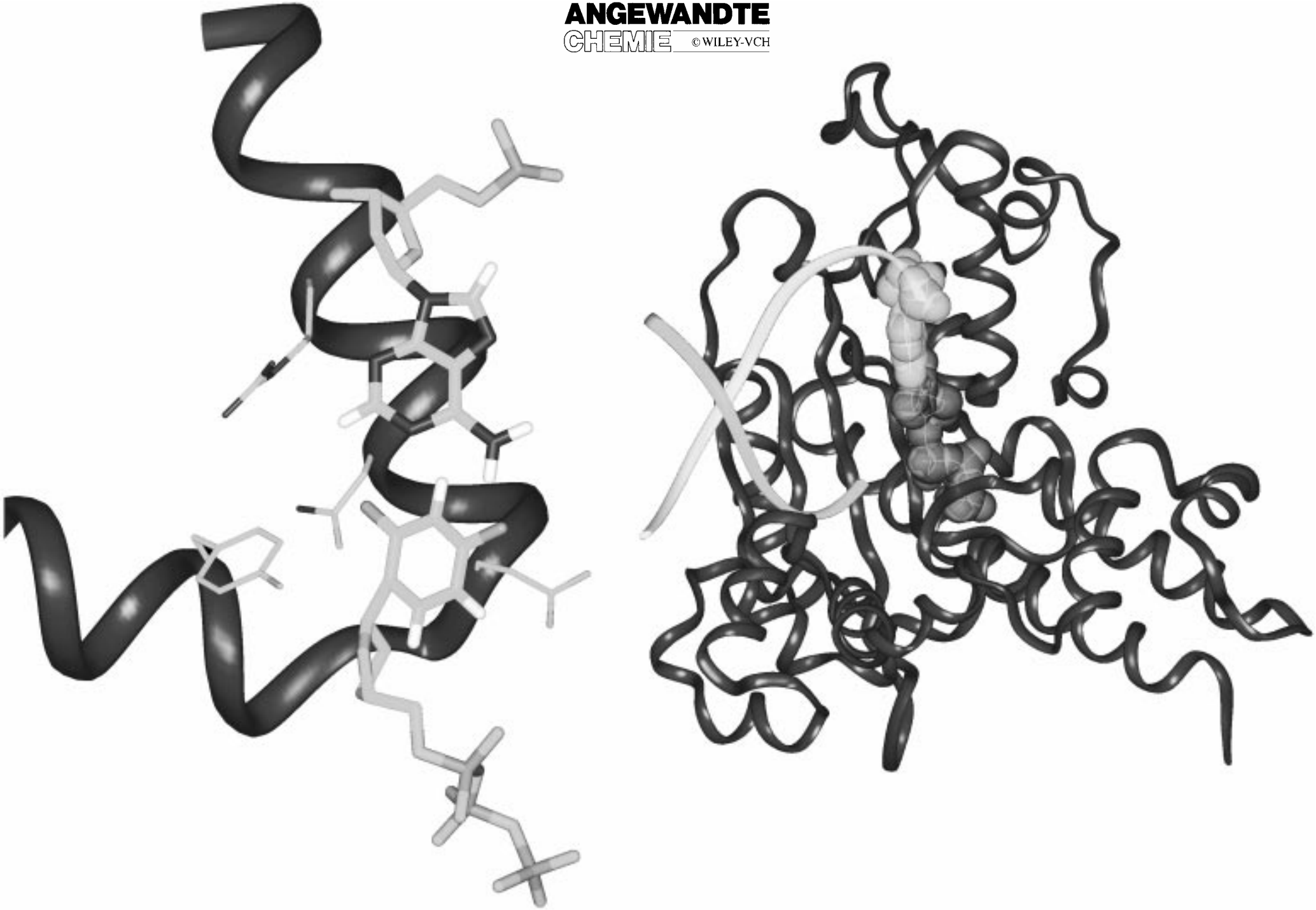
ANGEWANDTE
CHEMIE



Mimicking the Structure and Function of DNA: Insights into DNA Stability and Replication

Eric T. Kool,* Juan C. Morales, and Kevin M. Guckian

The physical and chemical factors that allow DNA to perform its functions in the cell have been studied for several decades. Recent advances in the synthesis and manipulation of DNA have allowed this field to move ahead especially rapidly during the past fifteen years. One of the most common chemical approaches to the study of interactions involving DNA has been the use of DNA base analogues in which functional groups are added, deleted, blocked, or rearranged. Here we describe a different strategy, in which the polar natural DNA bases are replaced by nonpolar aromatic molecules of the same size and shape. This allows the evaluation of polar interactions (such as hydrogen bonding) with little or no interference from steric effects. We have used these nonpolar nucleoside isosteres as probes of noncovalent interactions such as DNA base pairing and protein–DNA recognition. We have found that, while base-pairing selectivity does depend on Watson–Crick hydrogen bonding in the natural pairs, it is possible to design new bases that pair selectively and stably in the absence of hydrogen bonds. In addition, studies have been carried out with DNA polymerase enzymes to investigate the importance of Watson–Crick hydrogen bonding in enzymatic DNA replication. Surprisingly, this hydrogen bonding is not necessary for efficient enzymatic synthesis of a base pair, and significant levels of selectivity can arise from steric effects alone. These results may have significant impact both on the study of basic biomolecular interactions and in the design of new, functionally active biomolecules.

Keywords: base pairing · DNA polymerase · hydrogen bonds · nucleobases

1. Introduction

1.1. Overview

We describe a series of studies involving a new class of nucleoside mimics. These molecules, called nonpolar nucleoside isosteres, imitate the shape of the natural nucleosides and nucleotides as closely as possible, but they lack Watson–Crick hydrogen-bonding ability. This makes them especially valuable as mechanistic tools for the study of the importance of these noncovalent bonds in the various functions of nucleic acids.

These studies have led to some surprising observations about the mechanism whereby polymerase enzymes copy a strand of DNA. The results have potentially important implications in the basic scientific study of biochemical and biophysical structure and function, and may lead to practical applications as well. In this paper, we first briefly review the basic structure formed by DNA and the functions that it serves in a cell. Then we describe how the structure and functions of DNA can be mimicked successfully by consideration of simple steric factors. Finally, we discuss the implications and possible future directions for this work.

[*] Prof. E. T. Kool,[+] Dr. J. C. Morales,[+] K. M. Guckian
Department of Chemistry
University of Rochester
Rochester, NY 14627 (USA)
[+] Current address:
Department of Chemistry
Stanford University
Palo Alto, CA 94305-5080 (USA)
Fax: (+1) 650-725-0259
E-mail: kool@leland.stanford.edu

1.2. The Structures and Functions of DNA

The three dimensional structure of the DNA double helix was correctly elucidated by Watson and Crick more than four decades ago.[1] The central features of this structure include an anionic sugar–phosphate backbone that winds around the outside of the cylindrical helix, and neutrally charged nucleobases that occupy the center.[2] The paired bases are roughly perpendicular to the helical axis, and are stacked in van der Waals contact with one another. The right-handed twist of the backbone (in B-form DNA) defines two grooves in the helix (the major and the minor) which are lined by the

  1433-7851/00/3906-0991 $ 17.50+.50/0

sugar–phosphate backbone (serving as the walls of the grooves) and the edges of the bases (which act as floors).

The chemical moieties that make up the primary structure of DNA are shown in Scheme 1. The phosphodiester backbone includes 2′-deoxyribose residues which serve to carry and orient the DNA bases. The backbone as a whole is highly polar because of the negative charge carried on each phosphate diester at neutral pH. The nucleobases themselves are aromatic nitrogen heterocycles comprised of monocyclic pyrimidines and bicyclic purines. They are generally quite polar along their edges, carrying multiple substituents such as amino, imino, and carbonyl groups.

The primary function of DNA is the storage of genetic information; the three billion base pairs in each human cell encode the synthesis of approximately 100000 proteins. It is, of course, crucial that not only are these proteins synthesized faithfully, but also that this genetic information is transferred accurately from generation to generation in a line of organisms. Central to these abilities is the double helical structure of DNA. Its antiparallel, C2-symmetrical structure with complementary hydrogen-bonded base pairs provides an extraordinarily effective means to carry out storage and transfer of the encoded messages.

DNA also acts as a structural scaffold for the arrangment of multiprotein complexes in a cell. This water-soluble scaffold is inherently rather rigid in a local sense, but proteins can use their binding energy to remodel the duplex by bending, winding, or unwinding it. In addition, chemists are currently making use of the structural properties of DNA to build artificial scaffolds for use in materials science and biomedical applications.[3–11]

1.3. Testing the Paradigms for DNA Stability and Replication

The aim of this paper is to examine some of the most basic questions of DNA structure and function. To begin to answer such questions, the chemist turns naturally to the idea of using nucleoside analogues. However, the large majority of existing analogues of natural DNA bases have significant structural changes relative to their natural congeners. Even a seemingly small perturbation such as alkylation of a base leads to significant steric changes in the molecule. For example, methylation of a hydrogen-bonding group might decrease a


E. T. Kool

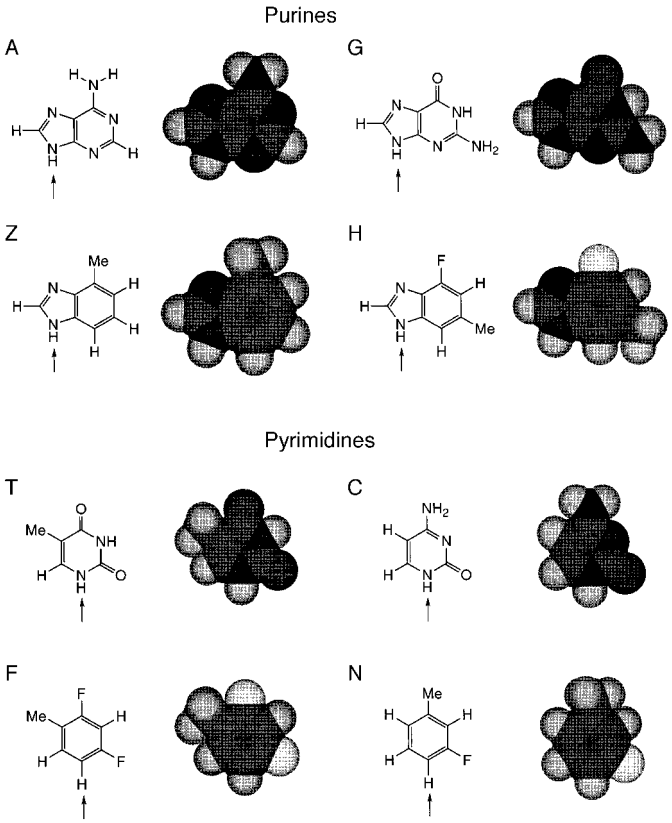

J. C. Morales


K. M. Guckian

Eric T. Kool is Professor of Chemistry at Stanford University in Stanford, CA (USA). He received his Ph.D. in 1988 as an NSF Predoctoral Fellow at Columbia University, working in the research group of Professor Ronald Breslow. Kool continued his training at the California Institute of Technology as an NIH Postdoctoral Fellow with Professor Peter B. Dervan. In 1990 he joined the faculty at the University of Rochester in New York, and he moved to Stanford University in 1999. His research interests fall in the field of bioorganic chemistry, and include the design of molecules which can recognize DNA and RNA sequences and structures, and of molecules that mimic biological functions. More than 70 research publications to date describe this work, and he has presented over 120 invited lectures. He has received several awards in recognition of his work, including the Arthur C. Cope Scholar Award of the American Chemical Society, an Alfred P. Sloan Foundation Fellowship, the Young Investigator Award of the US Office of Naval Research, the US Army Research Office Young Investigator Award, the Arnold and Mabel Beckman Foundation Young Investigator Award, an American Cyanamid Faculty Award, and the Camille and Henry Dreyfus Foundation Teacher–Scholar Award.

Juan C. Morales was born in Madrid (Spain) in 1969. He received his B.Sc. in chemistry from the Universidad de Alcalá in 1992. He gained his Ph.D. under the supervision of Professor S. Penadés at the Universidad Autónoma de Madrid in 1996, by the work carried out at the Instituto de Química Orgánica (C.S.I.C.) on molecular recognition of carbohydrates using macrocyclic receptors. He is currently a postdoctoral researcher in the research group of Prof. E. T. Kool. His research interests focus on the understanding of molecular recognition in DNA–protein and carbohydrate–protein systems and its application to drug design.

Kevin M. Guckian was born and raised in Rochester, NY (USA). He obtained his B.Sc. in chemistry from Oswego State University in 1995. Since 1995 he has studied the structure and function of DNA using nonpolar base analogues under the supervision of Prof. E. T. Kool at the University of Rochester. His research interests focus on protein–DNA interactions.



Scheme 1. Chemical formulae and space-filling models of the four natural bases and their corresponding nonpolar isosteres. Arrows indicate point of attachment of deoxyribose.

protein's affinity for the DNA, but the reason for this could be either loss of the hydrogen bond or steric repulsion.

For this reason, we designed a series of compounds that are meant to mimic DNA base structure but to lack Watson–Crick hydrogen-bonding ability.[12, 13] These "nonpolar nucleoside isosteres" (Scheme 1) are simple compounds and are generally prepared in short syntheses. In this design, substituted benzenes mimic the pyrimidine bases, and indoles or benzimidazoles replace the purines. Of the four shown in Scheme 1, we consider two to be nearly perfect isosteres (H—an analogue of G, and F–—an analogue of T). We consider N and Z less ideal, because in these a nitrogen in the aromatic ring is replaced by the sterically larger C-H group. This adds approximately 0.7 Å of steric bulk to the molecules at this position. Nonetheless, these are the closest possible isosteres available which lack the ability to form hydrogen bonds.

1.4. Possible Outcomes

The intended use for these molecules was to substitute them for the natural nucleosides in DNA, and evaluate whether they function properly. One possible outcome of such experiments is negative; that is, the molecular mimics may lack essentially all ability to function properly. However, it is sometimes difficult to draw confident conclusions when a negative result is found, because it is usually true that in changing a molecule, more than one property is affected.

Probably more useful and important is the second possible outcome, retention of activity. For example, deletion of a hydrogen-bonding group in DNA might conceivably result in little loss of stability or activity. In such a case one can be somewhat more confident in concluding that the hydrogen bond was probably not energetically important to the stabilization of a ground or transition state. It is notable that similar analyses have been carried out for years in site-directed mutagenesis studies of proteins.

The basic tenets of science become paradigms when they have been repeatedly tested and questioned. Here we ask some of these questions again.

Are hydrogen bonds necessary for stabilization of the DNA helix? Since Watson and Crick's elucidation of the duplex structure, it has been generally accepted that the hydrogen bonds between the bases are critical to the stability of DNA. A Watson–Crick pair has two or three such bonds, and these add approximately 0.5–1.8 kcal mol^{-1} of stabilization per base pair of DNA.[14–17] The uncertainties in this value arise from the fact that the experimentally measured value varies greatly depending on the molecular context. Of course, this relatively low energetic stabilization reflects the competition with water, because breaking each hydrogen bond in a base pair (by separating the bases) also means the simultaneous formation of two hydrogen bonds with water molecules. In the gas phase, where such competition does not occur, experiments show that individual hydrogen bonds add stabilization of 6–7 kcal mol^{-1} with the kinds of functional groups found in DNA.[18, 19] Although DNA has been studied for decades, the role of hydrogen bonds in stabilization of the helix and selective base-pair formation is still unclear.

Is stacking of the aromatic bases less important to helix stability than hydrogen bonds? In general, the number of hydrogen bonds in the base pairs have received more attention than base stacking in explaining DNA function. However, in B-form DNA, neighboring base pairs make contact over approximately 115 Å^2, by overlap of one π-system with the next (Figure 1). The formation of a helix from a random coil therefore involves removing as much as 230 Å^2

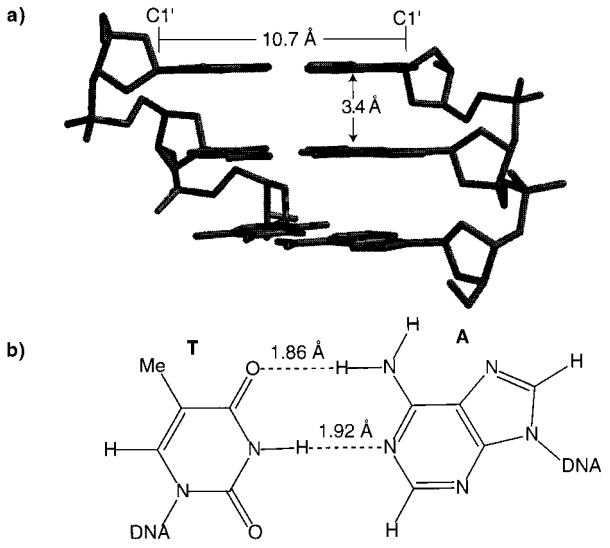
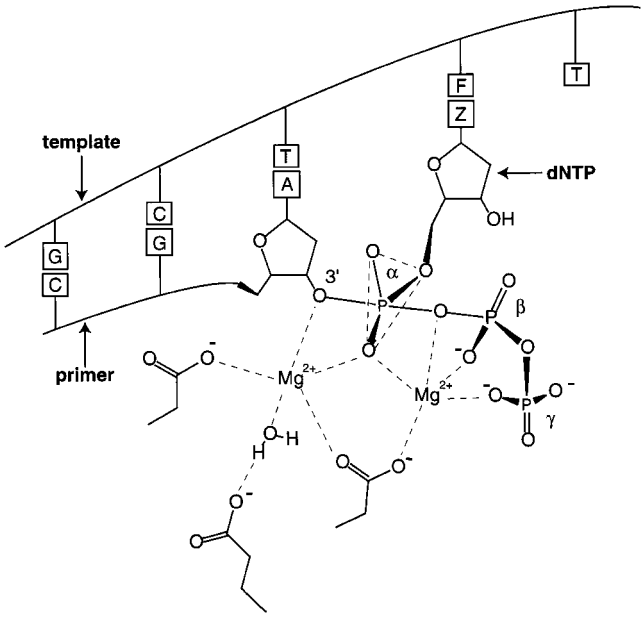


Figure 1. a) Model of a three base-pair B-form DNA duplex with distances between strands and bases. b) Watson–Crick A–T base pair.

of base π-surface area from the solvent per pair, assuming that no such stacking occurs in the single-stranded coil. The specific noncovalent forces that stabilize this stacking are relatively poorly understood. Although permanent electrostatic effects, induced dipole attractions, and solvophobic effects may all play a role, the relative contributions of these effects are still not clear.[20–26]

Are hydrogen bonds necessary for the specificity of base pairing? Hydrogen bonds are widely believed to be one of the most important factors in the specificity of biomolecular recognition. Again, if one examines the existing data on natural and modified DNAs, one would tend to support this paradigm. For example, synthetic DNA duplexes have been widely studied for the thermodynamics of stabilization. When a complementary base pair is mutated to a noncomplementary (mismatched) one, the energetic penalty is quite large. For example, a T–C mismatch destabilizes the helix by approximately 3.2–5.8 kcal mol^{-1} relative to a T–A matched pair.[27–29] However, if only hydrogen bonding is disrupted by mismatches such as T–C, then why does it destabilize the DNA so much? It is reasonable to assume that the stacking of the bases continues to stabilize even with this mismatch. One possible answer to this question might be that the T–C pair might lead to structural distortion of the DNA beyond that pair alone. Indeed, structural studies of mismatched pairs in DNA, such as G–T, and T–T pairs,[30–32] suggests that they may well distort the DNA. Significantly, they are usually hydrogen bonded, but to form the orientation necessary for hydrogen bonding, the pair usually adopts a geometry different than in the normal pairs, thus distorting the DNA backbone as well.

Are Watson-Crick hydrogen bonds essential to DNA replication? The enzymatic replication of DNA occurs with high efficiency (approximately 20–1000 nucleotides per second) and with very high fidelity (with an error only once in 10^3–10^5 nucleotide insertions).[33–40] The idea that Watson–Crick hydrogen bonds are primarily, even solely, responsible for these effects has been taught in biochemistry and molecular biology classes for decades. Commonly used textbooks in these fields nearly always cite the high specificity of hydrogen bonds in explaining DNA synthesis fidelity.[41, 42] However, as mentioned above, several forces are involved in the stabilization of DNA in the absence of enzymes. It would seem reasonable to assume that these may also affect the stability of the transition state of the polymerase catalyzed reaction (Scheme 2). Thus, in addition to hydrogen bonds, what are the effects of permanent dipolar or electrostatic attractions in the stacking of the incoming nucleotide? What are the effects of solvation on this process? And importantly, what are the effects of steric size and shape in the polymerase active site?



Scheme 2. Schematic representation of the transition state of the DNA polymerase reaction. A detailed description of the two metal ion mechanism is given by Steitz and co-workers.[142, 143]

2. Mimicking the Primary Structure

2.1. Shape and Conformation

To address such questions more directly, we have studied nonpolar nucleoside isosteres, which mimic the shape of the natural bases more closely than the previously known analogues (Scheme 1). The structures of these nucleosides have been studied computationally and experimentally. X-ray crystal structures of dF and dZ have been published, with comparisons made to their natural counterparts dT and dA[43, 44] The structure of dF[43] in the crystal is remarkably close to that of thymidine[45, 46] (Figure 2). The two have identical sugar puckers and very similar glycosidic torsional angles. The structure of dZ differs somewhat from that of deoxyadenosine in the solid state.[44] The two sugars have different conformations (although both are skew conformers), and the glycosidic orientation of dZ is *syn*, while that of dA is *anti*.

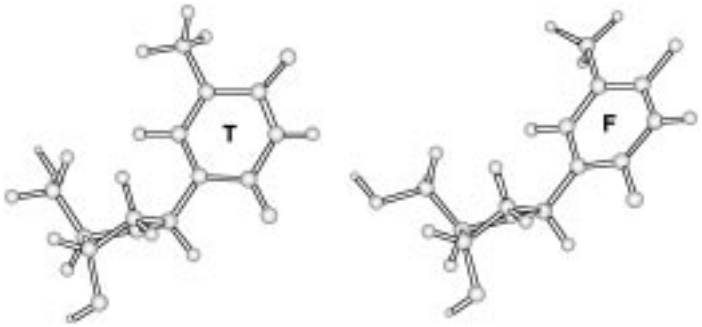


Figure 2. Single-crystal X-ray structure of difluorotoluene deoxynucleoside (right) compared to the published structure of thymidine (left).[45, 46]

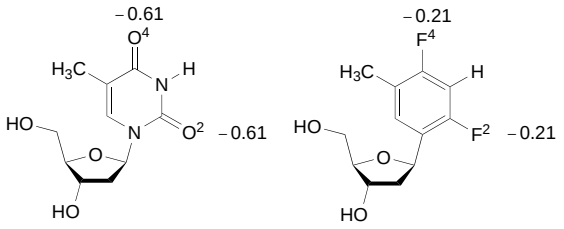
More relevant to the biochemical activities are the solution structures of these analogues, which have also been studied in D_2O by NMR spectroscopic methods. Here it was shown that analogue dF has a sugar conformation quite similar to that of dT in solution, with dF being classified as 90–100% in the skew conformation and dT as 70%.[43] Both nucleobases are

oriented *anti*, or away from the sugar, as nucleosides normally are in DNA. In solution, the sugars of dA and dZ are almost identical (60% and 70% south conformation, respectively), and the bases are both *anti*. Thus, in solution both analogues are very nearly perfect in their steric and conformational mimicry of the natural nucleosides.

2.2. Electronic Factors

There has been some discussion in the literature as to whether organofluorine compounds can form hydrogen bonds with good hydrogen-bond donors.[47, 48] This question is of particular interest in the present case, because difluorotoluene might conceivably form two weakened hydrogen bonds with adenine. These would be a C–F···H–N and a C–H···N interaction. However, Thalladi et al. have concluded that fluorines attached to sp2 carbons are extremely poor as hydrogen-bonding acceptors.[49] Similarly, studies in crown ether systems have concluded that, if such bonds exist at all, they must be very weak.[50]

A number of computational studies have addressed the electronic properties of the nonpolar nucleoside analogues, especially for the difluorotoluene compound.[51–54] For example, a few studies have examined the partial charge on F4, which is analogous to O4 on T (one of the Watson–Crick hydrogen-bonding groups). The calculated partial charge on F4 is −0.21, while for the analogous O4 it is −0.61 (Scheme 3). Based on simple electrostatic (point charge)



Scheme 3. Calculated partial charges for difluorotoluene and thymidine. Data were taken from Wang and Houk.[53]

calculations, one might expect the F–A pair to be approximately one third as stable in enthalpic terms as T–A in the gas phase. Ab initio calculations of the relative F–A and T–A pair stabilities also fall in this range, and so the pair is predicted to be quite weak.[52, 54] In addition, it has been pointed out that screening of the electrostatic attractions by water's high dielectric constant would lead to a significant reduction in this attractive force.[53] Most of the theoretical studies have concluded that this weak noncovalent interaction between F and A is not sufficient to explain later results in DNA replication (see Section 5.3).

2.3. Do Nonpolar Nucleoside Isosteres Form Hydrogen Bonds?

Despite the calculations cited above, the question of whether analogues such as F can undergo base pairing remains an interesting one for experimental testing. The polarity of dT and dF were compared by water–octanol partitioning studies (unpublished results). The results showed that dT is quite hydrophilic, preferring water over octanol ($\lg P = -1.27$; P = partition coefficient). By contrast, dF is lipophilic, preferring octanol over water by a similar margin ($\lg P = 1.39$). Since the sugar is the same in the compounds, it actually masks an even larger difference between the bases T and F. Thus, if F were proficient at hydrogen bonding with water, it should have shown a preference for partitioning in the aqueous phase.

A more direct study of the hydrogen-bonding question was carried out in deuteriochloroform, by measuring any complexes formed between 9-ethyladenine and difluorotoluene or, for comparison, a uracil derivative, U.[55] Complexation was observed by following chemical shifts of adenine in response to the addition of increasing concentrations of F or U. The results showed that while U is indeed able to form a complex, no differences in chemical shifts were observed in the presence of F, even at the highest concentrations studied. Thus, even in chloroform, where hydrogen-bonded complexes are much more stable than in water, we find no evidence for hydrogen bonding by F.

Also relevant are some observations made in an NMR spectroscopy study of duplex DNAs containing T or F paired opposite A (discussed in Section 3).[56] We observed that, for the natural pair, the H3 proton of T was shifted downfield by approximately 3.0 ppm on forming the duplex from the single strands. This is attributed to the effects of hydrogen bonding with the N1 of the adenine opposite. By contrast, we observed that the H3 of F in the same situation was actually shifted slightly upfield (by 0.15 ppm) on forming the duplex from the single-stranded state at 3 °C. Overall, NMR spectroscopy shows no evidence for hydrogen bonds of either the C–H···N or F···H–N type between F and A.

3. Mimicking secondary structure

The observation that analogues such as dF and dZ mimic the structures of the natural nucleosides well in solution does not guarantee that this is the case in the context of double helical DNA. This is especially true in the light of experimental observations (discussed in Section 4.1) which show that the pairing properties of the nonpolar and polar compounds in DNA are very different. One might well expect that a substitution that is very destabilizing to DNA (as is true for some F→T substitutions) might also cause it to be structurally distorted. We were therefore interested in carrying out a structural study of a short DNA duplex containing pairs with F or Z.

This was carried out by use of 2D-NMR spectroscopic methods combined with restrained molecular dynamics and energy minimization.[56] The duplex studied was a twelve base-pair sequence corresponding to the binding site of a known transcription factor. The data showed, surprisingly, that this duplex is structurally identical to what one would expect with a T–A pair at the same site (Figure 3). Thus, although the F–A pair is destabilizing to the DNA (see Section 4.1.2),

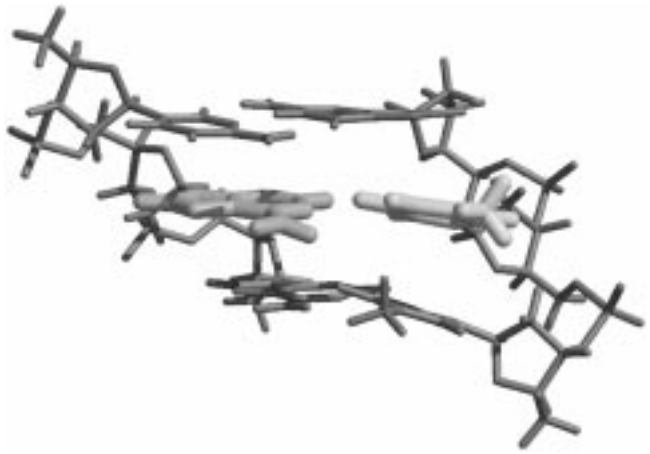


Figure 3. Side view of the central three base pairs of a twelve base-pair duplex which contains an adenine–difluorotoluene pair in the center. Analogue F is on the right.[55]

it adopts a geometry virtually indistinguishable from that of a T–A pair. Later studies were carried out with the same duplex sequence, but in which both T and A were exchanged for F and analogue Z (unpublished results). Those studies have also indicated very little structural distortion of the DNA.

4. Noncovalent Interactions and DNA Stability

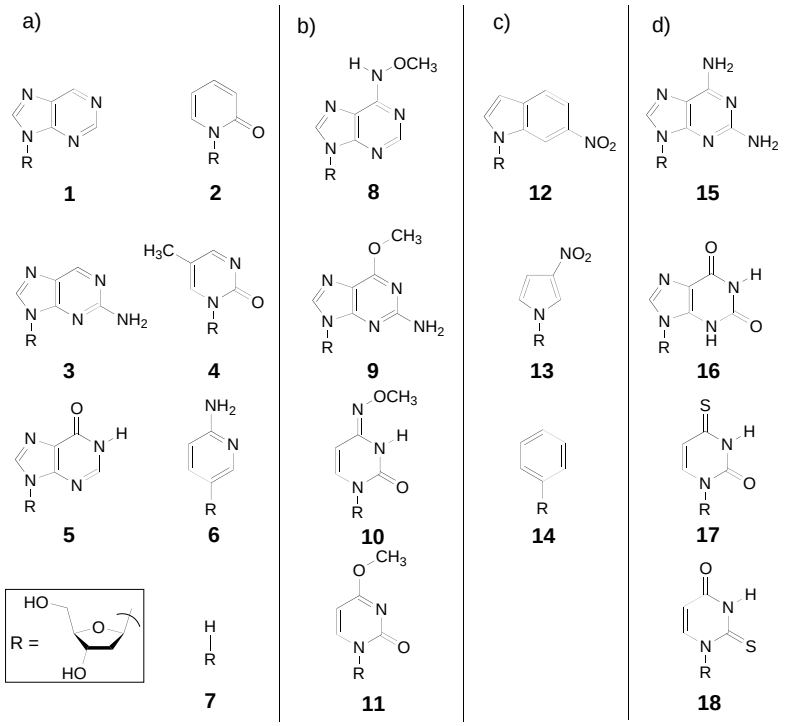
4.1. Thermodynamic Consequences of Removal of Hydrogen Bonds

4.1.1. Nonisosteric Base Analogues

Many analogues which delete, block, modify, or rearrange the hydrogen-bonding groups of natural DNA bases have been studied in the past.[57–68] Some of these analogues are shown in Scheme 4. Deletion of hydrogen-bond donor or acceptor groups (compounds **1**–**6**),[57–60] or deletion of the entire base (abasic site, **7**),[57] has been found to result in a destabilization of the DNA (Table 1), although with some variability. The only exception is the compound 2-aminopurine (**3**),[61] which maintains good stability, presumably due to the fact that two hydrogen bonds can still be made without disrupting the Watson–Crick geometry. A second class of analogues has been prepared by blocking a hydrogen-bond donor or acceptor group (compounds **8**–**11**);[62–65] examples are O-methylated versions of thymine and guanine. These also destabilize DNA, as do the members of the third group (compounds **12**–**14**), analogues that do not resemble the purine or pyrimidine bases.[66–68] Examination of the pairing data generated for the compounds that completely remove the hydrogen bonds between the bases (Table 1) shows that all the analogues do appear to destabilize the DNA. Upon examining the data closely, one observes that the loss of stability on making these substitutions is surprisingly large.

Is this destabilization due to the lack of hydrogen bonds? Or are other factors affecting the stability of the DNA? Two possible explanations must be considered. First, there may be a significant degree of structural distortion with these analogues, which influences not only the pair in question but also the DNA around it. This distortion might well arise from the fact that the analogues are sterically different than the natural molecules, leading to steric or geometric penalties. A second answer may be that solvation effects may be energetically costly; this effect is discussed in Section 4.3, 5.6, and in the following sections.

Another class of analogues contain modifications in their hydrogen-bonding ability (compounds **15-18**).[58, 69] For example, a carbonyl group is switched with an amino group, a C=O bond is replaced by a C=S bond, or an amino group is added. Some of these modified bases, like 2,6-diaminopurine (**15**),[69] or 2-thiothymine (**18**)[58] appear to stabilize the DNA, due to



Scheme 4. Nucleoside analogues with altered, blocked, or rearranged hydrogen-bonding groups.

their retention of the normal Watson–Crick hydrogen bonds and maintaining the correct geometry, whereas others are slightly destabilizing.

One further class of analogues are those that rearrange hydrogen-bond donors and acceptors to form new base pairs. Two new Watson–Crick base pairs have been designed with a hydrogen-bonding pattern different from that of A–T and G–C pairs (Scheme 5). These rearrangements were found to not significantly affect the DNA stability.[70, 71]

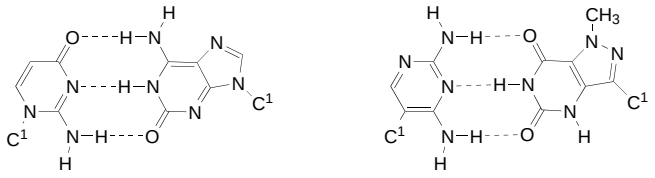
4.1.2. Isosteric Nonpolar Base Analogues.

The nonpolar nucleoside isosteres, analogues unlike those discussed above, do not change the steric size of the DNA

Table 1. Melting temperatures for DNA duplexes which contain non-isosteric base analogues.

Base Pair X – Y	T_m [°C]	ΔT_m [°C]
A – T[a,b]	42.5	–
A – C	36.0	6.5
A – ϕ	26.0	16.5
G – ϕ	30.0	12.5
C – ϕ	26.0	16.5
T – ϕ	28.0	14.5
A – B[a,b]	32.0	10.5
G – B	29.0	13.5
C – B	25.5	17.0
T – B	29.0	13.5
G – C[c,d]	45.4	–
O^{6}meG – A	19.4	26.0
O^{6}meG – G	22.6	22.8
O^{6}meG – C	25.2	20.2
O^{6}meG – T	21.9	23.5
A – T[e,f]	35.0	–
A – C	12.0	23.0
N – A	18.0	17.0
N – G	19.0	16.0
N – C	13.0	22.0
N – T	29.0	6.0

[a] Measured in the sequence d(AACCAGTACGXTGAG) · d(CTCAYCGTACTGGTT) in a buffer which contained 9.5 mM NaCl, 2 mM HEPES and 20 mM EDTA (pH 7.0). B is compound **14**. [b] Data taken from Millican et al.[66] [c] Measured in the sequence d(GGTTXTTGG) · d(CCAAYAACC) in a buffer which contained 100 mM NaCl, 0.1 mM EDTA and 10 mM phosphate (pH 7.0). O^{6}meG is compound **9**. [d] Data taken from Gaffney and Jones.[64] [e] Measured in the sequence d(CGGXGGC) · d(GCCYCCG) in a buffer which contained 150 mM NaCl and 50 mM phosphate (pH 7.0). N is compound **1**. [f] Data taken from Eritja et al.[57] HEPES = 2-[4-(2-hydroxyethyl)-1-piperazinyl]-ethanesulfonic acid, EDTA = ethylenediaminetetraacetate.



Scheme 5. Structures of novel base pairs with rearranged hydrogen-bonding patterns as reported by Benner and coworkers.[70, 90, 91]

base; thus the importance of hydrogen bonding can be probed in the absence of significant steric effects. We therefore carried out studies of the stability of DNA duplexes containing analogues F or Z.[44, 55] The data are shown in Table 2.

Both of these analogues are quite destabilizing to duplex DNA. The single substitutions with F or Z result in a loss of 4 – 5 kcal mol^{-1} in stability, which is a significant fraction of the entire free energy of the duplex. In addition, we observed that both these analogues were almost entirely nonselective in their base pairing, giving roughly equal amounts of destabilization when paired with any of the four natural bases. This latter result suggests that hydrogen bonds do play a significant role in the selectivity of base pairing of natural bases. Thymine shows a selectivity of 3 – 4 kcal mol^{-1} in pairing with A rather than T, C, or G, while molecule F, which has the same size and shape, shows little or no selectivity. This observation also

Table 2. Free energies and melting temperatures for DNA duplexes which contain isosteric nonpolar bases F and Z.

Base Pair X – Y[a]	T_m [°C]	ΔT_m [°C]	$-\Delta G^0_{25}$ [kcal mol^{-1}]
T – A[b]	39.4	–	12.3
T – C	26.4	13.0	8.7
T – G	30.7	8.7	9.3
T – T	27.1	12.3	8.9
F – A[b]	21.4	18.0	7.4
F – C	25.0	14.4	8.2
F – G	23.0	16.4	8.0
F – T	20.3	19.1	7.3
F – F	28.6	10.8	8.9
A – T[c]	39.8	–	12.4
Z – A	25.3	14.5	8.0
Z – C	24.4	15.4	7.9
Z – G	23.8	16.0	7.6
Z – T	20.8	19.0	7.5
Z – F	30.3	9.5	8.9

[a] Measured in the sequence d(CTTTTCXTTCTT) · d(AAGAAY-GAAAAG) in a buffer which contained 100 mM NaCl, 10 mM $MgCl_2$ and 10 mM Na-PIPES (pH 7.0). [b] Data taken from Schweitzer and Kool[55]. [c] Data taken from Guckian et al.[44] Na-PIPES = sodium piperazinebis-(ethanesulfonate).

strongly argues that F is not capable of forming a measurably hydrogen-bonded complex with A in this context.

At a second glance, these data prompt a new question: why are these analogues so destabilizing to the helix? Simple removal of the base pair in question actually results in less destabilization than observed with F – A, for example.[72] This suggests that the net interaction between F and A (or T and Z) is actually repulsive by more than 2 kcal mol^{-1} in water. One possible explanation is that these analogues stack very poorly, this however does not seem to be the case (see Section 4.2). Another experiment shows that mutating an F – A pair to an F – Z or F – F pair actually is a little less destabilizing.[44, 55] This suggests that a reason for the energetic penalty of pairing F with A is the cost of removing the waters of solvation strongly bound to A.[73, 74] Since the 2D-NMR spectroscopically determined structure of F opposite A in DNA shows the two next to one another in the center of the duplex, this suggests that approximately two solvating waters might be lost on forming this duplex, with little energetic compensation (since F cannot hydrogen bond well with A). Of course, another contributing factor to the destabilization is the lack of hydrogen bonds between F and A. A third explanation is that the increased flexibility of these nonpolar base pairs (due to the lack of hydrogen bonds) may disrupt stacking of neighboring base pairs. The structure discussed above containing the F – A pair shows that the neighboring T – A base-pair signal broadens at a lower temperature than others in the sequence, indicating increased movement at this base pair.

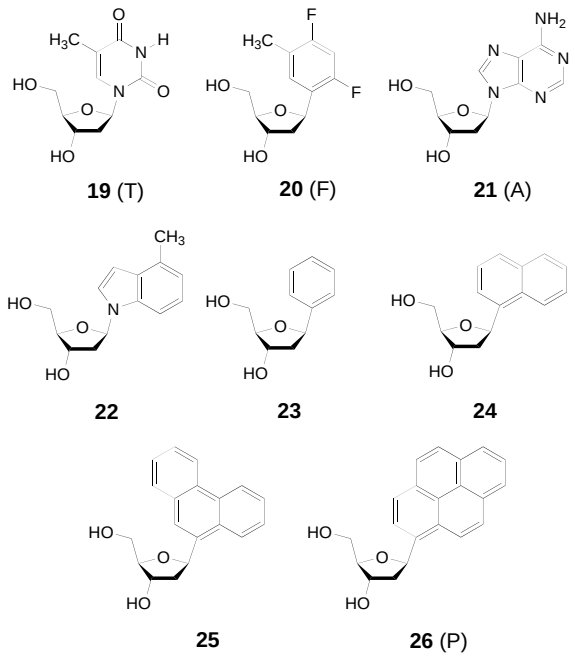
4.2. Forces Contributing to Aromatic Stacking in DNA

Of course, stacking of the bases is a very significant factor in the stabilization of the double helix, and the phenomenon is still relatively poorly understood from a noncovalent bonding standpoint. Several research groups have investigated the stacking phenomenon and the results have spurred a debate in

the literature.[20–26, 75, 76] Some have asserted that electrostatic interactions are a primary driving force for base stacking, while others cite the importance of hydrophobic effects. However, while there has been much discussion on the topic, there has been relatively little study of stacking in the context of DNA itself. For this reason we undertook a quantitative study of the relative stacking abilities of natural DNA bases and nonnatural analogues in the context of double-stranded DNA.[77] The thermodynamic comparisons were carried out by use of a self-complementary six base pair duplex having the sequence 5′-dXCGCGCG in each strand. When the expected B-form duplex is formed, this leaves residue X as an overhanging (dangling) base that can stack on the helix but cannot pair with the opposite strand. Similar studies have been carried out previously with the natural bases in RNA.[78] In the context of DNA, however, there exists (until recently) only one published example of a dangling thymidine in which full thermodynamic parameters were reported,[79] while in another study only T_m values for the four natural bases were reported.[80] In all previous work no unnatural analogues were studied.

The bases studied in our experiments[77] are shown in Scheme 6, and the data are given in Table 3. Of the natural bases, A stacks most strongly, but interestingly, both of the nonpolar isosteres F and Z stack more strongly than any of the natural bases. This is in marked contrast to their strong destabilization when F and Z are in the middle of a duplex opposite a natural base.[44, 55] Clearly, the structural context is important for explaining these differences.

A simple series of unsubstituted aromatic hydrocarbons was studied for comparison, and it was found that all of them stack relatively strongly, with increasing size generally leading to increasing stacking ability. The overall relative stacking abilities of this series of molecules do not correlate well with polarizability or with electrostatic polarization. Instead, we concluded that the most important single factor in this series is the hydrophobic effect, with the largest and most nonpolar molecules giving the best stacking.



Scheme 6. Structures of natural and nonnatural nucleosides studied for stacking propensity.[77]

Table 3. Free energies and melting temperatures for DNA duplexes which contain a dangling base at the 5′end.

Dangling residue X[a,b]	T_m [°C]	ΔT_m [°C]	$-\Delta G^0_{37}$ [kcal mol^{-1}][c]
–	41.0	–	8.1
thymine	48.1	7.1	9.2
difluorotoluene	54.4	13.4	10.7
trimethylbenzene[d]	51.4	9.7	9.7
cytosine	46.2	5.2	9.1
adenine	51.6	10.6	10.1
guanine	51.5	10.5	9.4
4-methylindole	54.6	13.6	11.1
pyrrole	46.6	4.9	8.9
benzene	48.3	7.3	9.4
naphthalene	56.2	15.2	10.9
phenanthrene	57.3	16.3	10.7
pyrene	64.1	23.1	11.4

[a] Measured in the sequence d(XCGCGCG)·d(GCGCGCX) in a buffer which contained 1M NaCl and 10 mM Na-Phosphate (pH 7.0). DNA concentration: 5.0 μM. [b] Data taken from Guckian et al. [77]. [c] Values obtained by plotting $1/T_m$ against $\ln(C_T)$ (C_T = total concentration) with data from at least five concentrations. [d] DNA concentration: 6 μM.

Some useful conclusions and applications arise from these data. First, it seems that the natural DNA bases do not stack very strongly, and it is simple to find compounds that are more efficient at stacking. In retrospect, this seems reasonable, since the DNA helix must be unwound at various points during gene expression and replication. A second observation is that the relatively nonpolar amino acid side-chains of phenylalanine and tryptophan would be expected to stack well with DNA (see the phenyl and indole nucleoside data in Table 3), and such side-chains would therefore make good intercalators in protein–DNA recognition.[81, 82] Finally, a third lesson to be taken from the results is that a useful and simple way to stabilize nucleic acid helices is to place a strongly stacking group at the end. Indeed, the pyrene nucleoside is seen to increase the melting temperature of the test sequence by 23 °C.

The ability to stack strongly can be useful in the stabilization of structures beyond simple double helices. Hairpin secondary structures contain elements of helix as well as loops bridging the ends. The residues in loops often undergo stacking even when they are not strongly contributing to pairing interactions,[83] and so increasing the stacking ability can stabilize looped structures. This has been demonstrated in the context of loops bridging DNA double helices.[84] Table 4 shows the results of one such study. Pentanucleotide loops were used in a duplex sequence, and the effects of replacing the loop residues with nonpolar nucleosides were investigated. The results showed that in all cases the replacement leads to increased stability of the structures.

4.3. The Importance of Solvation Effects in Base Pairing and Stacking

Small nonpolar analogues such as F and Z can stabilize or destabilize DNA structures, depending on where they are substituted. They are generally quite proficient as stacking

Table 4. Free energies and melting temperatures for DNA duplexes which contain tetranucleotide loops.

DNA sequence[a–c]	T_m [°C]	ΔT_m [°C]	$-\Delta G^0_{60}$ [kcal mol^{-1}]
5'-GCAATTGA^{T}T 3'-CGTTAACG$_{T}$T	67.7	–	1.3
5'-GCAATTGA^{F}F 3'-CGTTAACG$_{F}$F	78.5	10.8	3.6
5'-GCAATTGA^{A}A 3'-CGTTAACG$_{A}$A	62.9	–	0.4
5'-GCAATTGA^{D}D 3'-CGTTAACG$_{D}$D	67.2	4.3	1.1

[a] Measured in buffer which contained 100 mM NaCl and 10 mM phosphate (pH 7.0). DNA concentration: 5.0 μM. [b] D is compound **22**. [c] Data taken from Ren et al.[84]

species, but are very poor as base-pairing species. Interestingly, however, there is one situation in which residues such as F and Z *appear* to pair well. When F, for example, is placed opposite A at the end of a duplex, the resulting pair is actually more stabilizing than a natural T–A base pair (unpublished results). We attribute this stabilization to the stronger stacking of F relative to T, without a penalty for desolvation of A. At the end of the helix we surmise that water is able to remain hydrogen bonded to the polar groups on A.

Thus we attribute a number of effects—strong stacking, poor pairing in the center of the duplex, and stable pairing at the end—to solvation effects in water. The strong stacking of nonpolar F and Z (and aromatic hydrocarbons) results from the favorable avoidance of solvent-exposed surface area by adopting the stacked structure. It may be expected that strongly stacking residues may actually stack even in the single-stranded state, which then lowers the entropic cost of duplex formation. It is well known that a significant degree of single-strand stacking occurs within runs of purines in DNA,[85–87] and we have seen that nonpolar nucleosides stack even better than the purines.[77] When a nonpolar "base" is present in the middle of a strand, it may hide both its flat faces from solvent by stacking with the two neighboring bases. At the end, it can still remove one of its faces from solvent by stacking.

An interesting question arises at this point: if only the flat π-surfaces of the nonpolar analogues are removed from solvent upon stacking, then why are the π-surfaces of the natural bases not able to stack with the same affinity? Are not all aromatic π-surfaces the same in solvation and stacking ability? Clearly the edges of the natural bases are much more polar than those of the nonpolar isosteres, but one would guess that the edges of all bases remain exposed to solvent in single-stranded DNA. This implies that the flat surfaces of the aromatic DNA bases are more polar than the analogous surfaces of the nonpolar isosteres, and the data imply that the flat surfaces of DNA bases are more favorably solvated than those of the very nonpolar molecules. It would appear, therefore, that solvent molecules align dipoles to interact more favorably (or less unfavorably) with the local bond dipoles that exist at the flat π-surfaces of the natural bases. In the nonpolar analogues these dipoles are generally small or are absent.

Almost all of the nonpolar nucleoside analogues exhibit strong destabilization in the center of a DNA duplex when paired opposite natural DNA bases, as mentioned above. This is very likely due in part to the cost of desolvation of the polar edge of the natural base[74] when the nonpolar base is forced against it in the helix. This is despite the very strong stacking of the nonpolar molecules, which is apparently not large enough to compensate for the cost of desolvation.

5. Testing DNA Replication Mechanisms

5.1. Previous Studies

This work has been aimed at investigating the precise noncovalent interactions that govern the formation of base pairs. The work described above focuses on the inherent properties of DNA bases and analogues in forming base pairs. Judging by the results with analogues F and Z, it appears that Watson–Crick hydrogen bonds do add a significant level of stabilization to DNA duplexes, and that they are very important to the selectivity of base pair formation. Due to their lack of hydrogen-bonding groups, F and Z show 1) inherently strong destabilization in their interaction with natural bases, and 2) very little, if any, pairing selectivity.

Once the inherent properties of the nonpolar nucleoside mimics were determined, we began using them to test the mechanisms governing the insertion of nucleotides into DNA by polymerase enzymes. When this work began, it was widely believed that Watson–Crick hydrogen bonds are essential for replication of a DNA base pair by polymerases. For example, recent editions of biochemistry and molecular biology textbooks argue that hydrogen bonds alone may explain the observed fidelity.[41, 42]

If one examines the data gathered by use of modified nucleosides up to 1995, one can find a number of examples of nucleosides such as *O*-6-methylguanosine (O^6meG, **9**; Scheme 4). This compound has been altered from natural guanine by methylation. This blocks Watson–Crick hydrogen bonding by the carbonyl, and changes N1 from a hydrogen-bond donor to an acceptor. Studies with the Klenow fragment of *E. coli* DNA polymerase I, perhaps the most well studied of DNA polymerases, have shown that this compound is a poor substrate.[88] When placed in a template strand of DNA, the enzyme inserts nucleotides opposite with efficiency (v_{max}/K_m; v_{max} = maximum rate of reaction, K_m = Michaelis constant) that is 5000 times less than insertion of C opposite unmodified G (Figure 4). Moreover, when nucleotides are inserted, there is a tendency for insertion of both C and T.

It has been common to attribute these changes in the polymerase substrate properties of compounds like O^6meG to their altered hydrogen-bonding properties. Indeed, since hydrogen bonding to cytosine is blocked, this might be a viable explanation for such a result. However, one must also consider other factors as possible explanations. Specifically, the steric size and shape of the molecule is altered by such substitution; the methyl group adds 2–3 Å of bulk to the O6

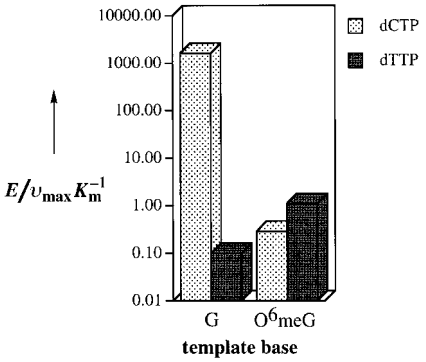
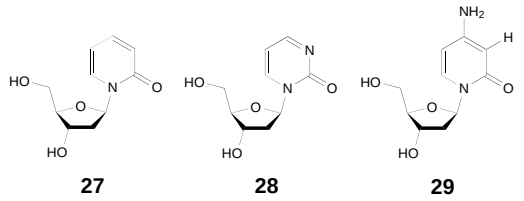


Figure 4. Histogram showing the steady state insertion of dCTP and dTTP opposite O^{6}meG and, for comparison, G. Data was taken from Levy and Spratt.[88]

position, and the loss of the proton removes approximately 0.7 Å of bulk at N1. Since both the hydrogen bonding and the sterics are affected, one cannot be confident in attributing the polymerase activities with this compound to only one of these noncovalent interactions.

A previous series of studies was carried out to address the importance of hydrogen bonds in DNA replication.[89] It involved new pyrimidine analogues that were missing a single hydrogen-bonding group. These analogues are shown in Scheme 7. Primer extension experiments were carried out

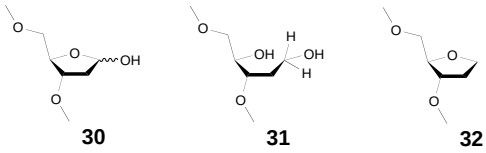


Scheme 7. Deoxycytidine analogues studied by Strazewski and Tamm[89] for their ability to be accepted as polymerase substrates.

with these analogues both in the template and as nucleoside triphosphates. As triphosphates none of these compounds were found to extend the primer. In the template, however, compound **29** (a hydrogen bonding impaired analogue) was found to code for dCTP incorporation. Unfortunately, no quantitative data are available for these interesting analogues, and so we must rely on qualitative observations. The general conclusion made from this work is that the chief factor governing polymerase activity is "the number and strength of hydrogen bonds" between bases in a pair.[89] As we have pointed out, however, some of the steric changes made to these analogues may confuse the situation somewhat. Lastly, some of the changes lead to possible tautomerization of the DNA base, which confuses both the issues of hydrogen bonding and steric effects. In addition, partial deletion of hydrogen-bonding groups also raises issues of solvation that may lead to problems (see Section 4.3).

Another important set of studies has been carried out by Benner and co-workers,[70, 90, 91] who have made analogues following a different strategy. They have designed pairs in which hydrogen bonds are not deleted, but are rather rearranged (Scheme 5) to be different from that in A–T and C–G (purine–pyrimidine geometry is maintained). Two such base pairs have been shown qualitatively to be replicated efficiently by DNA polymerase (Klenow enzyme). This is a nice example suggesting the possibility that base pairs other than the natural ones might be used to expand the genetic code. These analogues contain hydrogen bonds and maintain Watson–Crick geometry; therefore they do not allow us to address the importance of hydrogen bonds and geometric effects in replication.

Other potentially relevant modified nucleotides that can help address the importance of hydrogen bonds in DNA replication are the abasic nucleosides, such as **30**–**32** (Scheme 8).[92–94] Naturally occurring abasic sites are one of the most common forms of DNA damage and are formed by alkylation damage to DNA bases followed by hydrolysis of the glycosidic bond, base excision repair, and spontaneous hydrolytic loss of purines.[95–97]



Scheme 8. Structure of the abasic site (**30**) and abasic site analogues (**31**, **32**).

Abasic sites have been termed "non-instructional" because of their lack of hydrogen-bonding information.[98, 99] Interestingly, it has generally been observed that deoxyadenosine is preferentially inserted opposite abasic sites in a template strand by most DNA polymerases.[94, 100–103] This default choice by the enzyme is commonly referred to as the "A rule". The A rule has been variously explained by an inherent preferential interaction with A by the enzyme, or by the lower polarity or stronger base stacking ability by A relative to the other nucleotides. Kinetics studies have shown that polymerases insert dA opposite abasic sites with efficiency approximately three to four orders of magnitude lower than for adenine opposite T, which would tend to argue that hydrogen bonding does provide a significant benefit. However, the abasic site may lead to DNA distortion,[104–107] which could also explain the low efficiency. Also of interest is the magnitude of selectivity observed for inserting A opposite abasic sites. The kinetics experiments show that the selectivity is actually rather low, with an approximately tenfold preference for dA rather than dG insertion.[92] Although abasic sites do provide interesting information on what happens without hydrogen bonds, the lack of any base structure to test steric effects, and the large structural perturbation caused by the lack of a base makes it difficult to judge the effects of sterics and hydrogen bonding.

To summarize, in the general literature on polymerase mechanisms and fidelity, Watson–Crick pairing has most commonly been cited as the important factor which drives incorporation of a nucleotide and governs the specificity of this process.[41, 42] It is worth noting that some workers have also pointed out the relevance of maintaining a geometry similar to standard base pairs.[108–110]

5.2. Predictions of the Hydrogen-Bonding Model

Before experiments with DNA polymerases were carried out, we considered what the standard model would predict. A purely hydrogen-bonding model for replication would suggest that analogues such as F would be very poor substrates for polymerases. In addition, if they were made to be substrates under forcing conditions, they would be expected to be handled with little or no selectivity. This would not be surprising, since the inherent property of F is very unstable and nonselective pairing. Indeed, such thermodynamic base-pairing measurements (done in the absence of enzymes) are routinely carried out as predictors of polymerase activity with the same base pair.

5.3. Studies with Nonpolar Thymine Analogue F

The difluorotoluene analogue was placed in synthetic DNA templates and tested for its ability to be accepted by a DNA polymerase. For the initial studies we carried out experiments with the Klenow fragment of *E. coli* DNA polymerase I, which is among the best studied of DNA polymerases.[34, 111–113] Immediately we observed a number of surprising results. In an experiment with all four nucleotides present, the polymerase proceeded past F with no apparent hesitation, and one of the natural nucleotides was inserted directly across from F (Figure 5).[114] Single nucleotide additions were then carried out, and these revealed that it was adenine that was selectively inserted. Quantitative steady-state kinetics were then measured for the A–F pair, and it was found that efficiency of the base-pair synthesis was only approximately fourfold lower than that of the natural pair (Figure 6). Even more surprisingly, the selectivity of adenine insertion (rather than the other nucleotides) was nearly as high as that of the natural pair. Since we were not able to measure any hydrogen-bonding ability for F, we proposed that effects other than hydrogen bonding should be considered as important to polymerase activity. Clearly the enzyme enforces properties on F that it does not display on its own in the absence of enzymes. We proposed at the time that steric effects might be sufficient to explain the results.[114]

Of course, when such a surprising result is observed, then alternative explanations must be considered. For example, one might conclude that since F does not hydrogen bond well, it acts essentially as a non-instructional base, and the enzyme simply follows the A rule when it selectively inserts A. We pointed out, however, that the A–F pair is considerably more efficiently processed than occurs with the A rule, and that the A–F pair does not cause strong pauses as the A–abasic pair does[114]. Nonetheless, the best way to disprove the A rule hypothesis is to turn the base pair around. We therefore synthesized the nucleoside triphosphate derivative of F, and studied its insertion by the polymerase.[115] It was found that insertion was still highly efficient, and that the selectivity was as high as for the natural pair (Figure 7). We concluded that the A rule could not be invoked here. We proposed that hydrogen bonding in DNA synthesis was not as important as

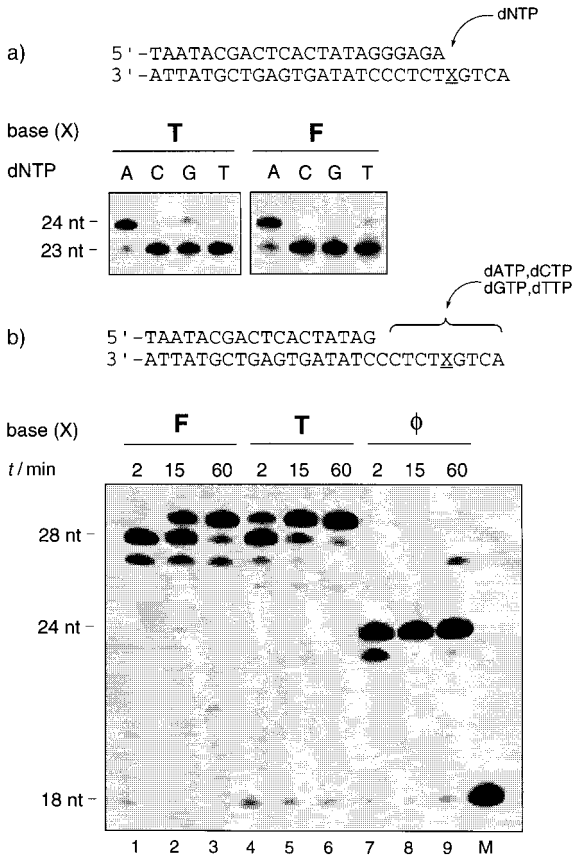


Figure 5. Polymerase insertion of nucleotides opposite difluorotoluene, thymine, or an abasic site. a) Autoradiogram showing single nucleotide insertions of the four natural bases on 23/28 nucleotide primer–template duplexes (d(TAATACGACTCACTATAGGGAGA)-d(ACTGXTCTCC-CTATAGTGAGTCGTATTA)) which contain either thymine or difluorotoluene in position X.[114] b) Autoradiogram showing running-start multiple nucleotide insertion experiments on 18/28 nucleotide primer–template duplexes (d(TAATACGACTCACTATAG-d(ACTGXTCTCCCTATAGT-GAGTCGTATTA)), with mixtures of the four natural nucleosides.[114]

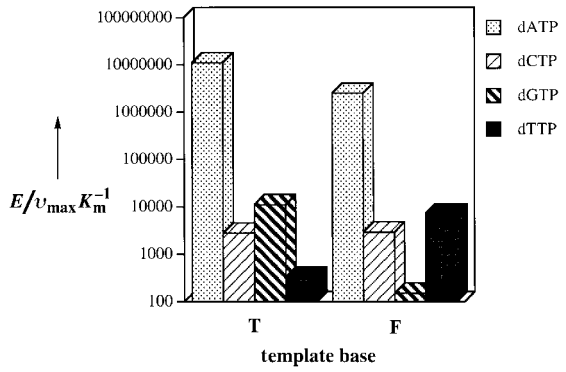


Figure 6. Histogram of fidelity of nucleotide insertion opposite T and isostere F using the primer–template shown in Figure 5a.[114]

commonly believed, and that perhaps steric effects might play a significant role.

It was later shown that the F–A pair could replace every T–A pair in a 48 base-pair duplex, with faithful synthesis occurring in six different DNA contexts.[116] Ironically, the DNA being synthesized was less stable than the natural DNA by a significant amount. Clearly, however, the DNA polymerase does not sense this destabilization until after the transition

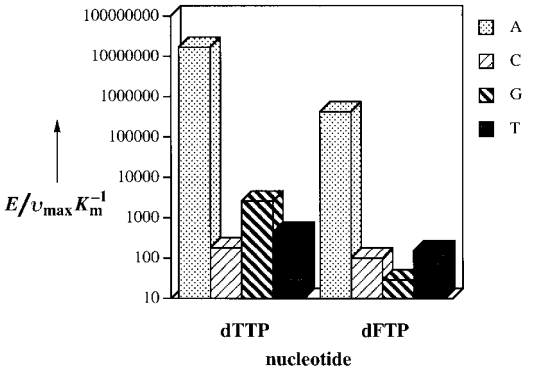


Figure 7. Histogram of selectivity of dTTP and dFTP insertion opposite the four natural bases using the primer–template shown in Figure 5a.[115]

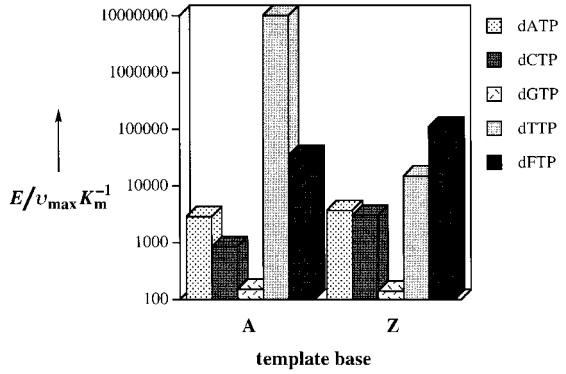


Figure 8. Histogram of fidelity of nucleotide insertion opposite A and isostere Z using the primer–template shown in Figure 5a.[117]

state for nucleotide insertion. Interestingly, the DNA polymerase does apparently sense destabilization when multiple F–A pairs are being synthesized in a row.[116] We suspect that this might arise from the enforced desolvation of the F–A pairs when they become buried further into the active site.

Later studies have also tested various polymerase enzymes for their ability to process the F–A pair (unpublished results). It was found that *Taq* polymerase, T7 DNA polymerase, and HIV reverse transcriptase can all handle the pair qualitatively very well. Quantitative data are not yet available.

5.4. Studies with Nonpolar Adenine Analogue Z

The results with analogue F raised questions about generality and mechanism. Would similar results be found for other nucleoside isosteres? Or were the fluorine substituents responsible for the special activity of F? Indeed, one possible explanation for the high efficiency of the F–A base pair in replication is that F forms a hydrogen-bonded pair with A. In addition, another possible influence on the selectivity of the F–A pair might arise from solvation effects; since A is the least strongly solvated of the natural bases, this would favor its pairing with a nonpolar analogue such as F.

Thus, we carried out a series of replication experiments with a new analogue, Z (Scheme 1), which has a 4-methylbenzimidazole group replacing adenine in dA.[44] Models suggest that it can fit opposite F in a DNA duplex with only a relatively small structural distortion arising from the bulk of the H1 proton on Z (a group that is absent in A). Both qualitative and quantitative experiments were carried out with the Klenow fragment enzyme. Consistent with the previous results with analogue F, we observed relatively efficient base-pair synthesis when Z was present in a template or as a deoxynucleoside triphosphate (Figures 8, 9). For example, F and Z are replicated very well against one another, despite the indisputable lack of hydrogen bonds between the two. Also interesting is the fact that analogue Z is not as well inserted opposite another Z. This provides an example of selectivity in the absence of hydrogen bonding. Thus we concluded that efficient base-pair synthesis by this polymerase does not require Watson–Crick hydrogen bonding.[117] In addition, steric effects alone (in the absence of these hydrogen

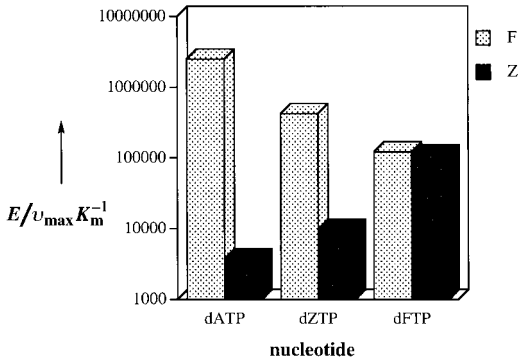


Figure 9. Histogram of selectivity of dATP, dZTP, and dFTP insertion opposite isosteres Z and F using the primer–template shown in Figure 5a.[117]

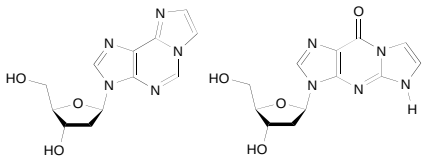
bonds) can result in significant pairing selectivity. Importantly, the selectivity observed with analogues F and Z (using the Klenow fragment enzyme) is much higher than the very low or nonexistent pairing selectivity observed by these "bases" in the absence of the enzyme.[44] Thus, it appears that the enzyme provides an active site which closely conforms to the Watson–Crick duplex DNA geometry, and which strongly restricts the flexibility of the DNA backbone and bases.

In our studies with analogues F and Z we observed one way in which the members of this pair behave very differently than the natural bases T and A. It was seen that dFTP is actually inserted relatively efficiently opposite another F in the template (Figure 9).[115, 117] This is in contrast to the analogous natural pair, T–T, which is synthesized 900-fold less efficiently by the KF enzyme. Possible reasons for this difference are discussed in Section 5.7.

5.5. Other Hydrogen Bonding Impaired Nucleotides

The activity of DNA polymerases in handling other hydrogen bonding impaired nucleotides has been recently reviewed,[118] and so the data will be discussed here only briefly. Previous to the work with difluorotoluene and benzimidazole analogues, no polymerase studies had been carried out with nucleotide analogues that completely lack hydrogen-bonding ability. The ethenoadenine and ethenoguanine deoxynucleoside adducts **33** and **34**, respectively, are potentially of interest

(Scheme 9), but εG **34** can still potentially form one Watson–Crick bond,[119] and εA **33** can form hydrogen-bonded complexes by adopting a *syn* glycosidic orientation.[120] As a result, the only entirely convincing results in the absence of hydrogen



33 **34**

Scheme 9. Structures of 1,6-*N*-ethenodeoxyadenosine **33** and 1,2-*N*-ethenoguanine **34**, which are formed upon alkylation of dA and dG with vinyl chloride.

bonds between the bases comes from the previously mentioned work with abasic sites[92–94] and the nonpolar analogues described above. Interestingly, there exists one isolated 1973 report of the study of benzimidazole deoxynucleoside being incorporated into the DNA in cells.[121] No studies were carried out with a purified polymerase or with an isolated DNA sequence, and so no quantitative data are available for that analogue. Another interesting report was published in 1984, in which benzene deoxynucleoside was incorporated into synthetic DNA.[66] No polymerase work was carried out. Interestingly, the workers reported that it formed very unstable pairs with natural bases, but this was attributed to poor stacking ability, something that we know now not to be the case.[77]

5.6. Discussion of Possible Explanations

The work with the dF and dZ analogues provokes an important and provocative question: Is it true that Watson–Crick hydrogen bonds are completely dispensable in efficient and specific replication of DNA base pairs? Before this can be safely concluded, we must consider alternative explanations that might conform to our results. Perhaps analogue F functions correctly because it is in fact forming efficient hydrogen bonds with A.[51] However, it seems clear that this is not the case. First, we see that there is no measurable complex between difluorotoluene and adenine even in a solvent that strongly encourages such bonding.[55] Second, dF is highly hydrophobic, while dT is the opposite, highly hydrophilic. Third, F shows no tendency to base pair with adenine, even when forced to be opposite it in a DNA helix. Fourth, F shows no measurable selectivity in pairing with A, while thymine displays 3–4 kcal mol^{-1} of selectivity in the same context. Fifth, although calculations suggest that there might be one weakened hydrogen bond between F and A in the gas phase,[52, 54] there appears to be general agreement among most computational scientists that in water there will not be nearly enough favorable interaction to explain our results.[52–54, 122] Finally, hydrogen bonds are completely ruled out in cases such as the pair between F and Z, which is still processed relatively efficiently by the enzyme.

A second explanation for why adenine is preferentially inserted opposite F starts with the opposite assumption, that F is nonpolar and does not form hydrogen bonds. One might argue that F is another non-instructional lesion, analogous to abasic sites. Once again, experiments do not support this. First, when A is inserted opposite abasic sites there is a strong pause, while no such pause is observed for A insertion opposite F.[114] Second, the insertion of A opposite F is two orders of magnitude more efficient than insertion of A opposite abasic sites, and the specificity opposite F is two orders of magnitude higher. So in these respects, F is clearly not behaving in a manner analogous to other non-instructional lesions in DNA. Probably the most convincing result, however, comes when the F–A pair is turned around.[115] The data show that F is very well inserted opposite A, and once again it is processed with very high specificity. In this case adenine is not being inserted, and the A rule cannot be invoked. For these reasons, this second explanation does not hold water.

The final alternative explanation to be considered is that solvation effects might favor enzymatic pairing of F with A and vice versa. Indeed, calculations suggest that the Watson–Crick edge of adenine is less strongly solvated than the analogous parts of the other DNA bases.[74] For this reason, it might cost less energy to enzymatically pair F and A together as opposed to F with T, C, or G. However, if this explanation were correct, then why does F not show selective pairing with A in the absence of enzymes? In addition, when analogue Z is studied in the template strand, the polymerase selectively inserts T,[117] although it is among the most strongly solvated of the four bases.[74] Thus one suspects that most of the available results do not support this solvation effect as the most important answer, superseding steric effects. In any case, probably the best way to rule out such effects is to look at pairs in which neither partner carries strongly solvated hydrogen-bonding groups. Thus, we find that the pair of F with Z is indeed handled with good efficiency and with significant selectivity (the results discussed in Section 5.4).

Thus, overall we feel confident in stating that hydrogen bonds are not necessary to achieve highly efficient base-pair synthesis, and that significant levels of selectivity can be achieved in the complete absence of these bonds. However, it is important to note that the replication of Z opposite F is not entirely as efficient as a natural base pair, and that the misinsertion of Z opposite Z is more favorable than the analogous A opposite A.[117] Thus, one should not yet rule out a smaller favorable effect of hydrogen bonding on efficiency and especially specificity.

5.7. The Steric Exclusion Model for DNA Replication

The results in Section 5.6 led us to propose a model for replication of DNA base pairs which does not rely on Watson–Crick hydrogen bonds to explain efficiency and selectivity. In this steric exclusion model,[118] the polymerase active site is relatively rigid and presents a binding pocket of specific size and shape. This active-site pocket has four possible shapes, since it is defined both by the structure of the

polymerase and that of the template base immediately downstream of the primer. The incoming nucleotide is fixed in position for polymerization (phosphodiester bond formation) by the binding of the triphosphate moiety and by proper fit of its DNA base. The forces that support binding of the nucleoside triphosphate at the transition state are binding of the triphosphate and sugar, and stacking of the base on the primer 3′ end base. In the simple steric exclusion model, hydrogen bonding between the two bases does not add much energy of stabilization at the transition state. The rationalization for this is that both the template base and the incoming nucleotide base are already hydrogen bonded to water prior to the reaction, and so there is little stabilization to be gained by exchanging these for Watson – Crick bonds.

How does the steric exclusion model address the issue of fidelity of DNA synthesis, that is, the high selectivity in choosing to insert the correct nucleotide rather than the other three? Again, while the steric exclusion model does not rule out additional contributions from other factors such as Watson – Crick hydrogen bonding, it is worthwhile to consider how steric effects alone might function to give high selectivity. It is clear that the assumption of a rigid active site which provides a close fit to duplex DNA geometry[39, 123–131] will provide an explanation for why purines are not misinserted opposite other purines; they are simply too large to fit opposite one another (Figure 10). In addition, one can use a steric exclusion argument to explain why purine – pyrimidine mismatches (G – T and A – C) are not well synthesized. If one tries to enforce a standard base-pair geometry on them, there will be steric overlap of two protons in either type of pair. In addition, if a wobble geometry is adopted, then the active site will again provide a steric clash between the protein and the base pair. Thus, for these types of mismatches a simple steric exclusion argument fits the data well.

However, there appears to be a serious shortcoming in this simplistic model: how does one explain why the sterically small pyrimidines are not misinserted opposite other pyrimidines? Clearly, the low frequency of these mispairings[132] deserves an explanation, and steric effects as described at the beginning of this section would not account for this, since two pyrimidines can easily fit into the space normally occupied by a pyrimidine paired with a purine. Here we suggest that one must take into consideration the tightly bound solvent molecules that associate themselves strongly with the polar groups on DNA bases. Each of these groups is hydrogen bonded to at least one water molecule. One such hydrogen bond will be bound by 6 – 7 kcal mol^{-1} in terms of gas phase enthalpy.[18, 19] While such a water molecule is exchanging rapidly with others in solution, the solvating water is essentially never lost because of this energetic cost of breaking the bond. As a result, the only time that a solvating water is lost in aqueous solution is when it is

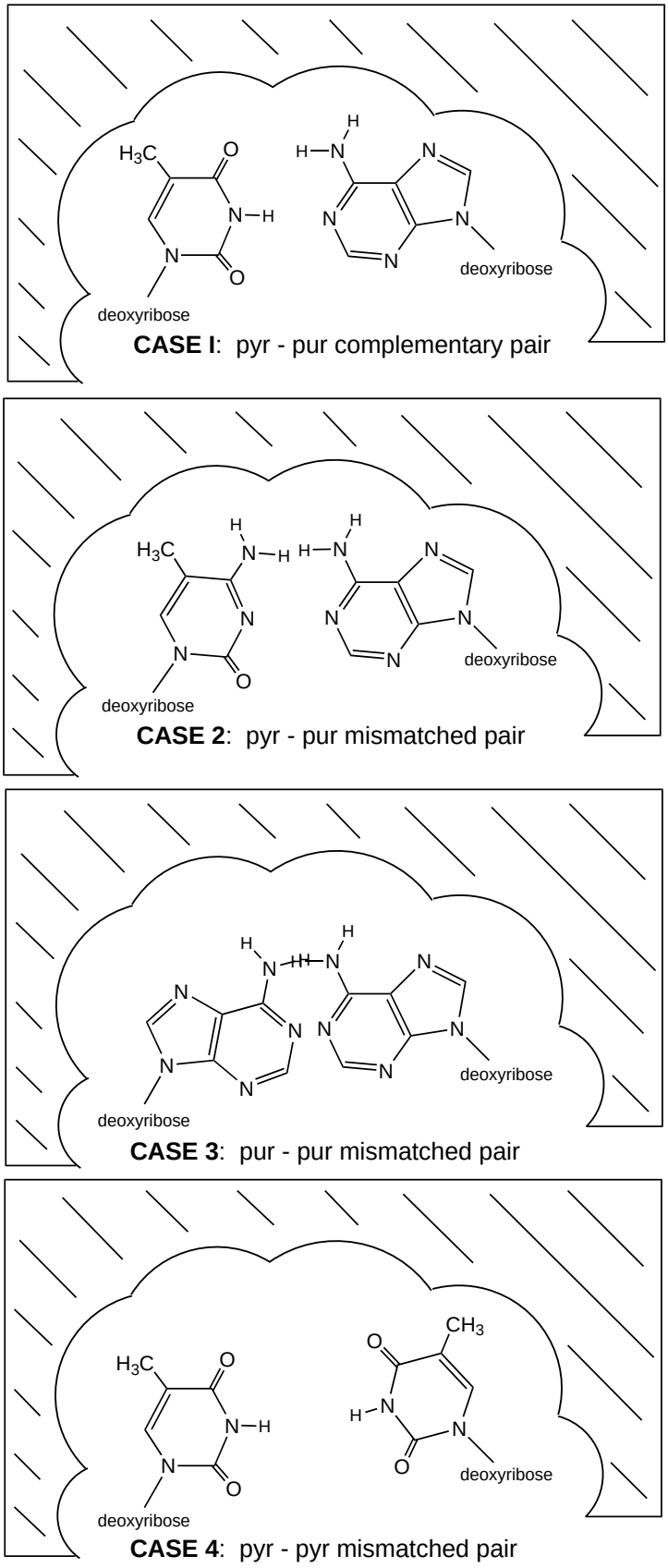


Figure 10. Representation of steric clashes that could lead to the exclusion of mismatches from the polymerase active site in at least three of the four possible classes of mismatches. The polymerase DNA binding cleft tightly surrounds the template, leaving an active-site cavity for the triphosphate, defined by the polymerase and the template base.

simultaneously exchanged for another complementary, energetically equivalent hydrogen bond.

We argue that the strongly bound waters of solvation along the Watson–Crick pairing edge of a natural DNA base serve to make the base sterically larger than we commonly draw it (Figure 11). Therefore, pyrimidines (indeed, all bases) are

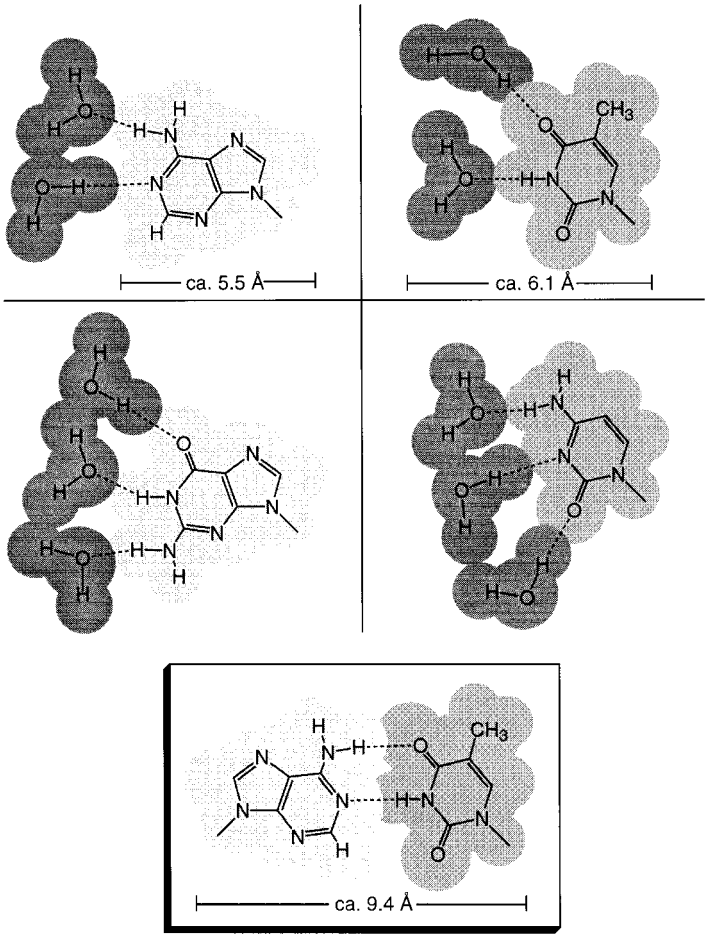


Figure 11. Schematic representation of possible geometries of solvating waters on the four DNA bases. Also shown are the estimated sizes of the van der Waals surface of a solvated pyrimidine and purine base; this shows how solvation greatly increases the size of the DNA bases in water. A solvated pyrimidine may be two-thirds the size of an entire base pair!

larger than we commonly envision them, and this added steric size can have energetically important effects at the transition state for DNA polymerization. Thus, it is argued that pyrimidines are not misinserted opposite other pyrimidines because they are sterically too large. They will not give up these waters (to become small enough to fit) because they cannot exchange them for energetically equivalent hydrogen bonds.

Is there any evidence available for the steric effect of waters of solvation? Here we describe two pieces of experimental data, one with polymerases and one with small molecules. First, we consider the classic measure of the steric size of a functional group in organic chemistry: the *A* value, which is measured for substituted cyclohexanes. The *A* value represents the magnitude of steric clashes with axial protons, and it has been recognized for some time that polar groups give larger values in protic than in aprotic solvents.[133] For example, the alcohol of cyclohexanol has an *A* value of 0.52 in aprotic solvent and 0.87 in protic solvent. A similar trend is seen for the amine of cyclohexylamine, which has an *A* value of 1.2 in aprotic solvent and 1.6 in protic solvent.

The second piece of evidence supporting the influence of solvating waters on the steric size of DNA bases comes from study of our nonpolar isosteres. As mentioned above, we observe that F is inserted opposite another F 900-fold more efficiently than T is inserted opposite another T.[115] We attribute the facile F→F base-pair synthesis to the fact that F is small enough to be accommodated opposite another F residue. Conversely, T with its waters of solvation is too large to fit opposite another T (which is also solvated).

The "jigsaw puzzle" prediction

This steric exclusion model for DNA polymerase activity leads to some testable predictions about new experiments that could be performed.[118] First, it predicts that in general it will be possible to replicate base pairs between nonpolar, non hydrogen bonding residues as long as they are sterically well matched in the double helix. The chief factors in their efficiency will be their avoidance of steric clashes and their stacking ability. Second, at least some measure of selectivity among various non hydrogen bonding bases will be achievable by careful selection of their molecular shapes. If a selective pair that does not suffer from self-pairing (as occurs with F) is the goal, then a self-exclusive shape must be part of the design. Thus, a successful base pair will be matched like two adjacent pieces of a jigsaw puzzle. Finally, the shape exclusion hypothesis predicts that purine and pyrimidine shapes will not be a requirement for efficient synthesis; any flat aromatic shape should be acceptable as long as the pair fits within the overall context of duplex DNA.

5.8. Second Generation Base-Pair Design

Figure 12 shows the structures involved in a new base pair that tests some of the above predictions. This pair, P–ϕ, is a somewhat extreme example testing the concept that pyrimidine and purine shapes might not be required for polymerase

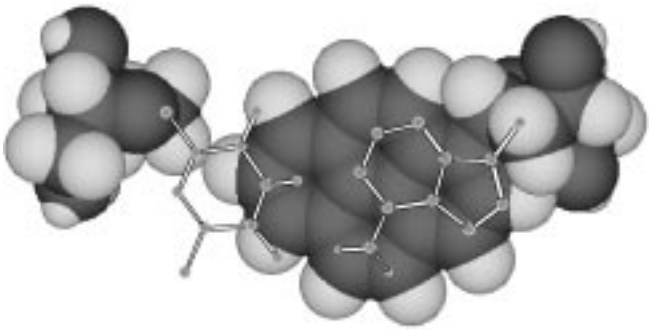


Figure 12. Space-filling model of a pyrene abasic pair within the context of B-form duplex DNA. For comparison a Watson Crick A–T base pair is superimposed on top of the space-filling model.

activity. The pair consists of pyrene deoxynucleoside **26** and an abasic nucleoside **32**. In this pair, one partner is sterically as small as possible (the "base" is a proton), and the other is nearly as large as possible, filling the void left when a base is removed. This novel example has now been tested, both as a base pair in the absence of enzymes,[72] and more recently, as a substrate for DNA synthesis by polymerases.[134]

Table 5 shows data illustrating the pairing properties of pyrene and the abasic nucleoside[72] in a DNA duplex without enzymes. Interestingly, the pyrene – abasic "pair" is nearly as stabilizing as an A – T pair in the same sequence. This is surprising because abasic sites are notoriously destabilizing to DNA when paired opposite natural bases[66] (see examples in the Table). As a result of the P – ϕ stability, one notes that the abasic site actually shows high selectivity in its pairing preference for pyrene over natural bases. On the other hand, pyrene shows only a small preference for pairing with ϕ rather than the natural bases. We presume that in the latter case, pyrene is intercalated between bases on the opposite strand (structural studies are underway to address this question).

Table 5. Free energies and melting temperatures for DNA duplexes which contain pyrene and abasic sites.

Base pair X – Y[a]	T_m [°C][b]	ΔT_m [°C]	$-\Delta G^0_{37}$ [kcal mol^{-1}][c]
T – A	43.2	–	12.6
P – A	38.7	4.5	10.8
P – C	37.6	5.6	10.5
P – T	36.4	6.8	9.6
P – G	38.2	5.0	10.7
P – ϕ	41.0	2.2	11.6
P – P	42.9	0.3	11.8
ϕ – A	22.0	21.2	7.0
ϕ – C	20.9	22.3	6.7
ϕ – T	18.0	25.2	6.2
ϕ – G	23.1	20.1	7.2

[a] Measured in the sequence d(CTTTTCXTTCTT) · d(AAGAAY-GAAAAG) in a buffer which contained 100 mM NaCl, 10 mM $MgCl_2$ and 10 mM Na-PIPES (pH 7.0). DNA concentration: 5.0 μM. Error in Tm is ±5%. [b] Data taken from Matray and Kool.[72] [c] Values obtained from the fit of at least five denaturation curves.

We then examined the more difficult questions: would the unusually large, non hydrogen bonding pyrene nucleotide behave as a substrate for DNA polymerases? Would the abasic site act as a viable template? And would the efficiency be low or high? Would any selectivity be observed, since there are no hydrogen bonds in the pair? To test these questions we synthesized the deoxynucleoside 5′-triphosphate derivative of P (dPTP), as well as template DNAs containing the abasic analogue ϕ. These were studied with the Klenow DNA polymerase as well as T7 DNA polymerase.

Surprisingly, the results show that dPTP is, in fact, a substrate for DNA polymerases (Figure 13).[134] More surprisingly, it is a very good substrate, with efficiency approaching that of a natural base pair, and a very selective one, being inserted much more efficiently opposite ϕ than opposite natural bases or another pyrene (Figure 14). Once again, the polymerase is clearly enforcing properties of pyrene base that it does not display in DNA without an enzyme.

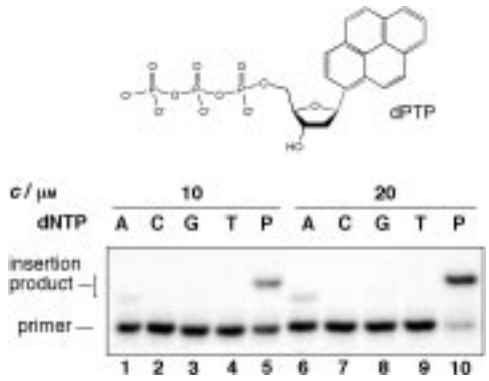


Figure 13. a) Structure of dPTP, which was shown to be an efficient substrate for DNA polymerases. b) Autoradiogram showing single nucleotide insertions of the four natural nucleotides and dPTP opposite the natural bases or an abasic site. For sequences see Figure 5a.[134]

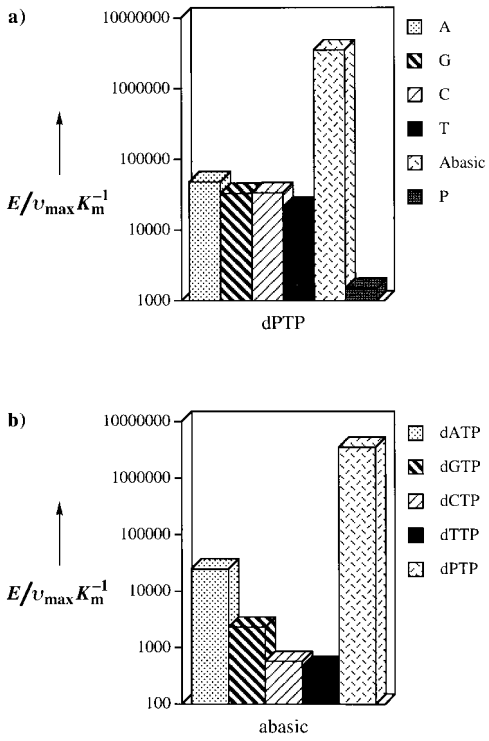


Figure 14. Histograms showing the steady state efficiency of insertion of dPTP: a) Selectivity for insertion of dPTP, showing preferential insertion opposite the abasic site rather than the four natural bases or pyrene itself.[134] b) Fidelity of insertion of the four natural nucleotides and dPTP on the primer – template shown in Figure 5a where X is the abasic site.[134]

The clear conclusions of this work are, first, that hydrogen bonds are not required to reach very high base-pair synthesis efficiencies. Second, significant levels of selectivity can be achieved without hydrogen bonds; in the present case, they apparently arise from simple steric effects alone. Third, base structures other than ones resembling purines and pyrimidines can still be very effective as substrates, as long as they are sterically matched against an appropriate partner. This bodes well for the design of new base pairs in the future.

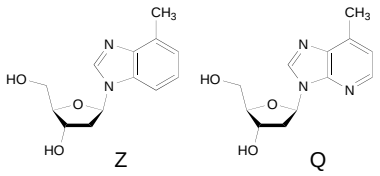
It should be pointed out that the P – ϕ pair is quite different from a natural pair in at least one important way. After the pyrene is inserted, the DNA synthesis is halted; that is, the

strand is extended no further.[134] It is not yet known why this happens, but it may be due in part to some property of the abasic site. For example, extension of the DNA strand containing an abasic residue also proceeds very poorly even when a natural base (adenine) is inserted opposite it.[135] In the pyrene study, we took advantage of this chain termination after pyrene insertion by using it as a DNA sequencing agent for detecting abasic sites in DNA.

5.9. Other Polymerase Effects

The results described in this review tend to support the idea that Watson–Crick hydrogen bonds are not essential to DNA replication, at least for an isolated base pair. In some ways, this makes good sense, because such bonds are internal to the DNA helix and so their presence or absence cannot be easily sensed structurally by the enzyme. Recent X-ray crystal structures have shown, however, that polymerase enzymes do, in fact, make use of hydrogen bonds between the protein and the DNA.[123, 128, 129, 131, 136] These interactions occur between hydrogen-bond donor side-chains and acceptor groups in the DNA minor groove (this is referred to as the "minor groove binding track").[137]

This fact may help explain why some nonpolar base analogues, such as the benzimidazole analogue Z, are inserted well but then give significant inhibition of further chain elongation.[117] We proposed that moving Z downward in the active site (so that the next nucleotide can be inserted) causes the hydrogen-bond donor in the active site to lose its hydrogen bond with the previous base. In effect, it becomes desolvated, which is energetically costly, and this causes the enzyme to dissociate. We proved this by synthesizing a new analogue, Q, which is the same as Z (it still lacks Watson–Crick hydrogen-bonding groups) but has one nitrogen engineered back into the minor groove (Scheme 10). This modified base is a very good substrate, both for insertion and for further chain elongation (Figure 15).[138]



Scheme 10. Structure of the nonpolar isostere Q and, for comparison, the previously described nonpolar isostere Z. These analogues are used to probe DNA minor groove interactions with polymerases.

6. Conclusions

In the absence of enzymes, it appears that hydrogen bonding may indeed add a significant degree of stabilization to existing DNA helices, and that it does add a strong component of pairing specificity. However, it appears that base stacking is also a very important factor for stabilization of the double helix, even for the natural DNA bases, which stack relatively weakly as compared to other less polar molecules.

Figure 15. Autoradiogram showing minor groove interaction effects on extension of normal and modified base pairs.[138] Note the higher efficiency of primer extension with Q (lanes 15–18) as compared with Z (lanes 9–12).

In DNA replication, we conclude that Watson–Crick hydrogen bonds are not required to replicate a base pair with very high efficiency. In addition, significant selectivity can also be achieved from steric effects alone. It remains to be seen, however, whether steric effects in the absence of hydrogen bonding can give the very high levels of fidelity observed for normal hydrogen-bonded base pairs. Regardless, it is clearly true that one must pay close attention to steric and geometric factors in the DNA if one is to understand how base pairs are replicated, and if one aims to design new replicable base pairs.

Moreover, it is clear that polymerase enzymes closely enforce geometric constraints on DNA beyond what it feels from its own backbone constraints. Thus, base-pairing stabilities in the absence of enzyme correlate poorly with the ability to be replicated by an enzyme. This means that even base pairs that are very unstable and nonselective in DNA alone may be handled with high specificity and efficiency by an enzyme.

7. Future Work

There remains much work to be done in this field. Synthesis of a number of new analogues is warranted, to test the effects of subtle changes in molecular shape on selectivity in base-pair synthesis. In addition, studies with a greater number of polymerases would test the generality of the effects observed. Testing other classes of polymerases, such as RNA polymerases and reverse transcriptases, would also be of interest.

These analogues could also be valuable in the study of other activities of polymerase enzymes. For example, extension of base pairs is distinct from insertion of base pairs, and this involves different interactions with the enzyme. In addition,

the proofreading (mismatch correction) activity of polymerases could also be tested. In this case, a similar question can be addressed: is it hydrogen bonding or base-pair shape, which governs the rate at which a base pair is removed by the proofreading apparatus of the polymerase?[139, 140]

Other interactions of proteins and DNA are also of interest. Since the difluorotoluene–adenine pair does not distort the DNA,[56] the analogue could be very useful in probing the importance of hydrogen bonding and solvation on protein recognition of the DNA structure. In addition, it could serve as an NMR structural probe; analogue F contains two fluorines that are located in the center of each groove of the DNA and this provides a useful ^{19}F probe for NOE studies.[56]

DNA repair mechanisms are also currently under intensive study. For example, the base excision repair enzymes remove mismatched DNA bases by "flipping" them out of the helix and hydrolyzing the C–N glycosidic bond.[141] Analogues such as F would not be hydrolyzable by such enzymes, and so they might serve as specific inhibitors in synthetic DNA substrate sequences.

In addition to using such analogues as tools for studying mechanism and structure, we envision a number of potentially useful practical applications in biomedicine. For example, the use of analogues such as the pyrene nucleoside P might give strong stabilization of helices because of its robust stacking ability.[77] Hybridization is widely used in high-throughput genomics strategies.

We are grateful to the past and present workers in the Kool laboratory who have contributed to this research. Special recognition in this regard is deserved by Dr. Barbara Schweitzer, Dr. Sean Moran, Dr. Dongyu Liu, Dr. Tracy Matray, Dr. Rex Ren, Dr. Narayan Chaudhuri, and Robin Carroll. This work was supported by the National Institutes of Health (GM52956). JCM thanks the Spanish Ministry of Education for a postdoctoral fellowship.

Received: June 4, 1999 [A 348]

[1] J. D. Watson, F. H. Crick, *Nature* **1953**, *171*, 737–738.
[2] W. Saenger, *Principles of Nucleic Acid Structure*, Springer, New York, **1984**.
[3] N. C. Seeman, *Curr. Opin. Struct. Biol.* **1996**, *6*, 519–526.
[4] C. Mao, W. Sun, N. C. Seeman, *Nature* **1997**, *386*, 137–138.
[5] N. C. Seeman, *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 225–248.
[6] N. C. Seeman, *Angew. Chem.* **1998**, *110*, 3408–3428; *Angew. Chem. Int. Ed.* **1998**, *37*, 3220–3238.
[7] E. Winfree, F. Liu, L. A. Wenzler, N. C. Seeman, *Nature* **1998**, *394*, 539–544.
[8] C. Mao, W. Sun, Z. Shen, N. C. Seeman, *Nature* **1999**, *397*, 144–146.
[9] R. L. Letsinger, S. Chaturvedi, *Bioconjugate Chem.* **1998**, *9*, 826–830.
[10] C. A. Mirkin, R. L. Letsinger, R. C. Mucic, J. J. Storhoff, *Nature* **1996**, *382*, 607–609.
[11] A. P. Alivisatos, K. P. Johnsson, T. E. Wilson, C. J. Loweth, M. P. Bruchez, P. G. Schultz, *Nature* **1996**, *382*, 609–611.
[12] N. C. Chaudhiri, R. X.-F. Ren, E. T. Kool, *Synlett* **1997**, 341.
[13] B. A. Schweitzer, E. T. Kool, *J. Org. Chem.* **1994**, *59*, 7238–7242.
[14] F. H. Martin, M. M. Castro, F. Aboul-ela, I. Tinoco, Jr., *Nucleic Acid Res.* **1985**, *13*, 8927–8938.
[15] Y. Kawase, S. Iwai, H. Inoue, K. Miura, E. Ohtsuka, *Nucleic Acid Res.* **1986**, *14*, 7727–7736.
[16] B. L. Gaffney, L. A. Marky, R. A. Jones, *Tetrahedron* **1984**, *40*, 3–13.
[17] D. H. Turner, N. Sugimoto, R. Kierzek, S. D. Dreiker, *J. Am. Chem. Soc.* **1987**, *109*, 3783–3785.
[18] I. K. Yanson, A. B. Teplitsky, L. F. Sukhodub, *Biopolymers* **1979**, *18*, 1149–1170.
[19] A. R. Fersht, *Trends Biochem. Sci.* **1987**, *12*, 301–304.
[20] I. R. Gould, P. R. Kollman, *J. Am. Chem. Soc.* **1994**, *116*, 2493–2499.
[21] Y.-P. Pang, J. L. Miller, P. A. Kollman, *J. Am. Chem. Soc.* **1999**, *121*, 1717–1725.
[22] R. A. Friedman, B. Honig, *Biophys. J.* **1995**, *69*, 1528–1535.
[23] L. F. Newcomb, S. H. Gellman, *J. Am. Chem. Soc.* **1994**, *116*, 4993–4994.
[24] C. A. Hunter, *J. Mol. Biol.* **1993**, *230*, 1025–1054.
[25] D. B. Smithrud, T. B. Wyman, F. A. Diederich, *J. Am. Chem. Soc.* **1991**, *113*, 5420–5426.
[26] V. M. Rotello, E. A. Viani, G. Deslongchamps, B. A. Murray, J. Rebek, Jr., *J. Am. Chem. Soc.* **1993**, *115*, 797–798.
[27] E. T. Kool, *J. Am. Chem. Soc.* **1991**, *113*, 6265–6266.
[28] F. Aboul-ela, D. Koh, I. Tinoco, Jr., F. H. Martin, *Nucleic Acid Res.* **1985**, *13*, 4811–4824.
[29] H. Werntges, G. Steger, D. Riesner, H. J. Fritz, *Nucleic Acid Res.* **1986**, *14*, 3773–3791.
[30] M. Kouchakdjian, B. F. Li, P. F. Swann, D. J. Patel, *J. Mol. Biol.* **1988**, *202*, 139–155.
[31] W. N. Hunter, G. Kneale, T. Brown, D. Rabinovich, O. Kennard, *J. Mol. Biol.* **1986**, *190*, 605–618.
[32] H. T. Allawi, J. SantaLucia, Jr., *Nucleic Acid Res.* **1998**, *26*, 4925–4934.
[33] L. A. Loeb, T. A. Kunkel, *Annu. Rev. Biochem.* **1982**, *51*, 429–457.
[34] A. Kornberg, T. A. Baker, *DNA replication*, 2nd ed., Freeman, New York, **1992**.
[35] S. J. Benkovic, C. E. Cameron, *Methods Enzymol.* **1995**, *262*, 257–269.
[36] K. A. Johnson, *Annu. Rev. Biochem.* **1993**, *62*, 685–713.
[37] M. F. Goodman, S. Creighton, L. B. Bloom, J. Petruska, *Crit. Rev. Biochem. Mol. Biol.* **1993**, *28*, 83–126.
[38] H. Echols, M. F. Goodman, *Annu. Rev. Biochem.* **1991**, *60*, 477–511.
[39] C. M. Joyce, T. A. Steitz, *Annu. Rev. Biochem.* **1994**, *63*, 777–822.
[40] C. M. Joyce, T. A. Steitz, *J. Bacteriol.* **1995**, *177*, 6321–6329.
[41] L. Stryer, *Biochemistry*, Freeman, New York, **1995**.
[42] J. D. Watson, M. H. Hopkins, J. W. Roberts, J. A. Steitz, A. M. Weiner, *Molecular Biology of the Gene*, Benjamin/Cummings, Menlo Park, CA, **1987**.
[43] K. M. Guckian, E. T. Kool, *Angew. Chem.* **1997**, *109*, 2942–2945; *Angew. Chem. Int. Ed. Engl.* **1997**, *36*, 2825–2828.
[44] K. M. Guckian, J. C. Morales, E. T. Kool, *J. Org. Chem.* **1998**, *63*, 9652–9656.
[45] A. N. Chekhlov, *J. Struct. Chem.* **1995**, *36*, 155–161.
[46] D. W. Young, P. Tollin, H. R. Wilson, *Acta Crystallogr. Sect. B* **1969**, *25*, 1423.
[47] L. Shimoni, J. P. Glusker, *Struct. Chem.* **1994**, *5*, 383.
[48] J. A. K. Howard, V. J. Hoy, D. O'Hagan, G. T. Smith, *Tetrahedron* **1996**, *52*, 12613–12622.
[49] V. R. Thalladi, H.-C. Weiss, D. Bläser, R. Boese, A. Nangia, G. R. Desiraju, *J. Am. Chem. Soc.* **1998**, *120*, 8702–8710.
[50] H. Plenio, R. Diodone, *Chem. Ber.* **1997**, *130*, 633–640.
[51] T. A. Evans, K. R. Seddon, *Chem. Commun.* **1997**, 2023–2024.
[52] M. Meyer, J. Suhnel, *J. Biomol. Struct. Dyn.* **1997**, *15*, 619–624.
[53] X. Wang, K. N. Houk, *Chem. Commun.* **1998**, 2631–2632.
[54] D. Barsky, E. T. Kool, M. Colvin, *J. Biomol. Struct. Dyn.* **1999**, *16*, 1119–1134.
[55] B. A. Schweitzer, E. T. Kool, *J. Am. Chem. Soc.* **1995**, *117*, 1863–1872.
[56] K. M. Guckian, T. R. Krugh, E. T. Kool, *Nat. Struct. Biol.* **1998**, *5*, 954–959.
[57] R. Eritja, D. M. Horowitz, P. A. Walker, J. P. Ziehler-Martin, M. S. Boosalis, M. F. Goodman, K. Itakura, B. E. Kaplan, *Nucleic Acid Res.* **1986**, *14*, 8135–8153.
[58] B. A. Connolly, P. C. Newman, *Nucleic Acid Res.* **1989**, *17*, 4957–4973.
[59] S. Hildbrand, A. Blaser, S. P. Parel, C. J. Leumann, *J. Am. Chem. Soc.* **1997**, *119*, 5499–5511.

