Chemistry & Biology, Vol. 10, 815–825, September, 2003, ©2003 Elsevier Science Ltd. All rights reserved. DOI 10.1016/j.chembiol.2003.08.008

# DNA Polymerase Template Interactions Probed by Degenerate Isosteric Nucleobase Analogs

**Natasha Paul,[1] Vishal C. Nashine,[1] