, **1990**, *249*, 73-75.

(38) Strobel, S. A.; Doucette-Stamm, L. A.; Riba, L.; Housman, D. E.; Dervan, P. B. *Science*, **1991**, *254*, 1639-1642.

(39) Lyamichev, V. I.; Mirkin, S. M.; Frank-Kamenetskii, M. D.; Cantor, C. R. *Nucleic Acids Res.*, **1988**, *16*, 2165-2178.

(40) Povsic, T. J.; Dervan P. B. *J. Am. Chem. Soc.*, **1989**, *111*, 3059-3061.

(41) Maher, L. J.; Dervan, P. B.; Wold, B. *Biochemistry*, **1990**, *29*, 8820, 8826.

(42) Plum, G. E.; Park, Y.-W.; Singleton, S. F.; Dervan, P. B.; Breslauer, K. J. *Proc. Natl. Acad. Sci. USA*, **1990**, *87*, 9436-9440.

(43) Povsic, T. J., Ph.D. Thesis, California Institute of Technology, 1991.

(44) Ono, A.; Ts'o, P. O. P.; Lou-Sing, K. *J. Am. Chem. Soc.* **1991**, *113*, 4032-4033.

(45) Koh, J. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 1470-1478.

(46) Krawczyk, S. H.; Milligan, J. F.; Wadwani, S.; Moulds, C.; Froehler, B. C.; Matteucci, M. D. *Proc. Natl. Acad. Sci. USA*, **1992**, *89*, 3761-3764.

(47) Singleton, S. F.; Dervan, P. B. *Biochemistry*, **1992**, *31*, 10995-11003.

(48) Singleton, S. F.; Dervan, P. B. Biochemistry, 1993, 32, 13171-13179.

(49) Krakauer, H.; Sturtevant, J. M. *Biopolymers*, **1968**, *6*, 491-512.

(50) Krakauer, H. *Biochemistry*, **1974**, *13*, 2579-2589.

(51) Lee, J. S.; Woodworth, M. L.; Latimer, L. J. P.; Morgan, A. R. *Nucleic Acids, Res.*, **1984**, *12*, 6603-6614.

(52) Latimer, L. J. P.; Hampel, K.; Lee, J. S. *Nucleic Acids Res.*, **1989**, *17*, 1549-1561.

(53) Pilch, D. S.; Brousseau, R.; Shafer, R. H. *Nucleic Acids Res.*, **1990**, *18*, 5743-5750.

(54) Shea, R. G.; Ng, P.; Bischotberger, N. *Nucleic Acids Res.*, **1990**, *18*, 4859-4866.

(55) Durand, M.; Peloille, S.; Thuong, N. T.; Maurizot, J. C. *Biochemistry*, **1992**, *31*, 9197-9204.

(56) Rougée, M.; Faucon, B.; Mergny, J. L.; Barcelo, F.; Giovannageli, C.; Garestier, T.; Hélène, C. *Biochemistry*, **1992**, *31*, 9269-9278.

(57) Manning, G. S. *Q. Rev. Biophys.*, **1978**, *11*, 179-246.

(58) Record, M. T., Jr.; Anderson, C. F.; Lohman, T. M. *Q. Rev. Biophys.*, **1978**, *11*, 103-178.

(59) Singelton S. F.; Dervan, P. B. *J. Am Chem. Soc.*, submitted.

(60) Griffin, L. C.; Dervan, P. B. *Science*, **1990**, *249*, 73-75.

(61) Yoon, K.; Hobbs, C. A.; Koch, J.; Sardaro, M.; Kutny, R.; Weis, A. *Proc. Natl. Acad. Sci. USA*, **1992**, *89*, 3840-3844.

(62) Griffin, L. C.; Kiessling, L. L.; Beal, P. A.; Gillespie, P.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 7976-7982.

(63) Kiessling, L. L.; Griffin, L. C.; Dervan. P. B. *Biochemistry* **1992**, *31*, 2829-2834.

(64) Horne, D. A.; Dervan, P. B. *J. Am. Chem. Soc.*, **1990**, *112*, 2435-2437.

(65) Horne, D. A.; Dervan, P. B. Unpublished data.

(66) Cooney, M.; Czernuszewicz, G.; Postel, E. H.; Flint, S. J.; Hogan, M. E. *Science*, **1988**, *241*, 456-459.

(67) Kohwi, Y.; Kohwi-Shigematsu, T. *Proc. Natl. Acad. Sci. USA*, **1988**, *85*, 3781-3785.

(68) Beal, P. A.; Dervan, P. B. *Science*, **1991**, *251*, 1360-1363.

(69) Durland, R. H.; Kessler, D. J.; Gunnell, S.; Duvic, M.; Hogan, M. E.; Pettitt, B. M. *Biochemistry*, **1991**, *30*, 9246-9255.

(70) Beal, P. A. Ph.D. Thesis, California Institute of Technology, 1993.

(71) Beal, P. A.; Dervan, P. B. *Nucleic Acids Res.* **1992**, *20*, 2773-2776.

(72) Stilz, H. U.; Dervan, P. B. *Biochemistry* **1993**, *32*, 2177-2185.

(73) Jayasena, S. D.; Johnston, B. H. *Biochemistry* **1992**, *31*, 320-327.

(74) Beal, P. A.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 4976-4982.

(75) Jayasena, S. D.; Johnston, B. H. *Nucl. Acids Res.* **1992**, *20*, 5279-5288.

(76) Jayasena, S. D.; Johnston, B. H. *Biochemistry* **1993**, *32*, 2800-2807.

CHAPTER TWO

Characterization of the Energetics of Formation of the Sixteen Pyrimidine•Purine•Pyrimidine Base Triplets by Quantitative Affinity Cleavage Titration

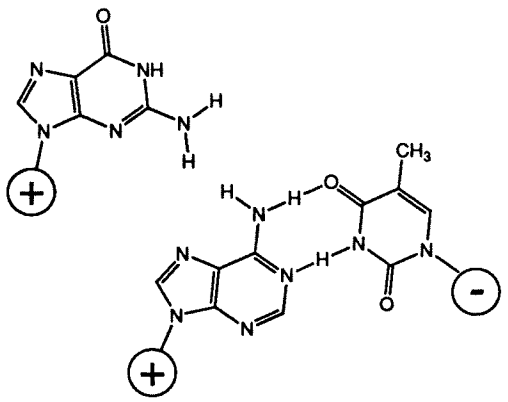
Introduction

Oligonucleotide-directed triple helix formation is a versatile method for accomplishing the sequence specific recognition of double helical DNA.[1-3] Triple helices can be classified into two structural motifs: those in which the third strand is primarily composed of pyrimidine bases,[4] and those in which the third strand is primarily composed of purine bases.[3-8] Triple helices composed of combinations of these two motifs can also be formed.[9-12] Triple helix formation in the more thoroughly characterized pyrimidine motif is sensitive to the backbone composition (DNA or RNA) of the three strands,[13-16] length of the third strand,[1, 17] single base mismatches,[1, 17-19] pH,[20-22] salt conditions of the buffer solution.[23,] and the structure of the polyvalent cations present in solution.[24] Oligonucleotide-directed triple helix formation has been used to mediate single site cleavage of human chromosomal DNA,[25] as well as interfere with the function of DNA binding proteins.[26-29] The ability to target a broad range of DNA sequences,[1, 30] and the high stability of the resulting local triple

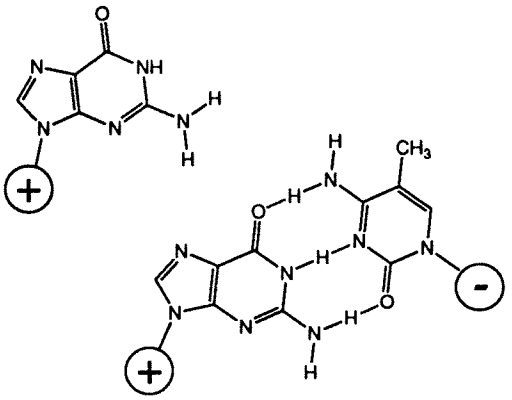
helical structures make this a powerful technique for the recognition of single sites within megabase segments of double helical DNA.

In an effort to understand the interactions which give rise to specific recognition of DNA in the triple helix, the specificity afforded by the natural bases in triple helical recognition has been studied by affinity cleavage,[31-33] melting temperature studies,[34-36] gel mobility shift assays,[37] and intramolecular triple helix formation.[38, 39] Key interactions shown by these studies to promote triple helix formation have been characterized by NMR spectroscopy. Hoogsteen type hydrogen bonds have been observed between T residues in the third strand and AT base pairs in the duplex,[40-42] and between N3 protonated cytosine or 5-methyl cytosine in the third strand and GC base pairs in the duplex.[18, 39, 43] Recent studies have extended the pyrimidine motif through the use of both non-pyrimidine and non-natural bases.[44-47] The best characterized of these are the G•TA, D_3•TA, and P_1•GC triplets, for which NMR studies have been completed.[48-55]

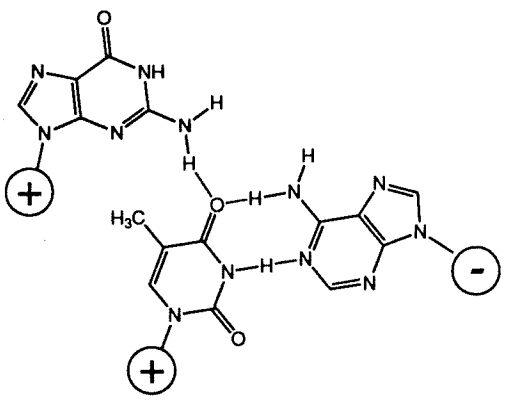
The study of the energetics of triple helix formation have proceeded in parallel with the structural studies. An early effort used affinity cleavage to explore the relative stability of base triplets consisting of the four natural bases and inosine across the four natural base pairs.[48] Other work made use of optical melting experiments[34-36] or gel mobility shift experiments[37] to explore the relative stability of triple helices containing various triplets. The introduction of the quantitative affinity cleavage titration (QACT) has allowed the facile determination of the equilibrium binding constants for the third strand of triple helical complexes.[17, 22, 23] QACT has since been used to explore the effects of pH[22] and cation valency and concentration[23] on triple helix formation. Here we



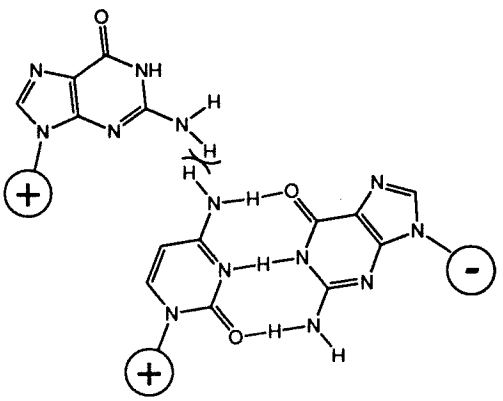
G•AT



G•GC

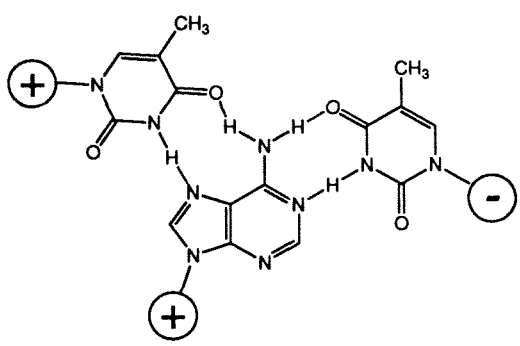


G•TA

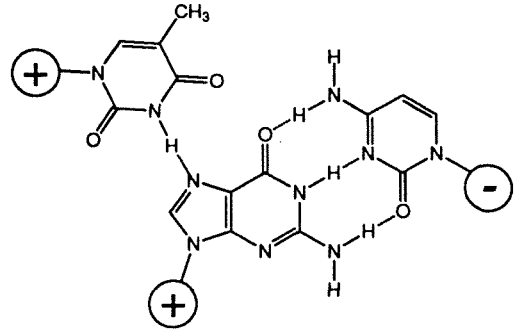


G•CG

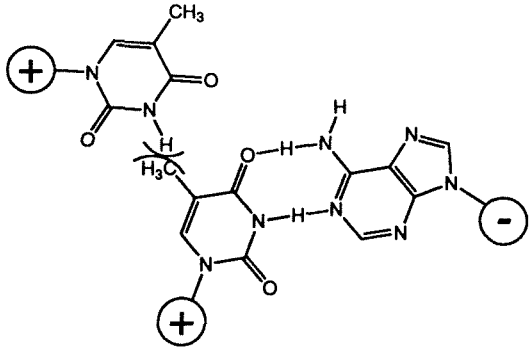
Figure 2.1: Structure of the four G•XY base triplets.



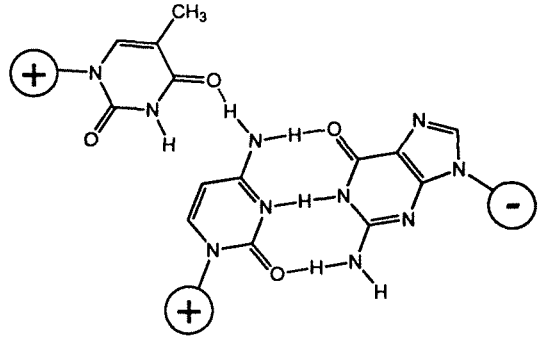
T•AT



T•GC



T•TA



T•CG

Figure 2.2: Structure of the four T•XY base triplets.

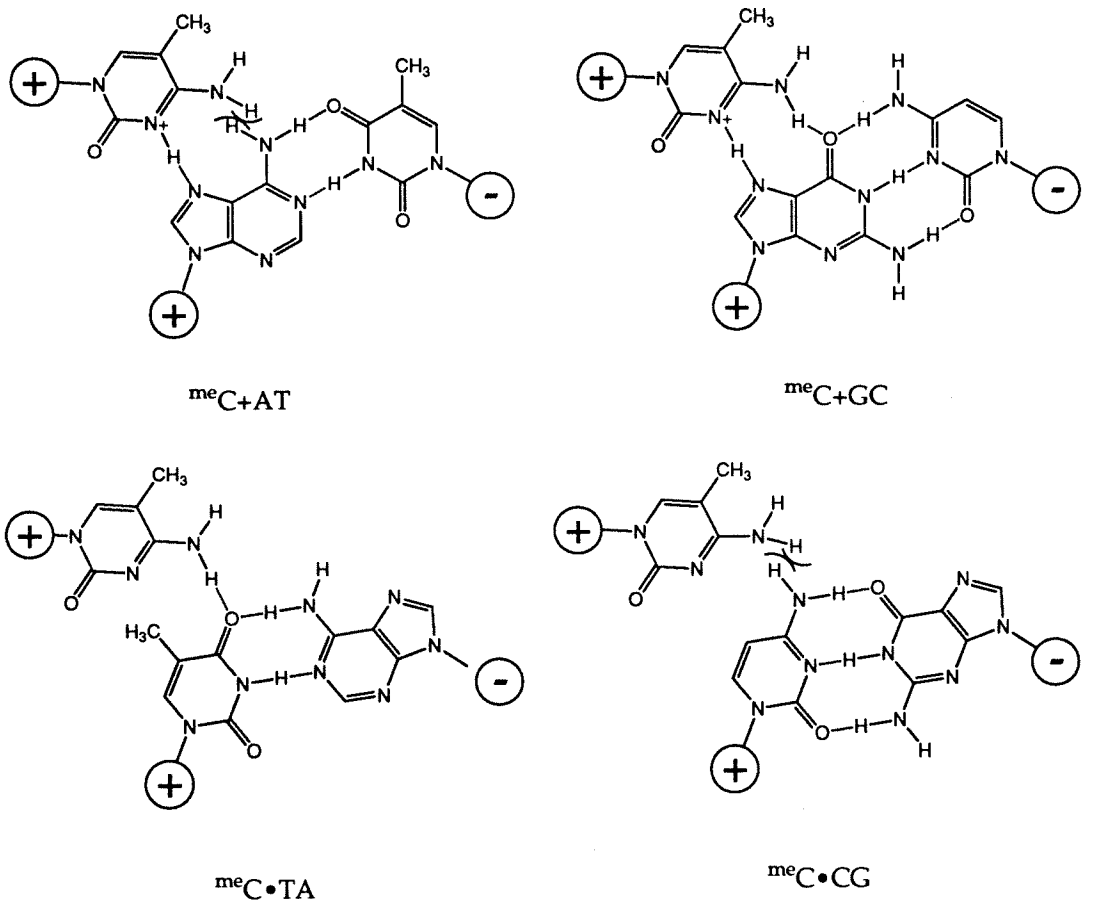
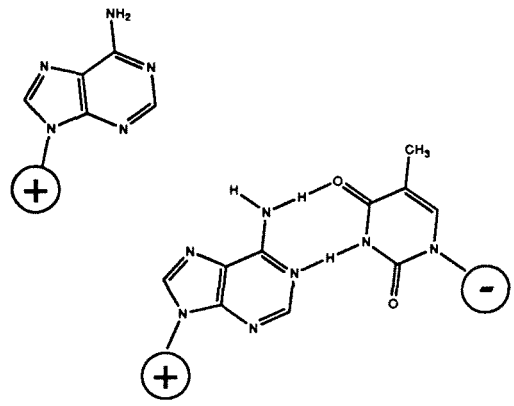
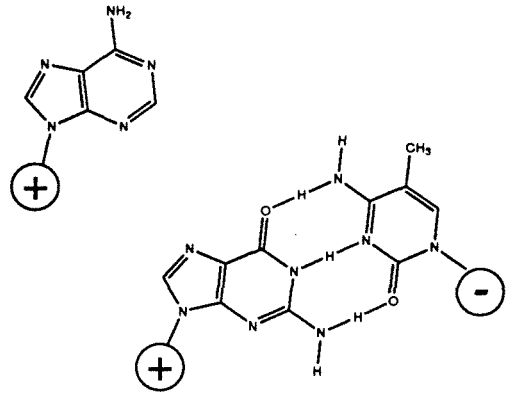


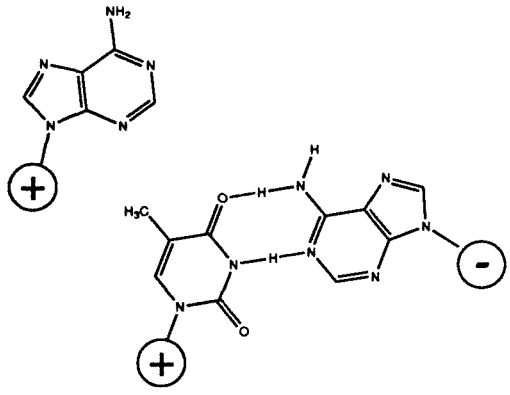
Figure 2.3: Structure of the four $^{me}C \bullet XY$ base triplets.



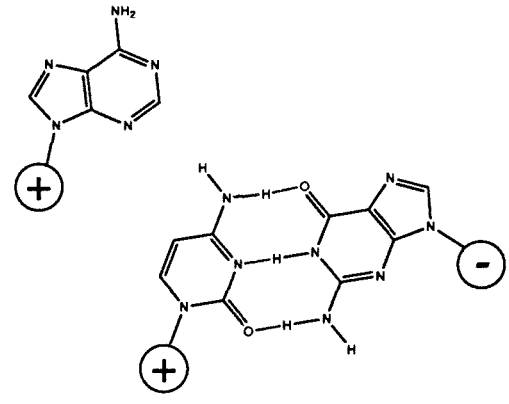
A•AT



A•GC



A•TA



A•CG

Figure 2.4: Structure of the four A•XY base triplets.

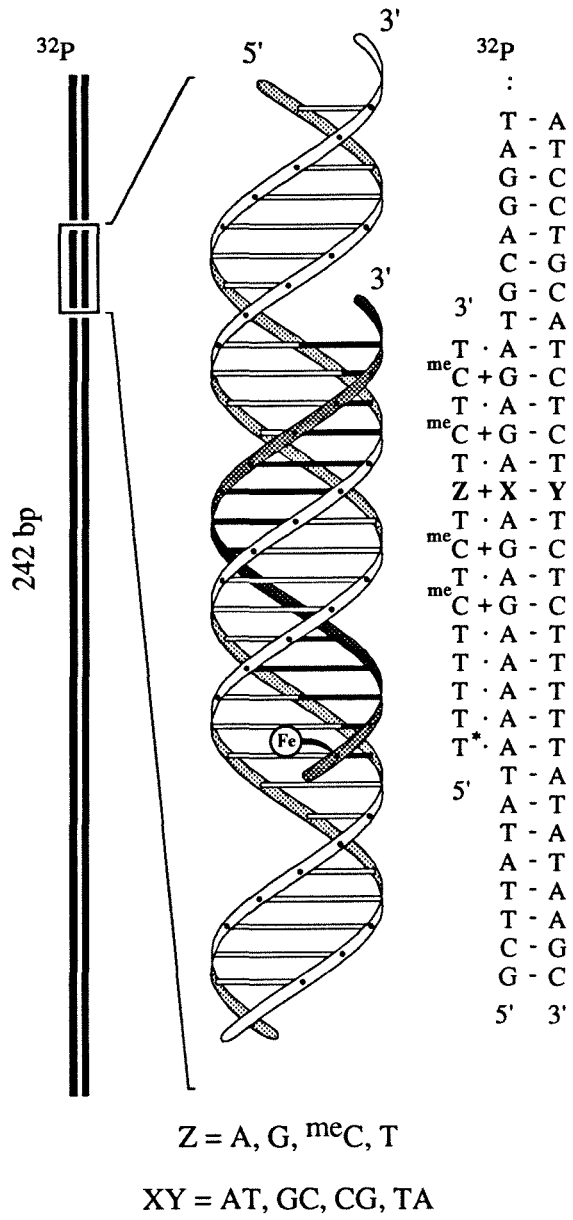


Figure 2.5: Ribbon model of the triple helix formed upon binding of the 15mer oligonucleotide probe to the 15 bp target site within a 242 bp restriction fragment. The Watson-Crick duplex strands are depicted as white ribbons while the oligonucleotide third strand is depicted as a dark ribbon.

report the use of QACT to determine the equilibrium binding constants and free energies of formation for triple helices containing the 16 natural pyrimidine•purine•pyrimidine base triplets.

Experimental Section

General. Sonicated, phenol extracted calf thymus DNA (Pharmacia) was dissolved in H_2O to a final concentration of 2.0 mM in base pairs and was stored at 0°C. Glycogen (20 mg/ml in water) and deoxynucleotide triphosphates were supplied by Boehringer Mannheim. α-^{32}P nucleotide triphosphates (3000 Ci/mmol) were purchased from Amersham or NEN-DuPont. α-^{35}S dATP for Sanger sequencing was from Amersham. Cerenkov radioactivity was counted on a Beckman LS 2801 scintillation counter. UV-visible spectroscopy was performed on a Hewlett-Packard 8452A diode array spectrophotometer. Restriction enzymes and the Klenow fragment of DNA polymerase were purchased from Boehringer-Mannheim and used in the buffers provided by the manufacturer. 5-methyl deoxycytidine phosphoramidite was purchased from Cruachem. Other phosphoramidites were purchased from ABI. Bis-tris was purchased from Sigma. All other chemicals were of reagent grade or better and were used as supplied.

Oligonucleotide preparation. T* phosphoramidite was prepared following established procedures[56] for use in Oligonucleotide synthesis. Oligonucleotides were synthesized on a 1 μmole scale using an ABI model 380B DNA synthesizer and the DNA synthesis protocols supplied by the manufacturer. After the final coupling, the dimethoxytrityl group was left on the oligonucleotide; deprotection and cleavage from the support was accomplished by treatment with 0.1 N NaOH at 55°C for 24 hours. The oligonucleotides were

desalted with a NAP 25 column (Pharmacia), frozen, and lyophilized. The residue was dissolved in 2 ml 100 mM triethyl amine acetate (TEAA) buffer (pH 7.0), filtered through a 0.45 μ cellulose acetate filter and purified by FPLC on a ProRPC HR 16/10 reversed phase column using a linear gradient from 0 - 40% acetonitrile in 100 mM TEAA pH 7.0. Fractions containing the fully deprotected oligonucleotide were combined and lyophilized. The residue was dissolved in water, frozen and lyophilized. Detritylation was accomplished by treatment of the residue with 1.5 ml of 40% acetic acid in H_2O (20 min, RT). The acetic acid was removed using a Speed Vac and the residue dissolved in 2 ml of 100 mM TEAA pH 7.0, filtered through a 0.45 μ filter and FPLC purified as above. Fractions containing the oligonucleotide were pooled, frozen and lyophilized. The residue was the lyophilized twice more from water. Prior to oligonucleotide packaging, the probes were desalted using a NAP 25 column. Oligonucleotides were packaged in 5 nmole aliquots according to the UV absorbance at 260 nm. Extinction coefficients were determined by adding the extinction coefficients of the individual monomers: T and T* , 8700 M^{-1}; $d^{me}C$, 5700 M^{-1} dG, 11,500 M^{-1}; dA, 15,400 M^{-1}. The resulting aliquots were lyophilized to dryness and stored at -20°C.

Plasmid preparation. The plasmids needed for the work reported in this paper were prepared by established procedures.[57] Briefly, oligonucleotides containing the desired insert sequences were synthesized, deprotected, and purified by reversed phase FPLC. The oligonucleotides were then annealed to give the desired insert duplex. The oligonucleotide duplex was added to *Bam* HI/*Hin*d III cleaved pUC 19; the resulting mixture was treated overnight with T4 DNA ligase. The ligation mixture was then used to transform Epicurean Coli XL1-Blue Supercompetent cells (Stratagene) according to the manufacturer's

1: 5'- *TTT TT^{me}C T^{me}CT **A**T^{me}C T^{me}CT -3'

2: 5'- *TTT TT^{me}C T^{me}CT **me**CT^{me}C T^{me}CT -3'

3: 5'- *TTT TT^{me}C T^{me}CT **G**T^{me}C T^{me}CT -3'

4: 5'- *TTT TT^{me}C T^{me}CT **T**T^{me}C T^{me}CT -3'

Figure 2.6: Sequences of the oligonucleotide probes prepared.

protocol. Transformants containing a plasmid with an insert were identified by α complementation, selected and grown overnight in 5 ml of liquid LB/ampicillin. The plasmids were isolated using a Qiagen miniprep kit and sequenced using a Sequenase 2.0 sequencing kit (U. S. Biochem). Large scale plasmid preparation was accomplished by the Qiagen maxiprep procedure. The sequence of the isolated plasmids was confirmed by repeating the sequencing protocol. The isolated plasmids were diluted to a concentration of ≈ 2.5 μg/μl and stored at -20°C.

DNA Labeling. The procedure used to prepare 3′ end labeled DNA was as follows: The initial restriction digest and 3′ end labeling were performed simultaneously. To 20 μg of plasmid was added: 10 - 20 units *Eco* RI, 10 units DNA polymerase, Klenow fragment, 10 μl (≈ 100 μCi) α-^{32}P dATP, 10 μl (≈ 100 μCi) α-^{32}P TTP, 10 μl 10x *Eco* RI reaction buffer, and sufficient water for a total reaction volume of 90 μl. After 2.5 hours incubation at 37°C, 10 μl of a solution of dNTPs (10 mM in each) was added, and the incubation allowed to continue for another 0.5 - 1 hour. Unincorporated radioactivity was removed using a NICK column (Pharmacia) and the linearized plasmid ethanol precipitated. After brief drying, the pellet was resuspended and digested with 20-30 units Pvu II (final reaction volume 100μl). After a 4 hour incubation at 37°C, 20 μl of 15% Ficoll loading buffer containing only bromphenol blue was added. The desired labeled fragment was isolated by preparative polyacrylamide gel electrophoresis (5% nondenaturing gel, 1:29 cross linked, 200 V, 1 hour). The desired 242 bp fragment was located by autoradiography; the band was excised from the gel and crushed. After addition of 1 ml elution buffer (10 mM Tris HCl pH 7.0, 10 mM EDTA, 250 mM NaCl, 0.1% SDS), the DNA was incubated overnight at 37°C. The elution buffer was filtered to remove polyacrylamide fragments and 700 μl of

isopropanol added to precipitate the DNA. The pellet was resuspended in 100 µl of 0.5 x TE pH 8.0. The resulting solution was extracted three times with phenol, once with 24:1 chloroform-isoamyl alcohol, and desalted on a NICK column using 0.5 x TE. The DNA was then precipitated with 50 µl 3M NaCl pH 4.0 and 1000 µl ethanol, resuspended in 100 µl 0.5 x TE and desalted once again with a NICK column in 0.5 x TE. The DNA was then counted and stored at -20°C until needed.

QACT experiments. Cleavage experiments were performed in sets of 18 microcentrifuge tubes. For each set of cleavage reactions, a cocktail containing 160 µl 5 x reaction buffer (500 mM NaCl; 50 mM Bis-tris, pH 7.0; 5 mM Spermine), 40 µl calf thymus DNA (2.0 mM bp), labeled restriction fragment (≈ 600,000 cpm), and sufficient water to make the total volume 640 µl was prepared. The cocktail was distributed 32 µl per tube, and necessary additional water added to the tubes. The dried aliquot of oligonucleotide was resuspended in 50 µl 250 mM $Fe(NH_4)_2(SO_4)_2 \cdot 6\ H_2O$ and allowed 15 min for equilibration. The oligonucleotide was then serially diluted and distributed to give final probe concentrations of 10 µM, 5 µM, 2.5 µM, 1 µM, 500 nM, 250 nM, 100 nM, 50 nM, 25 nM, 10 nM, 5 nM, 2.5 nM, 1 nM, 500 pM, 250 pM. Final equilibration volume was 40 µl. Final conditions were 100 mM NaCl, 25 mM Fe(II), 10 mM bis-tris•HCl pH 7.0, 250 µM spermine. After a 24 hour equilibration, cleavage reactions were initiated by the addition of 4 µl of 40 mM dithiothreitol (DTT). Cleavage reactions were allowed to proceed for 6 hrs and then terminated by the addition of 2 µl glycogen (2 mg/ml), 4 µl 3 M NaOAc, 0.1 M $MgCl_2$, pH 5.0, and 115 µl ethanol. The DNA was precipitated by storage at -20°C for 30 min, then centrifuged for 30 min at 4°C. The supernatant was removed and 60 µl of warm (50°C) water was added. The samples were then frozen, and the water removed

by lyophilization. The DNA was resuspended in 5 µl of formamide loading buffer (4:1 formamide/10 x TE) and counted. The samples were then heat denatured (90°C for 3 min) and loaded on a pre-electrophoresed polyacrylamide gel (8% polyacrylamide, 1:19 cross linked, 7 M urea, 0.4 mm thick). After loading, residual radioactivity in the microfuge tubes was counted. The polyacrylamide sequencing gels were run at ≈ 40 V/cm. When completed, the gel was transferred to filter paper, covered with plastic film and dried. After drying, the gel was pressed flat against a freshly erased phosphorimager screen.

Data collection and treatment. After 18-24 hours of exposure, the phosphor imaging plates were scanned using a Molecular Dynamics 400S PhosphorImager. The data were analyzed by performing volume integrations of the target site and the reference site using ImageQuant v. 3.15 software running on an AST Premium 386/33 computer. The results of the volume integration were transferred to a Macintosh computer, where subsequent data treatment was performed using Microsoft Excel v. 4.0 and Kaleidagraph v. 2.1 (Abelbeck Software). The site-specific cleavage was calculated using eq. 1:

$$I_{site} = I_{tot} - \lambda I_{ref} \qquad (1)$$

where I_{tot} and I_{ref} are the cleavage intensities in the site and reference blocks, respectively. For each experiment λ was determined by averaging the I_{tot} / I_{ref} values for data points with $\theta \approx 0$. A theoretical binding curve was fit to the data using eq. 2, with K_T and I_{sat} (the apparent maximum cleavage) as adjustable parameters:

$$I_{fit} = I_{sat}\left(\frac{K_T [O]_{tot}}{1 + K_T [O]_{tot}}\right) \qquad (2)$$

where $[O]_{tot}$ is the total concentration of the oligonucleotide probe. All data from a gel were used unless visual inspection revealed a flaw in the gel at either the target or reference sites , or the I_{site} value for a single data point was more than two standard errors away from the data points on either side. Data from experiments for which less than 80% of the lanes were usable were discarded. The goodness of fit of the binding curve was judged by the reduced chi-squared criterion.[58] Fits with values of $\chi^2_n > 1.5$ were discarded. Fits which met this criterion had correlation coefficients greater than 0.95.

Repeat experiments for a particular triplet used different serial dilutions of oligonucleotide probe prepared from a different aliquot of the probe. Uniquely prepared batches of 3'- end labeled DNA, buffer and calf thymus DNA were used. All K_T values reported in the text or tables are the means of three experimental observations plus or minus the standard error of the mean.

Results and Discussion

Methods. Previous work has demonstrated that the quantitative affinity cleavage titration can be used to measure the equilibrium association constant for the binding of an oligonucleotide probe to duplex DNA to form a local triple helix.

$$O + D \underset{}{\overset{K_T}{\rightleftharpoons}} T \xrightarrow{k_{cl}} P_{cl} \qquad (3)$$

A detailed description of this technique having been presented,[17] only a summary of the protocol is outlined below. Briefly, the fraction of the triple helix

Table 2.1: Association constants (K_T) for the formation of triple helical complexes containing the Z•XY triplets at 295 K 100 mM NaCl, 250 μM spermine, 10 mM bis tris pH 7.0.[a, b]

	G	T	^{me}C	A
AT	$2.1(\pm0.7) \times 10^5$	$5.6(\pm0.3) \times 10^7$	$3.8(\pm0.4) \times 10^6$	$4.8(\pm2.4) \times 10^5$
GC	$8.0(\pm2.8)\times10^5$	$4.3(\pm0.6) \times 10^5$	$4.1(\pm0.7) \times 10^7$	$2.7(\pm0.8) \times 10^5$
CG	$4.6(\pm2.5) \times 10^5$	$1.2(\pm0.4) \times 10^6$	$1.3(\pm0.7) \times 10^6$	$3.8(\pm0.8) \times 10^4$
TA	$3.2(\pm0.8) \times 10^6$	$2.0(\pm0.9) \times 10^5$	$1.5(\pm0.6) \times 10^5$	$2.1(\pm1.2) \times 10^5$

[a] K_T values are reported as the mean (± the standard error of the mean) of three measurements. The K_T values are reported in units of M^{-1}. [b] The identity of the base Z is indicated across the top of the columns; the identity of the Watson Crick base pair XY is indicated on the left side of the rows.

Table 2. Free energy (ΔG) of formation of triple helical complexes containing the Z•XY triplets. [a, b]

	G	T	meC	A
AT	-7.2 ± 0.2	-10.5 ± 0.1	-8.9 ± 0.1	-7.7 ± 0.3
GC	-8.0 ± 0.2	-7.6 ± 0.1	-10.3 ± 0.1	-6.0 ± 0.2
CG	-7.6 ± 0.3	-8.2 ± 0.2	-8.2 ± 0.3	-6.2 ± 0.2
TA	-8.8 ± 0.1	-7.2 ± 0.3	-7.0 ± 0.2	-7.2 ± 0.5

[a] Free energy values are calculated from the measured association constants at 295 K and are reported in kcal mol^{-1}. [b] The identity of the base in the third strand, Z, is indicated across the top row; the identity of the Watson Crick base pair, XY, is indicated to the left of the rows.

binding sites occupied by the oligonucleotide probe at a given concentration (θ_{app}) is proportional to amount of cleavage products (P_{cl}, eq. 3) which form after a fixed reaction time. After running a polyacrylamide gel to separate the cleavage products from the intact duplex (D) and triplex (T), an autoradiogram is obtained. The intensity of the cleavage bands proximal to the binding site (I_{tot}) can be related to the sequence specific cleavage at the site (I_{site}). Thus, for a set of DNA cleavage reactions in which the concentration of the oligonucleotide probe is varied while the solution conditions, concentration of duplex, reaction volume and reaction time are constant, the following relationship holds:

$$\theta_{app} = I_{site} \bullet I_{sat}^{-1} \qquad (4)$$

where I_{sat} is the intensity of cleavage at $\theta = 1$. Substitution of this expression into the relationship between θ_{app} and K_T, followed by rearrangement yields the following expression for I_{site}:

$$I_{site} = I_{sat}\left(\frac{K_T [O]_{tot}}{1 + K_T [O]_{tot}}\right) \qquad (5)$$

Thus, measuring the sequence-specific cleavage as a function of oligonucleotide concentration ($[O_{tot}]$) allows the determination of I_{sat} and K_T via the use of nonlinear least squares fitting of the empirical data.

Although the model used in the derivation of eq. 5 is a simple one (see eq. 3), which does not specifically address the choice of solution conditions, equilibration times or sequence composition effects, prior work has shown that, within experimental uncertainty, the quantitative affinity cleavage titration can give results identical to quantitative DNase I footprint titration methods.[17] The

substitution of T* for T also has no measurable effect on the binding constants.[17] In recent work QACT has been used to measure the association constants for triple helix formation under a variety of solution conditions.[22, 23] The technique has also been used to study systems other than the all DNA pyrimidine motif triple helix, including systems in which the composition of the sugar phosphate backbone has been varied,[59] systems with cooperatively binding oligonucleotides,[60-62] and a purine motif triple helix.[7]

Affinity. In this study, experimental conditions, such as temperature, salt concentration and pH, are chosen such that the lowest and highest equilibrium binding constants for the 16 triplets can be measured. We are primarily interested in the difference of free energy values ($\Delta\Delta G$), which are most relevant to the issue of specificity. The results of the affinity cleavage experiments are presented in Table 1. The values of the 16 equilibrium binding constants (K_T) range from 4×10^4 to 5×10^7 M^{-1}. Reports in the literature have suggested that four of the sixteen possible natural base triplets (T•AT, ^{me}C+GC, G•TA, and T•CG) are particularly stable.[31-39] An examination of the data confirms that this is the case, but that these interactions are not all of the same strength and that other "mismatched" triplets are similar in value (Table 2). As expected, the T•AT and ^{me}C+GC triplets are the most stable (ΔG = -10.5 and -10.3 kcal mol^{-1} respectively) interactions in the pyrimidine motif. In contrast, the G•TA base and T•CG base triplets are found to be of intermediate stability in this sequence context (ΔG = -8.8 and -8.2 kcal mol^{-1} respectively). While triple helices containing the T•AT and ^{me}C+GC base triplets at the variable position have approximately the same affinity, these triple helices are approximately 15-50 fold more stable than those which contain the G•TA and T•CG base triplets at that position.

Specificity. If specific local triple helices are to be useful structures for accomplishing the recognition of single sites in megabase DNA, it is important that the complex be stabilized via specific interactions and be sensitive to single base mismatches. The results presented here demonstrate that there is a unique best choice of base Z to place in the third strand of a triple helix to accomplish the sequence specific recognition of three of the four natural base pairs XY in duplex DNA (comparison across rows of Table 2). These interactions are also specific in the reciprocal sense, i. e. for three of the four bases Z studied, one of the four Z•XY triplets formed is particularly stable (comparison down columns of Table 2). For example, the T•AT triplet is 1.6 kcal mol^{-1} more stable than the next most stable Z•AT triplet. The T•AT triplet is 2.3 kcal mol^{-1} more stable than the T•CG triplet, the next most stable T•XY triplet. Similarly, ^{me}C+GC is 1.4 kcal mol^{-1} more stable than any other ^{me}C•XY and 2.3 kcal mol^{-1} more stable than any other Z•GC triplet.

Recognition of the TA and CG base pairs is more problematic. The G•TA triplet is 1.5 kcal mol^{-1} more stable than any other Z•TA triplet and 0.8 kcal mol^{1} more stable than any other G•XY triplet. These results show that G does offer a relatively specific means of recognizing a single TA base pair within this sequence context.[63] The low free energy value of the G•TA base triplet (relative to the T•AT base triplet) may result from the combination of the formation of one strong hydrogen bond, rather than the two seen in the T•AT and ^{me}C+GC base triplets, and the known distortion away from planarity of and loss of continuous stacking caused by the G•TA triplet.[51] Since the sequence used in this study provides the ideal flanking base pairs for the G•TA triplet,[63] the relative weakness of the G•TA interaction suggests that a novel base that can recognize TA base pairs with higher affinity than G

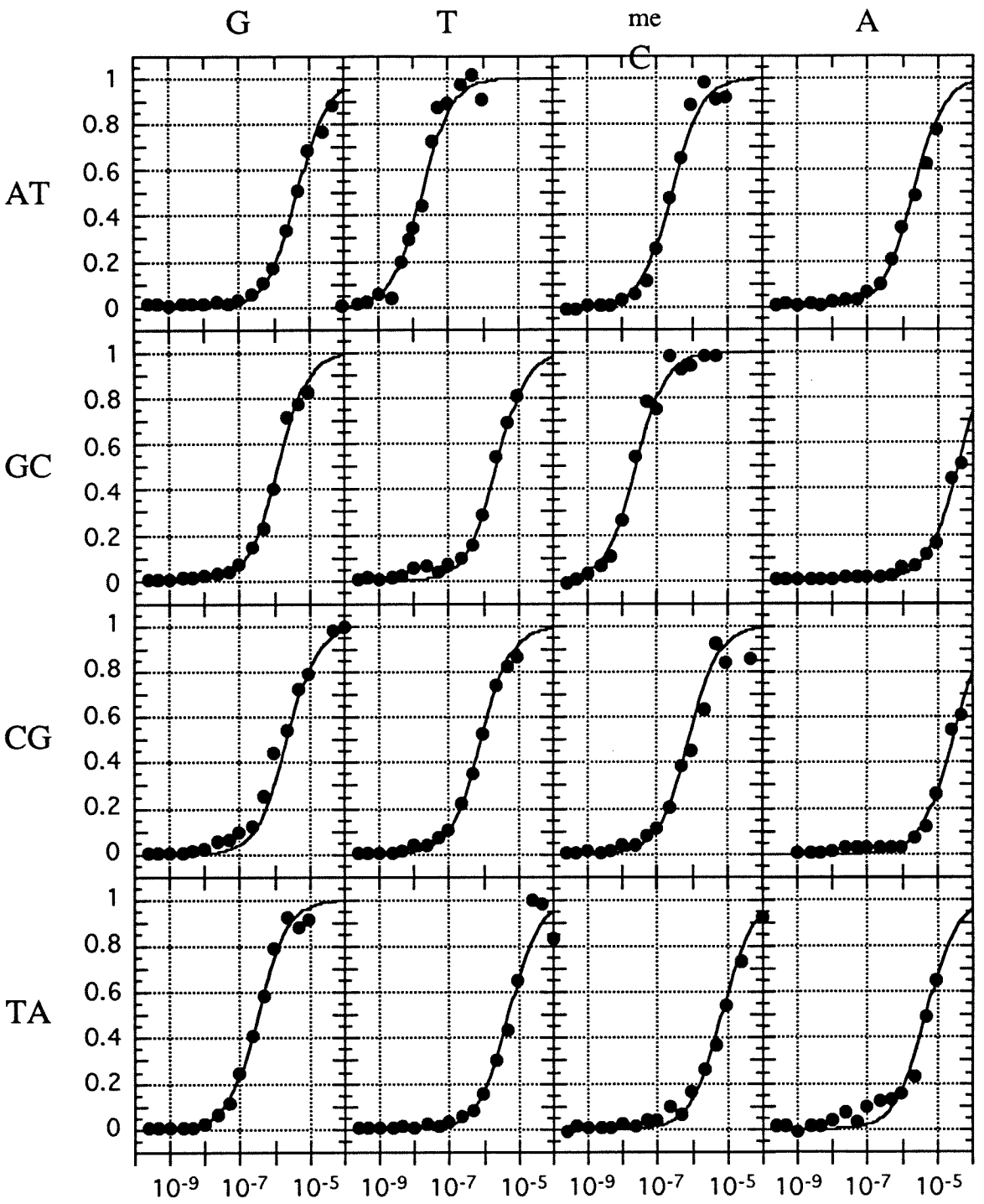


Figure 2.7: Binding isotherms for the sixteen Z•XY triplets studied. The base Z is indicated across at the top of the columns. The base pair XY is indicated to the left of each row. Each isotherm represents the average of three experiments conducted at 22°C, 10 mM bis tris, pH 7.0, 100 mM NaCl, 250 μM spermine.

is needed, particularly if sequences containing more than one TA base pair are to be targeted.

The data in Tables 1 and 2 reveal that, under these conditions and in this sequence context, there is no particularly favorable Z•CG triplet. While a comparison across the CG row of Table 2 shows that both T and ^{me}C form triplets with CG that are of intermediate stability, these interactions are not specific, since both T and ^{me}C can form other, energetically preferred base triplets (comparison down the columns of Table 2). This suggests that, while either of T or ^{me}C might be used to tolerate the presence of one CG base pair in a target site, such use of these bases may result in a decrease in the overall specificity of the oligonucleotide. Thus the specific recognition of the CG base pair remains a design-synthesis challenge.

Earlier, qualitative work had identified the major interactions (T•AT, ^{me}C+GC, G•TA) reported above as being energetically preferred. The relative ranking of the mismatched base triplets differs from one study to the next. This variation is probably due to the differences in salt conditions, pH and sequence studied. The earlier studies examined the stability of triple helices containing various mismatches by monitoring the thermal denaturation of triple helices by either optical spectroscopy,[34-36] NMR spectroscopy,[64] or gel mobility shift assay.[37] These studies have typically omitted polyvalent cations from the buffers used in the study. Since the presence of polyvalent cations favors the formation of triple helices, these studies have been performed at a lower pH in order to compensate for this omission.

Another complication in comparing the work reported here to the literature is the effect of sequence composition on the stability of various base triplets. It has been established that the identity of the neighboring triplets has

an effect, sometimes quite large, on the stability of a given triplet.[63] This effect is, in part, due to the importance of stacking interactions in stabilizing the triple helix. The need to protonate C or ^{me}C residues means that binding adjacent GC base pairs, which requires protonation at adjacent residues in the third strand, likely is energetically disfavored, while NMR studies have shown that some of the non-pyrimidine bases used in this motif interact with the duplex DNA either by intercalation or by forming additional hydrogen bonds with adjacent Watson-Crick base pairs of the double helix. Given the importance of the mechanisms such as those outlined above, the stability of a particular base triplet is likely to be a function of the overall sequence context in which the triplet is formed.

The association constants for the sixteen pyrimidine•purine• pyrimidine base triplets involving the natural bases have been determined. These results show that the T•AT and ^{me}C+GC base triplets are significantly more stable than the fourteen other triplets studied. The G•TA base triplet possesses a stability between that of the T•AT and ^{me}C+GC base triplets and the worst mismatches. Further, the T•AT, ^{me}C+GC, and G•TA base triplets are *highly specific interactions.* This means that oligonucleotides containing T, ^{me}C, and G are capable of recognizing target sites composed of AT, GC and TA base pairs in a specific manner; the energetic penalty for such an oligonucleotide binding to a site with a single mismatch is at least 0.5 kcal mol^{-1}. We would emphasize that our results are for one sequence composition context and that the dependence of the energetics of triple helix formation on sequence composition remains a key issue to be elucidated.

References

(1) Moser , H. E.; Dervan, P. B. *Science* **1987**, *238*, 645-650.

(2) LeDoan, T.; Perrouault, L.; Praseuth, D.; Habhoub, N.; Decout, J.-L.; Thuong, N. T.; Lhomme, J.; Hélène, C. *Nucleic Acids Res.* **1987**, *15*, 7749-7760.

(3) Cooney, M.; Czernuszewicz, G.; Postel, E. H.; Flint, S. J.; Hogan, M. E. *Science* **1988**, *241*, 456-459.

(4) Felsenfeld, G.; Davies, D.; Rich, A. *J. Am. Chem. Soc.* **1957**, *79*, 2023-2024.

(5) Beal, P. A.; Dervan, P. B. *Science* **1991**, *251*, 1360-1363.

(6) Beal, P. A.; Dervan, P. B. *Nucleic Acids Res.* **1992**, *20*, 2773-2776.

(7) Stilz, H. U.; Dervan, P. B. *Biochemistry* **1993**, *32*, 2177-2185.

(8) Malkov, V. A.; Voloshin, O. N.; Soyfer, V. N.; Frank-Kamenetskii, M. D. *Nucleic Acids Res.* **1993**, *21*, 585-591.

(9) Jayasena, S. D.; Johnston, B. H. *Biochemistry* **1992**, *31*, 320-327.

(10) Beal, P. A.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 4976-4982.

(11) Jayasena, S. D.; Johnston, B. H. *Nucl. Acids Res.* **1992**, *20*, 5279-5288.

(12) Jayasena, S. D.; Johnston, B. H. *Biochemistry* **1993**, *32*, 2800-2807.

(13) Roberts, R. W.; Crothers, D. M. *Science* **1992**, *258*, 1463-1466.

(14) Han, H.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA*, **1993**, *90*, 3806-3810.

(15) Escude, C.; François, J. C.; Sun, J. S.; Ott, G.; Sprinzl, M.; Garestier, T.; Hélène, C. *Nucleic Acids Res.* **1993**, *21*, 5547-5553.

(16) Han, H.; Dervan, P. B. *Nucleic Acids Res.* **1994**, *22*, 2837.

(17) Singleton, S. F.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 6957-6965.

(18) Mergny, J.-L.; Sun, J.-S.; Rougée, M.; Montenay-Garestier, T. Barcelo, F.; Chomilier, J.; Hélène, C. *Biochemistry* **1991**, *30*, 9791-9798.

(19) Rougée, M.; Faucon, B.; Mergny, J. L.; Barcelo, F.; Giovannageli, C.; Garestier, T.; Hélène, C. *Biochemistry* **1992**, *31*, 9269-9278.

(20) Povsic, T. J.; Dervan, P. B. *J. Am. Chem. Soc.* **1989**, *111*, 3059-3061.

(21) Xodo, L. E.; Manzini, G.; Quadrifoglio, F.; van der Marel, G. A.; van Boom, J. H. *Nucleic Acids Res.* **1991**, *19*, 5625-5631.

(22) Singleton, S. F.; Dervan, P. B. *Biochemistry* **1992**, *31*, 10995-11003.

(23) Singleton, S. F.; Dervan, P. B. *Biochemistry* , **1993**, *32*, 13171-13179.

(24) Thomas, T.; Thomas, T. J. *Biochemisty*, **1993**, *32*, 14068-14074.

(25) Strobel, S. A.; Doucette-Stamm, L. A.; Riba, L.; Housman, D. E.; Dervan, P. B. *Science* **1991**, *254*, 1639-1642.

(26) Maher, L. J.; Dervan, P. B.; Wold, B. *Science* **1989**, *245*, 725-730.

(27) Collier, D. A.; Thuong, N. T.; Hélène, C. *J. Am. Chem. Soc.* **1991**, *113*, 1457-1458.

(28) Maher, L. J.; Dervan, P. B.; Wold, B. *Biochemistry* **1992**, *31*, 70-81.

(29) Grigoriev, M.; Praseuth, D.; Robin, P.; Hemar, A.; Saison-Behmoaras, T.; Dautry-Varsat, A.; Thuong, N. T.; Hélène, C.; Harel-Bellan, A. *J. Biol. Chem.* **1992**, *267*, 3389-3395.

(30) Horne, D.A.; Dervan, P. B. *J. Am. Chem. Soc.* **1990**, *112*, 2435-2437.

(31) Griffin, L. C.; Dervan, P. B. *Science* **1989**, *245*, 967-971.

(32) Beal, P. A.; Dervan, P. B. *Nucleic Acids Res.* **1992**, *20*, 2773-2776.

(33) Shimizu, M.; Inoue, H.; Ohtsuka, E. *Biochemistry*, **1994**, *33*, 606-613.

(34) Mergny, J.-L.; Sun, J.-S.; Rougée, M.; Montenay-Garestier, T. Barcelo, F.; Chomilier, J.; Hélène, C. *Biochemistry* **1991**, *30*, 9791-9798.

(35) Miller, P. S.; Cushman, C. D.; *Biochemistry*, **1993**, *32*, 2999.

(36) Fossella, J. A.; Kim, Y. J.; Shih, H.; Richards, E. G.; Fresco, J. R. *Nucleic Acids Res.* **1993**, *21*, 4511-4515.

(37) Yoon, K.; Hobbs, C. A.; Koch, J.; Sardaro, M.; Kutny, R.; Weis, A. *Proc. Natl. Acad. Sci. USA*, **1992**, *89*, 3840-3844.

(38) Belotserkovskii, B. P.; Veselkov, A. G.; Filippov, S. A.; Dobrynin, V. N.; Mirkin, S. M.; Frank-Kamenetskii, M. D. *Nucleic Acids Res.* **1990**, *18*, 6621-6624.

(39) Xodo, L. E.; Alunni-Fabbroni, M.; Manzini, G.; Quadrifoglio, F. *Eur. J. Biochem.* **1993**, *212*, 395-401.

(40) de los Santos, C.; Rosen, M.; Patel, D. J. *Biochemistry* **1989**, *28*, 7282-7289.

(41) Rajagopal, P.; Feigon, J. *Nature* **1989**, *339*, 637-640.

(42) Rajagopal, P.; Feigon, J. *Biochemistry* **1989**, *28*, 7859-7870.

(43) Live, D. H.; Radhakrishnan, I.; Misra, V.; Patel, D. J. *J. Am. Chem. Soc.* **1991**, *113*, 4687-4688.

(44) Ono, A.; Ts'o, P. O. P.; Kan, L.; *J. Am. Chem. Soc.* **1991**, *113*, 4032-4033.

(45) Griffin,L. C.; Kiessling, L. L.; Beal, P. A.; Gillespie, P.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 7976-7982.

(46) Miller, P. S.; Bhan, P.; Cushman, C. D.; Trapane, T. L. *Biochemistry* **1992**, *31*, 6788-6793.

(47) Krawczyk, S. H.; Milligan, J. F.; Wadwani, S.; Moulds, C.; Froehler, B. C.; Matteucci, M. D. *Proc. Natl. Acad. Sci., USA*, **1992**, *89*, 3761-3764.

(48) Griffin, L. C.; Dervan, P. B. *Science* **1989**, *245*, 967-971.

(49) Radhakrishnan, I.; Gao, X.; de los Santos, C.; Live, D.; Patel, D. J. *Biochemistry*, **1991**, *30*, 9022-9030.

(50) Wang, E.; Malek, S.; Feigon, J. *Biochemistry* **1992**, *31*, 4838-4846.

(51) Radhakrishnan, I.; Patel, D. J.; Veal, J. M.; Gao, X. *J. Am. Chem. Soc.* **1992**, *114*, 6913-6915.

(52) Koh, J. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 1470-1478.

(53) Radhakrishnan, I.; Patel, D. J.; Priestley, E. S.; Nash, H.; Dervan, P. B. *Biochemistry* , **1993**, *32*, 11228-11234.

(54) Griffin, L. C.; Kiessling, L. L.; Beal, P. A.; Gillespie, P.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 7976-7982.

(55) Koshlap, K. M.; Gillespie, P.; Dervan, P. B.; Feigon, J. *J. Am. Chem. Soc.* **1993**, *115*, 7908-7909.

(56) Dreyer, J. B.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA*, **1985**, *82*,, 968-972.

(57) Sambrook, J.; Fritsch, E. F.; Maniatis, T. *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press: Cold Spring Harbor, N. Y., 1989.

(58) Bevington, P. R. *Data Reduction and Error Analysis for the Physical Sciences*; McGraw-Hill: New York, 1969.

(59) Han, H.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA*, **1993**, *90*, 3806-3810.

(60) Distefano, M. D.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA*, **1993**, *90*, 1179-1183.

(61) Colocci, N.; Distefano, M. D.; Dervan, P. B. *J. Am. Chem. Soc.* **1993**, *115*, 4468-4473.

(62) Colocci, N.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 785-786.

(63) Kiessling, L. L. Griffin, L. C.; Dervan. P. B. *Biochemistry* **1992**, *31*, 2829-2834.

(64) Macaya, R.; Gilbert, D. E.; Malek, S.; Sinsheimer, J. S.; Feigon, J. *Science*, **1991**, *254*, 270-274.

CHAPTER THREE

Sequence Dependent Energetics of Triple Helix Formation: Nearest Neighbor Analysis of Oligonucleotides Containing Deoxythymidine and 5-Methyldeoxycytidine

Introduction

Oligonucleotide-directed triple helix formation is a versatile means of accomplishing the sequence specific recognition of double helical DNA.[1-3] Triple helices can be classified into two motifs: those in which the third strand is primarily composed of pyrimidine bases,[4] and those in which the third strand is primarily composed of purine bases.[3, 5-8] Triple helices composed of combinations of these two motifs can also be formed.[9-12] Triple helix formation in the more thoroughly characterized pyrimidine motif is sensitive to the composition (DNA or RNA) of the three strands,[13-15] length of the third strand,[1, 16] single base mismatches,[16-18] pH,[19-21] salt conditions of the buffer solution,[22] and the structure of the polyvalent cations present in solution.[23] Oligonucleotide-directed triple helix formation has been used to mediate single site cleavage of human chromosomal DNA,[24] as well as interfere with the function of DNA binding proteins.[25-28] The ability to target a broad range of DNA sequences,[1, 29] and the high stability of the resulting local triple helical

structures make this a powerful technique for the recognition of single sites within megabase segments of double helical DNA.

In an effort to understand the interactions which give rise to specific recognition of DNA in the triple helix, the specificity afforded by the natural bases in triple helical recognition has been studied by affinity cleavage,[30-32] melting temperature studies,[33-35] gel mobility shift assays,[36] and intra-molecular triple helix formation.[37, 38] Some of the interactions shown by these studies to promote triple helix formation have been studied at higher resolution by NMR spectroscopy. Hoogsteen type hydrogen bonds have been observed between T residues in the third strand cytosine or 5-methyl cytosine in the third strand and GC base pairs in the duplex.[39-41] Recent studies have extended the pyrimidine motif through the use of both non-pyrimidine and non-natural bases.[42-45] The best characterized of these are the G•TA, D_3•TA, and P_1•GC triplets, for which NMR studies have been completed.[30, 43, 46-51]

The study of the energetics of triple helix formation have not proceeded as rapidly as the structural studies. An early effort used affinity cleavage to explore the relative stability of base triplets consisting of the four natural bases and inosine across the four natural base pairs.[30] Other work made use of optical melting experiments[33-35] or gel mobility shift experiments[36] to explore the relative stability of triple helices containing various triplets. The introduction of the quantitative affinity cleavage titration (QACT) has allowed the facile determination of the equilibrium binding constants for the third strand of triple helical complexes.[16, 21, 22] QACT has since been used to explore the effects of pH,[21] cation valency and concentration,[22] and a variety of mismatches[52] on triple helix formation. The triple helix has also been studied by quantitative DNase footprint titration.[10, 16, 53-55] These two techniques have been used to measure

binding constants that are the same, within experimental uncertainty, for several oligonucleotides.[16]

It has long been observed that the sequence composition of the oligonucleotide plays a key role in determining the stability of the triple helix which is formed.[56] In this paper, we determine the relative energetic contributions of the various dinucleotide units and present an algorithm for predicting the stability of triple helices formed at 22 °C, pH 7.0, 10 mM bis-tris, 100 mM NaCl, 250 μM spermine.

Experimental Section

General. Sonicated, phenol extracted calf thymus DNA (Pharmacia) was dissolved in H_2O to a final concentration of 2.0 mM in base pairs and was stored at 0°C. Glycogen (20 mg/ml in water) and deoxynucleotide triphosphates were supplied by Boehringer Mannheim. α-^{32}P nucleotide triphosphates (3000 Ci/mmol) were purchased from Amersham or NEN-DuPont. α-^{35}S dATP for Sanger sequencing was purchased from Amersham. Cerenkov radioactivity was counted on a Beckman LS 2801 scintillation counter. UV-visible spectroscopy was performed on a Hewlett-Packard 8452A diode array spectrophotometer. Restriction enzymes *Eco*R I, *Pvu* II and the Klenow fragment of DNA polymerase were purchased from Boehringer-Mannheim, and restriction enzyme *Xmn* I was purchased from New England Biolabs. Restriction enzymes were used in the buffers provided by the manufacturer. 5-methyldeoxycytidine phosphoramidite was purchased from Cruachem or Glen Research. Other phosphoramidites were purchased from ABI and Glen Research. Bis-tris was purchased from Sigma. All other chemicals were of reagent grade or better and were used as supplied.

Oligonucleotide preparation. Oligonucleotides were synthesized on a 1 μmole scale using an ABI model 380B DNA synthesizer and the DNA synthesis protocols supplied by the manufacturer. After the final coupling, the dimethoxytrityl group was left on the oligonucleotide; deprotection and cleavage from the support was accomplished by treatment with concentrated aqueous NH_3 at 55°C for 24 hours. The oligonucleotides were desalted with a NAP 25 column (Pharmacia), frozen, and lyophilized. The residue was dissolved in 2 ml 100 mM triethyl ammonium acetate (TEAA) buffer (pH 7.0), filtered through a 0.45 μ cellulose acetate filter and purified by FPLC on a ProRPC HR 16/10 reversed phase column using a linear gradient from 0 - 40% acetonitrile in 100 mM TEAA pH 7.0. Fractions containing the fully deprotected oligonucleotide were combined and lyophilized. The residue was dissolved in water, frozen and lyophilized. Detritylation was accomplished by treatment of the residue with 1.5 ml of 40% acetic acid in H_2O (20 min, RT). The acetic acid was removed *in vacuo* and the residue dissolved in 2 ml of 100 mM TEAA pH 7.0, filtered through a 0.45 μ filter and FPLC purified as above. Fractions containing the oligonucleotide were pooled, frozen and lyophilized. The residue was the lyophilized twice more from water. Prior to oligonucleotide packaging, the probes were desalted using a NAP 25 column. Oligonucleotides were packaged in 5 nmole aliquots according to the UV absorbance at 260 nm. Extinction coefficients were determined by adding the extinction coefficients of the individual monomers: T, 8700 M^{-1}; $d^{me}C$, 5700 M^{-1} dG, 11,500 M^{-1}; dA, 15,400 M^{-1}. The resulting aliquots were lyophilized to dryness and stored at -20°C.

Plasmid preparation. The plasmids needed for the work reported in this paper were prepared by established procedures.[57] Briefly, oligonucleotides containing the desired insert sequences were synthesized, deprotected, and

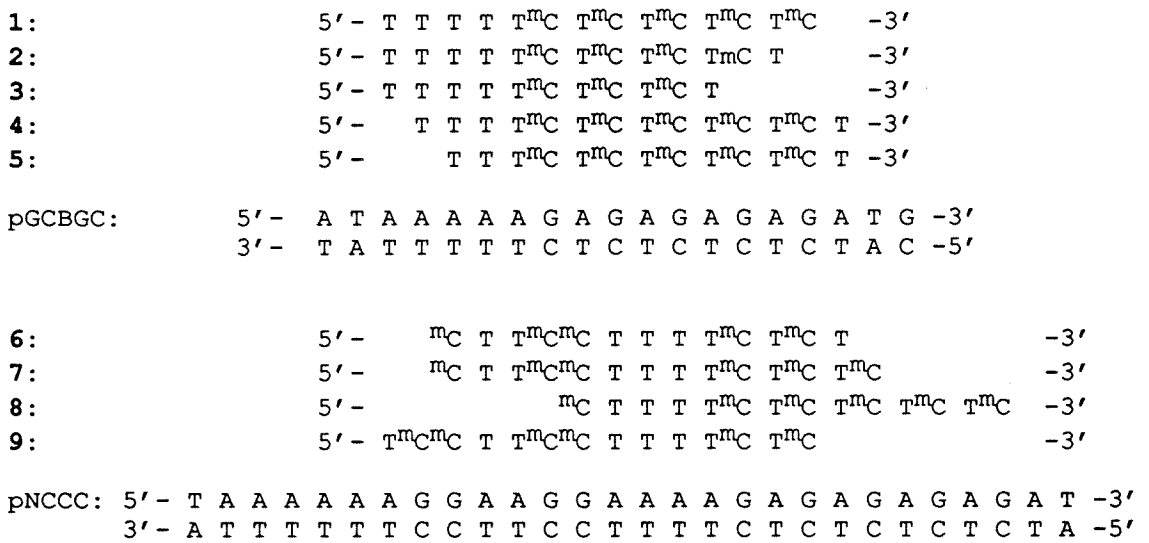


Figure 3.1: Sequences of the plasmid inserts and oligonucleotides used.

purified by reversed phase FPLC. The oligonucleotides were then annealed to give the desired insert duplex. The oligonucleotide duplex was added to *Bam* HI/*Hin*d III cleaved pUC 19; the resulting mixture was treated overnight with T4 DNA ligase. The ligation mixture was then used to transform Epicurean Coli XL1-Blue Supercompetent cells (Stratagene) according to the manufacturer's protocol. Transformants containing a plasmid with an insert were identified by α complementation, selected, and grown overnight in 5 ml of liquid LB/ampicillin. The plasmids were isolated using a Qiagen miniprep kit and sequenced using a Sequenase 2.0 sequencing kit (U. S. Biochem). Large scale plasmid preparation was accomplished by the Qiagen maxiprep procedure. The sequence of the isolated plasmids was confirmed by repeating the sequencing protocol. The isolated plasmids were diluted to a concentration of ≈ 2.5 μg/μl and stored at -20°C.

DNA Labeling. The procedure used to prepare 3′ end labeled DNA was as follows: The initial restriction digest and 3′ end labeling were performed simultaneously. To 20 μg of plasmid was added: 10 - 20 units *Eco*R I, 10 units DNA polymerase, Klenow fragment, 10 μl (≈ 100 μCi) α-^{32}P dATP, 10 μl (≈ 100 μCi) α-^{32}P TTP, 10 μl 10x *Eco* R I reaction buffer, and sufficient water for a total reaction volume of 90 μl. After 2.5 hours incubation at 37°C, 10 μl of a solution of dNTPs (10 mM in each) was added, and the incubation was allowed to continue for another 0.5 - 1 hour. Unincorporated radioactivity was removed using a NICK column (Pharmacia) and the linearized plasmid ethanol precipitated. After brief drying, the pellet was resuspended and digested with 20-30 units *Pvu* II or *Xmn* I (final reaction volume 100μl). After a 4 hour incubation at 37°C, 20 μl of 15% Ficoll loading buffer containing only bromphenol blue was added. The desired labeled fragment was isolated by preparative polyacrylamide gel

electrophoresis (5% nondenaturing gel, 1:29 cross linked, 200 V, 1 hour). The desired restriction fragment (242 bp, *Eco*R I/*Pvu* II; 840 bp, *Eco*R I/*Xmn* I) was located by autoradiography; the band was excised from the gel and crushed. After addition of 1 ml elution buffer (10 mM Tris HCl pH 7.0, 10 mM EDTA, 250 mM NaCl, 0.1% SDS), the DNA was incubated overnight at 37°C. The elution buffer was filtered to remove polyacrylamide fragments and 700 µl of isopropanol added to precipitate the DNA. The pellet was resuspended in 100 µl of 0.5 x TE pH 8.0. The resulting solution was extracted three times with phenol, once with 24:1 chloroform-isoamyl alcohol, and desalted on a NICK column using 0.5 x TE. The DNA was then precipitated with 50 µl 3M NaCl pH 4.0 and 1000 µl ethanol, resuspended in 100 µl 0.5 x TE and desalted once again with a NICK column in 0.5 x TE. The DNA was then counted and stored at -20°C until needed.

Quantitative DNase Footprint Titration Experiments. Cleavage experiments were performed in sets of 16 microcentrifuge tubes. For each set of cleavage reactions, a cocktail containing 160 µl 5 x reaction buffer (500 mM NaCl; 50 mM Bis-tris, pH 7.0; 5 mM Spermine), 40 µl calf thymus DNA (2.0 mM bp), labeled restriction fragment (≈ 600,000 cpm), and sufficient water to make the total volume 680 µl was prepared. The cocktail was distributed 36 µl per tube, and necessary additional water added to the tubes. The dried aliquot of oligonucleotide was resuspended in 50 µl H_2O and allowed 15 min for equilibration. The oligonucleotide was then serially diluted and distributed to give final probe concentrations of 10 µM, 5 µM, 2.5 µM, 1 µM, 500 nM, 250 nM, 100 nM, 50 nM, 25 nM, 10 nM, 5 nM, 2.5 nM, 1 nM, 500 pM, 250 pM. Final equilibration volume was 40 µl. Final conditions were 100 mM NaCl, 10 mM bis-tris•HCl pH 7.0, 250 µM spermine. After a 24 hour equilibration, the DNase

footprinting reaction was performed. A stock solution containing DNase (2.5 mU/μl), 100 mM $CaCl_2$, 100 mM $MgCl_2$, and 10 μM non specific oligonucleotide was prepared and distributed (4 μl/tube). The cleavage reactions were allowed to proceed for 7-8 minutes and then terminated by the addition of 2 μl glycogen (2 mg/ml), 4 μl 3 M NaOAc, 0.1 M $MgCl_2$, pH 5.0, and 115 μl ethanol. The DNA was precipitated by storage at -20°C for 30 min, then centrifuged for 30 min at 4°C. The supernatant was removed and 60 μl of warm (50°C) water was added. The samples were then frozen, and the water removed by lyophilization. The DNA was resuspended in 5 μl of formamide loading buffer (4:1 formamide/10 x TE) and counted. The samples were then heat denatured (90°C for 3 min) and loaded on a pre-electrophoresed polyacrylamide gel (8% polyacrylamide, 1:19 cross linked, 7 M urea, 0.4 mm thick). After loading, residual radioactivity in the microfuge tubes was counted. The polyacrylamide sequencing gels were run at ≈ 40 V/cm. When completed, the gel was transferred to filter paper, covered with plastic film and dried. After drying, the gel was pressed flat against a freshly erased phosphorimager screen.

Data collection and treatment. After 18-24 hours of exposure, the phosphor imaging plates were scanned using a Molecular Dynamics 400S PhosphorImager. The data were analyzed by performing volume integrations of the target site and a reference site (at which the DNase activity is invariant) using ImageQuant v. 3.15 software running on an AST Premium 386/33 computer. The results of the volume integration were transferred to a Macintosh computer, where subsequent data treatment was performed using Microsoft Excel v. 4.0 and Kaleidagraph v. 2.1 (Abelbeck Software).

The apparent occupancy of the site in a given lane was calculated from the cleavage data using the equation of Ackers and co-workers:[53-55]

$$Y_{app} = 1 - \frac{I_{site} / I_{ref}}{I°_{site} / I°_{ref}} \qquad (1)$$

where I_{site} and I_{ref} are the cleavage intensities at the target site and the reference block respectively and $I°_{site}$ and $I°_{ref}$ are the corresponding intensities from a control lane to which no oligonucleotide has been added. The data were then fit using an unweighted nonlinear least squares procedure with three adjustable parameters. The equation used in the fitting procedure was:

$$Y_{app} = \theta_{min} + (\theta_{max} - \theta_{min}) \frac{K_T [O]_{tot}}{1 + K_T [O]_{tot}} \qquad (2)$$

where θ_{min} and θ_{max} represent the experimentally observed values of Y_{app} at $\theta = 0$ and $\theta = 1$ respectively, K_T is the association constant for triple helix formation, and $[O]_{tot}$ is the total oligonucleotide concentration.

All data from a gel were used unless visual inspection revealed a flaw in the gel at either the target or reference sites, or the Y_{app} value for a single data point was more than two standard errors away from the data points on either side. Data from experiments for which less than 80% of the lanes were usable were discarded. The standard error of a given data point was estimated to be 20% of Y_{app}, with a minimum value of 0.1. The goodness of fit of the binding curve was judged by the reduced chi-squared criterion.[58] Fits with values of $\chi^2_n > 1.5$ were discarded. Fits which met this criterion had correlation coefficients greater than 0.95.

Repeat experiments for a particular oligonucleotide used different serial dilutions of the oligonucleotide probe prepared from a different aliquot of the probe. Uniquely prepared batches of 3'- end labeled DNA, buffer and calf

thymus DNA were used. All K_T values reported in the text or tables are the means of at least three experimental observations plus or minus the standard error of the mean.

Results and Discussion

Previous work has demonstrated that quantitative DNase footprint titration can be used to measure the equilibrium association constant for the binding of an oligonucleotide probe to duplex DNA to form a local triple helix.

$$O + D \underset{}{\overset{K_T}{\rightleftharpoons}} T \xrightarrow{k_{cl}} P_{cl} \qquad (3)$$

A detailed description of this technique having been presented,[53-55] only a summary of the protocol is outlined below. Briefly, the degree of protection from DNase cleavage afforded by triple helix formation is inversely proportional to the amount of cleavage products (P_{cl}, eq. 3) which form after a fixed reaction time. After running a polyacrylamide gel to separate the cleavage products from the intact duplex (D) and triplex (T), an autoradiogram is obtained. The amount of radioactivity in the cleavage bands at the target site and a distant reference site is quantitated and the apparent fraction of the site which is occupied (Y_{app}) can be calculated using equation 1. The Y_{app} data are fit and normalized to yield the final binding isotherm for a given gel. The association constants reported for a given oligonucleotide are the average of three or four independent measurements and are reported in Table 3.1.

Two different plasmids with sites amenable for triple helix formation were prepared for this study. Plasmid pGCBGC is a pUC 19 derivative which contains

a 15 base pair homopurine run. Plasmid pNCCC is a pUC 18 derivative which contains a 26 base pair homopurine run. The sequence of these inserts and the nine oligonucleotides used in this study are given in Figure 3.1. The sequence composition of the oligonucleotides at the mononucleotide level is presented in Table 3.2 (Columns 2 and 3). A comparison of the data in Tables 3.1 and 3.2 clearly shows that the sequence composition affects the association constant for triple helix formation. Oligonucleotides which are composed of the same number of T and ^{me}C residues have, in some cases, binding constants which vary by a factor of more than 30 (Table 3.3). This clearly suggests that the specific sequence of the bases in the oligonucleotides plays an important role in determining the stability of the triple helical complex.

The free energy of the binding event for each oligonucleotide is determined from the association constant. It is assumed that this free energy is sum of the energetic contributions of each dinucleotide stack in the triple helix, plus terms for the nucleotide at the 5′ and 3′ ends of the third strand:

$$\Delta G_{tot} = \sum_{XX} N_{XX} \Delta G_{XX} \qquad (4)$$

where N_{XX} is the number of times dinucleotide XX occurs in the oligonucleotide. For example N_{TmC} is the number of times the 5′-T^{me}C-3′ step is present in the oligonuleotide, and N_{5T} is equal to one if the nucleotide at the 5′ end of the oligonucleotide is a T and zero if it is not. Similarly, ΔG_{XX} is the contribution to the total free energy of binding made by an individual dinucleotide stack XX. This approach is similar to that used in the study of secondary structure of RNA[59, 60] and duplex DNA.[61, 62] Since there are eight ΔG_{XX} components, at least

Table 3.1: Measured association constants and free energies of triple helix formation.

Oligo	Association constant K_T[a]	ΔG[b]
1	3.1 (±0.5) x 10^8	-11.6 ±0.1
2	1.9 (±0.3) x 10^8	-11.3 ±0.1
3	3.8 (±0.7) x 10^6	-9.0 ±0.1
4	2.6 (±0.5) x 10^8	-11.5 ±0.1
5	6.0 (±0.5) x 10^7	-10.6 ±0.1
6	2.5 (±0.6) x 10^6	-8.7 ±0.1
7	7.5 (±0.8) x 10^6	-9.4 ±0.1
8	7.7 (±3.0) x 10^7	-10.8 ±0.2
9	2.0 (±0.3) x 10^6	-8.6 ±0.1

[a] K_T values are reported as the mean (± the standard error of the mean) of three measurements. The K_T values are reported in units of M^{-1}. [b] The free energy values are calculated from the association constants and are reported in kcal•mol^{-1}.

Table 3.2: Oligonucleotide Sequence Composition.

Oligo	# T	# ^{me}C	N_{TT}	N_{TmC}	N_{mCT}	N_{mCmC}	N_{5T}	N_{5mC}	N_{T3}	N_{mC3}
1	9	5	4	5	4	0	1	0	0	1
2	9	4	4	4	4	0	1	0	1	0
3	8	3	4	3	3	0	1	0	1	0
4	9	5	3	5	5	0	1	0	1	0
5	8	5	2	5	5	0	1	0	1	0
6	8	5	4	3	4	1	0	1	1	0
7	8	6	4	4	4	1	0	1	0	1
8	8	6	3	5	5	0	0	1	0	1
9	8	6	4	4	3	2	1	0	0	1

Table 3.3: Measured association constants and free energies of triple helix formation.

Oligo	# T	# ^{me}C	Association constant K_T[a]	ΔG[b]
1	9	5	3.1 (±0.5) x 10^8	-11.6 ±0.1
4	9	5	2.6 (±0.5) x 10^8	-11.5 ±0.1
5	8	5	6.0 (±0.5) x 10^7	-10.6 ±0.1
6	8	5	2.5 (±0.6) x 10^6	-8.7 ±0.1
7	8	6	7.5 (±0.8) x 10^6	-9.4 ±0.1
8	8	6	7.7 (±3.0) x 10^7	-10.8 ±0.2
9	8	6	2.0 (±0.3) x 10^6	-8.6 ±0.1

[a] K_T values are reported as the mean (± the standard error of the mean) of three measurements. The K_T values are reported in units of M^{-1}. [b] The free energy values are calculated from the association constants and are reported in kcal•mol^{-1}.

$$\begin{pmatrix} 4 & 5 & 4 & 0 & 1 & 0 & 0 & 1 \\ 4 & 4 & 4 & 0 & 1 & 0 & 1 & 0 \\ 4 & 3 & 3 & 0 & 1 & 0 & 1 & 0 \\ 3 & 5 & 5 & 0 & 1 & 0 & 1 & 0 \\ 2 & 5 & 5 & 0 & 1 & 0 & 1 & 0 \\ 4 & 3 & 4 & 1 & 0 & 1 & 1 & 0 \\ 4 & 4 & 4 & 1 & 0 & 1 & 0 & 1 \\ 3 & 5 & 5 & 0 & 0 & 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} \Delta G_{TT} \\ \Delta G_{TmC} \\ \Delta G_{mCT} \\ \Delta G_{mCmC} \\ \Delta G_{5T} \\ \Delta G_{5mC} \\ \Delta G_{T3} \\ \Delta G_{mC3} \end{pmatrix} = \begin{pmatrix} -11.6 \\ -11.3 \\ -9.0 \\ -11.5 \\ -10.6 \\ -8.7 \\ -9.4 \\ -10.8 \\ -8.6 \end{pmatrix}$$

Figure 3.2: Simultaneous equations used for finding the ΔG_{XX} components.

eight oligonucleotides must be used in order to solve for the values of the ΔG_{XX} components. The simultaneous equations are shown in matrix form in Figure 3.2.

If the free energy of binding of eight oligonucleotides are determined, the simultaneous equations can be solved and the value of the eight ΔG_{XX} components can be determined. The value of these components was determined by using the Solver feature of Microsoft Excel version 4.0 to minimize value of the following expression for the n oligonucleotides:

$$\sum_{i=1}^{n} \frac{(\Delta G_{obs} - \Delta G_{calc})^2}{s_{obs}^2} \qquad (5)$$

where s_{obs} is the standard error of the mean of the experimental data. Using the experimental data for oligonucleotides 1 - 8 allows determination of the value of the ΔG components used in this model. These are given in Table 3.4.

As the data in Table 3.4 show, the agreement between the measured and predicted values for the free energies of triple helix formation given by this model is quite good, the largest difference between the predicted and measured values being 0.3 kcal mol^{-1}. Using the values for the ΔG components provided by fitting data from oligonucleotides **1 - 8**, the free energy of triple helix formation for oligonucleotide **9** may be predicted. The predicted value of -8.3 kcal mol^{-1} is in good agreement with the measured value of -8.6 kcal mol^{-1}. Perhaps more importantly, the addition of the known binding constant for **9** to the data used in the minimization procedure does not greatly affect the values of the ΔG components. (see Table 3.5) This suggests that the value of the ΔG

components is likely to be relatively insensitive to the particular oligonucleotides used in the basis set.

Finally, we must address the use of a constant term for the initiation energy in the formation of the triple helix. That is to say, should the equation used for the prediction of the free energy be of the form:

$$\Delta G_{tot} = \Delta G_{init} + \sum_{XX} N_{XX} \Delta G_{XX} \qquad (6)$$

While a constant ΔG_{init} has been used by some workers,[62] others have found it to be unnecessary.[59-60] We find that, in this case, the use of a constant ΔG_{init} in the determination of the values of the ΔG_{XX} components, using data from all nine oligonucleotides, does not result in an improved prediction of the ΔG_{tot} for the various oligonucleotides. We find that the value of ΔG_{init} determined in the fitting process is dependent on the initial values used in the fitting procedure. Further, while the values of ΔG_{TT}, ΔG_{TmC}, ΔG_{mCT}, and ΔG_{mCmC} are unaffected by the inclusion of ΔG_{init}, the effect of inclusion of such a term is completely offset by changes in the values of ΔG_{5T}, ΔG_{5mC}, ΔG_{T3} and ΔG_{mC3}. For these reasons, we find that the inclusion of an explicit term for the free energy of initiation of triple helix formation is not necessary, and have fixed the value of ΔG_{init} at 0.

Examination of the values of the ΔG components suggests some interesting inferences. The first is that the 5′-TT-3′ stack makes the greatest positive contribution to the binding energy of the triple helix, ≈ -1.2 kcal•mol^{-1} dinucleotide^{-1}. The 5′-T^{me}C-3′ contributes ≈ -1.0 kcal•mol^{-1} dinucleotide^{-1}, while the 5′-^{me}CT-3′ stack contributes ≈ -0.9 kcal•mol^{-1} dinucleotide^{-1}. The inclusion of

Table 3.4: Values of ΔG components and predicted ΔG for triple helix formation based upon measured ΔG values for oligonucleotides **1 - 8**.

ΔG component	Value
ΔG_{TT}	-1.23[a]
ΔG_{TmC}	-1.01[a]
ΔG_{mCT}	-0.95[a]
ΔG_{mCmC}	0.66[a]
ΔG_{5T}	0.80[b]
ΔG_{5mC}	1.38[b]
ΔG_{T3}	0.86[b]
ΔG_{mC3}	1.33[b]

Oligo	ΔG predicted[b]	ΔG measured[b]
1	-11.6	-11.6
2	-11.1	-11.3
3	-9.1	-9.0
4	-11.8	-11.5
5	-10.6	-10.6
6	-8.8	-8.7
7	-9.4	-9.4
8	-10.8	-10.8
9	-8.3	-8.6

[a] expressed in kcal•mol^{-1} dinucleotide^{-1}. [b] expressed in kcal•mol^{-1}.

Table 3.5: Values of ΔG components and predicted ΔG for triple helix formation based upon measured ΔG values for oligonucleotides **1 - 9**.

ΔG component	Value
ΔG_{TT}	-1.22[a]
ΔG_{TmC}	-1.03[a]
ΔG_{mCT}	-0.93[a]
ΔG_{mCmC}	0.54[a]
ΔG_{5T}	0.76[b]
ΔG_{5mC}	1.45[b]
ΔG_{T3}	0.87[b]
ΔG_{mC3}	1.35[b]

Oligo	ΔG predicted[b]	ΔG measured[b]
1	-11.6	-11.6
2	-11.1	-11.3
3	-9.1	-9.0
4	-11.8	-11.5
5	-10.6	-10.6
6	-8.8	-8.7
7	-9.4	-9.4
8	-10.6	-10.8
9	-8.6	-8.6

[a] expressed in kcal•mol^{-1} dinucleotide^{-1}. [b] expressed in kcal•mol^{-1}.

a 5′-^{me}C^{me}C-3′ dinucleotide requires payment of an energetic penalty of ≈ 0.5 kcal•mol^{-1} dinucleotide^{-1}. The unfavorable energetics for including the 5′-^{me}C^{me}C-3′ dinucleotide in a triple helix are not particularly surprising, since the formation of ^{me}C+GC triplets requires protonation of the ^{me}C residue if two hydrogen bonds are to be made to the purine base. Neither protonation of adjacent residues in the third strand nor formation of triplets with a hydrogen bond donor acceptor mismatch is likely to be particularly favorable in energetic terms.

The terms for the nucleotide at either end of the triple helix are expected to be positive, since the termination of the triple helix disrupts the stacking which is present in the interior of the triple helix and exposes these residues to the solvated environment of the major groove. However, the penalty for using a ^{me}C residue in these positions is almost twice as great as that for using a T at the end of the oligonucleotide. This is likely due to the need for protonation of the meC residue in order to form two Hoogsteen hydrogen bonds with the edge of the purine base. It is known that the pK_a of ^{me}C residues in the triple helix is ≈ 5.7, which is ≈ 1.3 pK_a units higher than that found for C or ^{me}C residues in free oligonucleotides.[21] It seems likely that the pK_a of ^{me}C residues at either end of the triple helix would have an intermediate pK_a between those of residues found in the interior of the oligonucleotide in the triple helical complex and a single stranded oligonucleotide. The resulting decrease in the degree of protonation of the oligonucleotide results in the loss of a hydrogen bond, and increases the penalty for ending the triple helix.

The result of this greater energetic penalty for adding a ^{me}C to either end of an oligonucleotide is that the addition of a ^{me}C to the end of a triple helix increases the binding affinity of the oligonucleotide by a maximum of 0.5

kcal•mol^{-1}. This is in contrast to the addition of a T residue to the end of a triple helix, which results in an increase in the binding affinity of at least 1.2 kcal•mol^{-1}. The binding affinity of an oligonucleotide which has a 5′-^{me}C^{me}C-3′ dinucleotide at either end will actually be enhanced by the removal of the terminal ^{me}C residue.

The results presented above are in general agreement with a prior study of sequence composition effects.[56] In this study, the replacement of a T with a ^{me}C residue in a poly T oligonucleotide results in a small decrease in the affinity of the oligonucleotide. Introduction of a 5′-^{me}C^{me}C-3′ dinucleotide results in a substantial decrease in the affinity of the oligonucleotide for its target site; introduction of a 5′-^{me}C^{me}C^{me}C-3′ trinucleotide almost completely abolishes binding.

The results of a study of the sequence effects at the junction of cooperatively binding oligonucleotides is also qualitatively consistent with the results reported here.[63] In this study, the interaction energy of two oligonucleotides binding to adjacent sites on duplex DNA was found to be dependent on the identity of the bases at the junction between the two oligonucleotides. The highest favorable interaction energies were observed for junctions which resulted in the formation of 5′-T^{me}C-3′ and 5′-^{me}CT-3′ stacks. Those junctions which resulted in the formation of 5′-TT-3′ or 5′-^{me}C^{me}C-3′ base stacks had less favorable interaction energies. While the results may not be compared quantitatively due to the different solution conditions employed in the two studies, the results do agree in a qualitative sense. The more favorable interaction energies can be seen as arising from the combination of both the formation of a favorable dinucleotide *and* desolvation of and inclusion in a stacked complex of one energetically unfavorable ^{me}C terminal residue. The

junctions with less favorable interaction energies do not have both of these factors working in their favor. The 5′-^{me}C^{me}C-3′ junction does result in the inclusion of two ^{me}C residues in a stacked complex, but it results in the formation of a dinucleotide which does not interact in an energetically favorable manner with the double helix (5′-^{me}C^{me}C-3′). The 5′-TT-3′ junction does result in the formation of a dinucleotide which interacts favorably with the double helix. However, the exposure of the terminal T residues is not as energetically punitive as the exposure of ^{me}C residues, thus there is not as much to be gained by shielding these residues in a triple helical region.

The specific numerical predictions allowed by the model used in this study will not be general in nature. Since triple helix formation is dependent on the solution conditions used in a particular study, any numerical predictions will only be valid for the conditions used in this work (10 mM bis-tris, 100 mM NaCl, 250 μM spermine, pH 7.0, 22°C). The general agreement of this work with that reported in the literature suggests that the more general conclusions and trends derived from this study may be applicable to a wider range of solution conditions.

In conclusion, a model which assumes that the free energy of triple helix formation can be modeled as a linear combination of the various dinucleotide units in the third stand has been utilized. The free energy contributions to triple helix formation of these steps has been obtained. The 5′-TT-3′ dinucleotide makes the most favorable interaction with the double helix, while the inclusion of a 5′-^{me}C^{me}C-3′ dinucleotide requires the payment of an energetic penalty. Knowledge of the numerical values for these ΔG components allows the prediction of the free energy of binding for oligonucleotides under the conditions

used in this study. The trends observed in these values are consistent with and help explain previously published data.

References

(1) Moser, H. E.; Dervan, P. B. *Science* **1987**, *238*, 645-650.

(2) LeDoan, T.; Perrouault, L.; Praseuth, D.; Habhoub, N.; Decout, J.-L.; Thuong, N. T.; Lhomme, J.; Hélène, C. *Nucleic Acids Res.* **1987**, *15*, 7749-7760.

(3) Cooney, M.; Czernuszewicz, G.; Postel, E. H.; Flint, S. J.; Hogan, M. E. *Science* **1988**, *241*, 456-459.

(4) Felsenfeld, G.; Davies, D.; Rich, A. *J. Am. Chem. Soc.* **1957**, *79*, 2023-2024.

(5) Beal, P. A.; Dervan, P. B. *Science* **1991**, *251*, 1360-1363.

(6) Beal, P. A.; Dervan, P. B. *Nucleic Acids Res.* **1992**, *20*, 2773-2776.

(7) Stilz, H. U.; Dervan, P. B. *Biochemistry* **1993**, *32*, 2177-2185.

(8) Malkov, V. A.; Voloshin, O. N.; Soyfer, V. N.; Frank-Kamenetskii, M. D. *Nucleic Acids Res.* **1993**, *21*, 585-591.

(9) Jayasena, S. D.; Johnston, B. H. *Biochemistry* **1992**, *31*, 320-327.

(10) Beal, P. A.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 4976-4982.

(11) Jayasena, S. D.; Johnston, B. H. *Nucl. Acids Res.* **1992**, *20*, 5279-5288.

(12) Jayasena, S. D.; Johnston, B. H. *Biochemistry* **1993**, *32*, 2800-2807.

(13) Roberts, R. W.; Crothers, D. M. *Science* **1992**, *258*, 1463-1466.

(14) Han, H.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA*, **1993**, *90*, 3806-3810.

(15) Escude, C.; François, J. C.; Sun, J. S.; Ott, G.; Sprinzl, M.; Garestier, T.; Hélène, C. *Nucleic Acids Res.* **1993**, *21*, 5547-5553.

(16) Singleton, S. F.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 6957-6965.

(17) Mergny, J.-L.; Sun, J.-S.; Rougée, M.; Montenay-Garestier, T.; Barcelo, F.; Chomilier, J.; Hélène, C. *Biochemistry* **1991**, *30*, 9791-9798.

(18) Rougée, M.; Faucon, B.; Mergny, J. L.; Barcelo, F.; Giovannageli, C.; Garestier, T.; Hélène, C. *Biochemistry* **1992**, *31*, 9269-9278.

(19) Povsic, T. J.; Dervan, P. B. *J. Am. Chem. Soc.* **1989**, *111*, 3059-3061.

(20) Xodo, L. E.; Manzini, G.; Quadrifoglio, F.; van der Marel, G. A.; van Boom, J. H. *Nucleic Acids Research* **1991**, *19*, 5625-5631.

(21) Singleton, S. F.; Dervan, P. B. *Biochemistry* **1992**, *31*, 10995-11003.

(22) Singleton, S. F.; Dervan, P. B. *Biochemistry* , **1993**, *32*, 13171-13179.

(23) Thomas, T.; Thomas, T. J. *Biochemisty*, **1993**, *32*, 14068-14074.

(24) Strobel, S. A.; Doucette-Stamm, L. A.; Riba, L.; Housman, D. E.; Dervan, P. B. *Science* **1991**, *254*, 1639-1642.

(25) Maher, L. J.; Dervan, P. B.; Wold, B. *Science* **1989**, *245*, 725-730.

(26) Collier, D. A.; Thuong, N. T.; Hélène, C. *J. Am. Chem. Soc.* **1991**, *113*, 1457-1458.

(27) Maher, L. J.; Dervan, P. B.; Wold, B. *Biochemistry* **1992**, *31*, 70-81.

(28) Grigoriev, M.; Praseuth, D.; Robin, P.; Hemar, A.; Saison-Behmoaras, T.; Dautry-Varsat, A.; Thuong, N. T.; Hélène, C.; Harel-Bellan, A. *J. Biol. Chem.* **1992**, *267*, 3389-3395.

(29) Horne, D.A.; Dervan, P. B. *J. Am. Chem. Soc.* **1990**, *112*, 2435-2437.

(30) Griffin, L. C.; Dervan, P. B. *Science* **1989**, *245*, 967-971.

(31) Beal, P. A.; Dervan, P. B. *Nucleic Acids Res.* **1992**, *20*, 2773-2776.

(32) Shimizu, M.; Inoue, H.; Ohtsuka, E. *Biochemistry*, **1994**, *33*, 606-613.

(33) Mergny, J.-L.; Sun, J.-S.; Rougée, M.; Montenay-Garestier, T.; Barcelo, F.; Chomilier, J.; Hélène, C. *Biochemistry* **1991**, *30*, 9791-9798.

(34) Miller, P. S.; Cushman, C. D.; *Biochemistry*, **1993**, *32*, 2999.

(35) Fossella, J. A.; Kim, Y. J.; Shih, H.; Richards, E. G.; Fresco, J. R. *Nucleic Acids Res.* **1993**, *21*, 4511-4515.

(36) Yoon, K.; Hobbs, C. A.; Koch, J.; Sardaro, M.; Kutny, R.; Weis, A. *Proc. Natl. Acad. Sci. USA*, **1992**, *89*, 3840-3844.

(37) Belotserkovskii, B. P.; Veselkov, A. G.; Filippov, S. A.; Dobrynin, V. N.; Mirkin, S. M.; Frank-Kamenetskii, M. D. *Nucleic Acids Res.* **1990**, *18*, 6621-6624.

(38) Xodo, L. E.; Alunni-Fabbroni, M.; Manzini, G.; Quadrifoglio, F. *Eur. J. Biochem.* **1993**, *212*, 395-401.

(39) de los Santos, C.; Rosen, M.; Patel, D. J. *Biochemistry* **1989**, *28*, 7282-7289.

(40) Rajagopal, P.; Feigon, J. *Nature* **1989**, *339*, 637-640. (c) Rajagopal, P.; Feigon, J. *Biochemistry* **1989**, *28*, 7859-7870.

(41) Live, D. H.; Radhakrishnan, I.; Misra, V.; Patel, D. J. *J. Am. Chem. Soc.* **1991**, *113*, 4687-4688.

(42) Ono, A.; Ts'o, P. O. P.; Kan, L. *J. Am. Chem. Soc.* **1991**, *113*, 4032-4033.

(43) Griffin, L. C.; Kiessling, L. L.; Beal, P. A.; Gillespie, P.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 7976-7982.

(44) Miller, P. S.; Bhan, P.; Cushman, C. D.; Trapane, T. L. *Biochemistry* **1992**, *31*, 6788-6793.

(45) Krawczyk, S. H.; Milligan, J. F.; Wadwani, S.; Moulds, C.; Froehler, B. C.; Matteucci, M. D. *Proc. Natl. Acad. Sci. USA*, **1992**, *89*, 3761-3764.

(46) Radhakrishnan, I.; Gao, X.; de los Santos, C.; Live, D.; Patel, D. J. *Biochemistry*, **1991**, *30*, 9022-9030.

(47) Wang, E.; Malek, S.; Feigon, J. *Biochemistry*, **1992**, *31*, 4838-4846.

(48) Radhakrishnan, I.; Patel, D. J.; Veal, J. M.; Gao, X. *J. Am. Chem. Soc.* **1992**, *114*, 6913-6915.

(49) Koh, J. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 1470-1478.

(50) Radhakrishnan, I.; Patel, D. J.; Priestley, E. S.; Nash, H.; Dervan, P. B. *Biochemistry* , **1993**, *32*, 11228-1234.

(51) Koshlap, K. M.; Gillespie, P.; Dervan, P. B.; Feigon, J. *J. Am. Chem. Soc.* **1993**, *115*, 7908-7909.

(52) Best, G. C.; Dervan, P. B. manuscript in preparation.

(53) Brenowitz, M.; Senear, D. F.; Shea, M. A.; Ackers, G. K. *Meth. Enzymol.* **1986**, *130*, 132-181.

(54) Senear, D. F.; Brenowitz, M.; Shea, M. A.; Ackers, G. K. *Biochemistry*, **1986**, *25*, 7344-7354.

(55) Brenowitz, M.; Senear, D. F.; Shea, M. A.; Ackers, G. F. *Proc. Natl. Acad. Sci., USA*, **1986**, *83*, 8462-8466.

(56) Kiessling, L. L.; Griffin, L. C.; Dervan, P. B. *Biocehmistry*, **1992**, *31*, 2829-2834.

(57) Sambrook, J.; Fritsch, E. F.; Maniatis, T. *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press: Cold Spring Harbor, N. Y., 1989.

(58) Bevington, P. R. *Data Reduction and Error Analysis for the Physical Sciences*; McGraw-Hill: New York, 1969.

(59) Gralla, J.; Crothers, D. M. *J. Mol. Biol.* **1973**, *73*, 497-511.

(60) Tinoco, I., Jr.; Borer, P. N.; Dengler, B.; Levine, M. D.; Uhlenbeck, O. C.; Crothers, D. M.; Gralla, J. *Nature New Biol.* **1973**, *246*, 40-41.

(61) Marky, L. A.; Breslauer K. J. *Biopolymers*, **1982**, *21*, 2185-2194.

(62) Breslauer, K. J.; Frank, R.; Blöcker, H.; Marky, L. A. *Proc. Natl. Acad. Sci., USA*, **1986**, *83*, 3746-3750.

(63) Colocci, N.; Dervan, P. B. Manuscript in preparation.