[60] B. Gildea, L. W. McLaughlin, *Nucleic Acid Res.* **1989**, *17*, 2261 – 2280.
[61] S. M. Law, R. Eritja, M. F. Goodman, K. J. Breslauer, *Biochemistry* **1996**, *35*, 12329 – 12337.
[62] P. K. T. Lin, D. M. Brown, *Nucleic Acid Res.* **1989**, *17*, 10373 – 10383.
[63] B. F. L. Li, C. B. Beese, P. F. Swann, *Biochemistry* **1987**, *26*, 1086 – 1093.
[64] B. L. Gaffney, R. A. Jones, *Biochemistry* **1989**, *28*, 5881 – 5889.
[65] H. Nishio, A. Ono, A. Matsuda, T. Ueda, *Nucleic Acid Res.* **1992**, *20*, 777 – 782.
[66] T. A. Millican, G. A. Mock, M. A. Chauncey, T. P. Patel, M. A. Eaton, J. Gunning, S. D. Cutbush, S. Neidle, J. Mann, *Nucleic Acid Res.* **1984**, *12*, 7435 – 7453.
[67] D. Loakes, F. Hill, D. M. Brown, S. A. Salisbury, *J. Mol. Biol.* **1997**, *270*, 426 – 435.
[68] R. Nichols, P. A. Andrews, P. Zhang, B. E. Bergstrom, *Nature* **1994**, *369*, 492 – 493.
[69] A. Chollet, E. Kawashima, *Nucleic Acid Res.* **1988**, *16*, 305 – 317.
[70] J. A. Piccirilli, T. Krauch, S. E. Moroney, S. A. Benner, *Nature* **1990**, *343*, 33 – 37.
[71] C. Roberts, R. Bandaru, C. Switzer, *J. Am. Chem. Soc.* **1997**, *119*, 4640 – 4649.
[72] T. J. Matray, E. T. Kool, *J. Am. Chem. Soc.* **1998**, *120*, 6191 – 6192.
[73] V. A. Buckin, *Biophys. Chem.* **1988**, *29*, 283 – 292.
[74] J. Gao, *Biophys. Chem.* **1994**, *51*, 253 – 261.
[75] S. H. Gellman, T. S. Haque, L. F. Newcomb, *Biophys. J.* **1996**, *71*, 3523 – 3525.
[76] R. A. Friedman, B. Honig, *Biophys. J.* **1996**, *71*, 3525 – 3526.
[77] K. M. Guckian, B. A. Schweitzer, R. X.-F. Ren, C. J. Sheils, P. A. Paris, D. C. Tahmassebi, E. T. Kool, *J. Am. Chem. Soc.* **1996**, *118*, 8182 – 8183.
[78] D. H. Turner, N. Sugimoto, *Annu. Rev. Biophys. Biophys. Chem.* **1988**, *17*, 167 – 192.
[79] M. Senior, R. A. Jones, K. J. Breslauer, *Biochemistry* **1988**, *27*, 3879 – 3885.
[80] N. Sugimoto, A. Matsumura, K. Hasegawa, M. Sasaki, *Nucleic Acids Symp. Ser.* **1991**, *25*, 51 – 52.
[81] M. H. Werner, A. M. Gronenborn, G. M. Clore, *Science* **1996**, *271*, 778 – 784.
[82] C. A. Bewley, A. M. Gronenborn, G. M. Clore, *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 105 – 131.
[83] F. M. Jucker, A. Pardi, *Biochemistry* **1995**, *34*, 14416 – 14427.
[84] X.-F. Ren, B. A. Schweitzer, C. J. Sheils, E. T. Kool, *Angew. Chem.* **1996**, *108*, 834 – 837; *Angew. Chem. Int. Ed. Engl.* **1996**, *35*, 743 – 746.
[85] G. Felsenfeld, H. T. Miles, *Annu. Rev. Biochem.* **1967**, *36*, 407 – 448.
[86] D. Pörschke, *Eur. J. Biochem.* **1973**, *39*, 117 – 126.
[87] C. S. Olsthoorn, L. J. Bostelaar, J. F. De Rooij, J. H. Van Boom, C. Altona, *Eur. J. Biochem.* **1981**, *115*, 309 – 321.
[88] T. E. Spratt, D. E. Levy, *Nucleic Acid Res.* **1997**, *25*, 3354 – 3361.
[89] P. Stazewski, C. Tamm, *Angew. Chem.* **1990**, *102*, 37 – 59; *Angew. Chem. Int. Ed. Engl.* **1990**, *29*, 36 – 57.
[90] C. Switzer, S. E. Moroney, S. A. Benner, *J. Am. Chem. Soc.* **1989**, *111*, 8322 – 8323.
[91] J. Horlacher, M. Hottiger, V. N. Podust, U. Hubscher, S. A. Benner, *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 6329 – 6333.
[92] S. Shibutani, M. Takeshita, A. P. Grollman, *J. Biol. Chem.* **1997**, *272*, 13916 – 13922.
[93] M. Takeshita, C. N. Chang, F. Johnson, S. Will, A. P. Grollman, *J. Biol. Chem.* **1987**, *262*, 10171 – 10179.
[94] S. K. Randall, R. Eritja, B. E. Kaplan, J. Petruska, M. F. Goodman, *J. Biol. Chem.* **1987**, *262*, 6864 – 6870.
[95] C. W. Shearman, L. A. Loeb, *J. Mol. Biol.* **1979**, *128*, 197 – 218.
[96] R. M. Schaaper, B. W. Glickman, L. A. Loeb, *Cancer Res.* **1982**, *42*, 3480 – 3485.
[97] L. A. Loeb, B. D. Preston, *Annu. Rev. Genet.* **1986**, *20*, 201 – 230.
[98] B. Strauss, S. Rabkin, D. Sagher, P. Moore, *Biochimie* **1982**, *64*, 829 – 838.
[99] B. S. Strauss, *Bioessays* **1991**, *13*, 79 – 84.
[100] D. Sagher, B. Strauss, *Biochemistry* **1983**, *22*, 4518 – 4526.
[101] R. M. Schaaper, T. A. Kunkel, L. A. Loeb, *Proc. Natl. Acad. Sci. USA* **1983**, *80*, 487 – 491.
[102] C. W. Lawrence, A. Borden, S. K. Banerjee, J. E. LeClerc, *Nucleic Acid Res.* **1990**, *18*, 2153 – 2157.
[103] T. Paz-Elizur, M. Takeshita, Z. Livneh, *Biochemistry* **1997**, *36*, 1766 – 1773.
[104] R. D. Beger, P. H. Bolton, *J. Biol. Chem.* **1998**, *273*, 15565 – 15573.
[105] P. Cuniasse, L. C. Sowers, R. Eritja, B. Kaplan, M. F. Goodman, J. A. Cognet, M. LeBret, W. Guschlbauer, G. V. Fazakerley, *Nucleic Acid Res.* **1987**, *15*, 8003 – 8022.
[106] Z. Lin, K. N. Hung, A. P. Grollman, C. de los Santos, *Nucleic Acid Res.* **1998**, *26*, 2385 – 2391.
[107] P. Cuniasse, G. V. Fazakerley, W. Guschlbauer, B. E. Kaplan, L. C. Sowers, *J. Mol. Biol.* **1990**, *213*, 303 – 314.
[108] I. Wong, S. S. Patel, K. A. Johnson, *Biochemistry* **1991**, *30*, 526 – 537.
[109] D. L. Sloane, M. F. Goodman, H. Echols, *Nucleic Acid Res.* **1988**, *16*, 6465 – 6475.
[110] J. Petruska, M. F. Goodman, M. S. Boosalis, L. C. Sowers, C. Cheong, I. Tinoco, Jr., *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 6252 – 6256.
[111] R. D. Kuchta, V. Mizrahi, P. A. Benkovic, K. A. Johnson, S. J. Benkovic, *Biochemistry* **1987**, *26*, 8410 – 8417.
[112] R. D. Kuchta, P. Benkovic, S. J. Benkovic, *Biochemistry* **1988**, *27*, 6716 – 6725.
[113] K. Bebenek, C. M. Joyce, M. P. Fitzgerald, T. A. Kunkel, *J. Biol. Chem.* **1990**, *265*, 13878 – 13887.
[114] S. Moran, R. X.-F. Ren, S. Rumney IV, E. T. Kool, *J. Am. Chem. Soc.* **1997**, *119*, 2056 – 2057.
[115] S. Moran, R. X.-F. Ren, E. T. Kool, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 10506 – 10511.
[116] D. Liu, S. Moran, E. T. Kool, *Chem. Biol.* **1997**, *4*, 919 – 926.
[117] J. C. Morales, E. T. Kool, *Nat. Struct. Biol.* **1998**, *5*, 950 – 954.
[118] E. T. Kool, *Biopolymers* **1998**, *48*, 3 – 17.
[119] S. Langouet, M. Muller, F. P. Guengerich, *Biochemistry* **1997**, *36*, 6069 – 6079.
[120] V. Litinski, A. Chenna, J. Sagi, B. Singer, *Carcinogenesis* **1997**, *18*, 1609 – 1615.
[121] J. P. Seiler, *Mutat. Res.* **1973**, *17*, 21 – 25.
[122] U. Diederichsen, *Angew. Chem.* **1998**, *110*, 1745 – 1747; *Angew. Chem. Int. Ed.* **1998**, *37*, 1655 – 1657.
[123] H. Pelletier, M. R. Sawaya, A. Kumar, S. H. Wilson, J. Kraut, *Science* **1994**, *264*, 1891 – 1903.
[124] T. A. Kunkel, S. H. Wilson, *Nat. Struct. Biol.* **1998**, *5*, 95 – 99.
[125] W. A. Beard, S. H. Wilson, *Chem. Biol.* **1998**, *5*, R7 – R13.
[126] L. S. Beese, V. Derbyshire, T. A. Steitz, *Science* **1993**, *260*, 352 – 355.
[127] Y. Kim, S. H. Eom, J. Wang, D. S. Lee, S. W. Suh, T. A. Steitz, *Nature* **1995**, *376*, 612 – 616.
[128] S. Doublie, S. Tabor, A. M. Long, C. C. Richardson, T. Ellenberger, *Nature* **1998**, *391*, 251 – 258.
[129] J. R. Kiefer, C. Mao, J. C. Braman, L. S. Beese, *Nature* **1998**, *391*, 304 – 307.
[130] E. Arnold, J. Ding, S. H. Hughes, Z. Hostomsky, *Curr. Opin. Struct. Biol.* **1995**, *5*, 27 – 38.
[131] H. Huang, R. Chopra, G. L. Verdine, S. C. Harrison, *Science* **1998**, *282*, 1669 – 1675.
[132] B. T. Eger, R. D. Kuchta, S. S. Carroll, P. A. Benkovic, M. E. Dahlberg, C. M. Joyce, S. J. Benkovic, *Biochemistry* **1991**, *30*, 1441 – 1448.
[133] J. A. Hirsch, *Table of conformational energies*, *Vol. 1*, Interscience, New York, **1967**.
[134] T. J. Matray, E. T. Kool, *Nature* **1999**, *399*, 704 – 708.
[135] H. Cai, L. B. Bloom, R. Eritja, M. F. Goodman, *J. Biol. Chem.* **1993**, *268*, 23567 – 23572.
[136] S. H. Eom, J. Wang, T. A. Steitz, *Nature* **1996**, *382*, 278 – 281.
[137] K. Bebenek, W. A. Beard, T. A. Darden, L. Li, R. Prasad, B. A. Luton, D. G. Gorenstein, S. H. Wilson, T. A. Kunkel, *Nat. Struct. Biol.* **1997**, *4*, 194 – 197.
[138] J. C. Morales, E. T. Kool, *J. Am. Chem. Soc.* **1999**, *121*, 2323 – 2324.
[139] M. F. Goodman, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 10493 – 10495.
[140] S. Brenowitz, S. Kwack, M. F. Goodman, M. O'Donnell, H. Echols, *J. Biol. Chem.* **1991**, *266*, 7888 – 7892.
[141] S. S. Parikh, C. D. Mol, D. J. Hosfield, J. A. Tainer, *Curr. Opin. Struct. Biol.* **1999**, *9*, 37 – 47.
[142] T. A. Steitz, *Curr. Opin. Struct. Biol.* **1993**, *3*, 31 – 38.
[143] T. A. Steitz, S. J. Smerdon, J. Jager, C. M. Joyce, *Science* **1994**, *266*, 2022 – 2025.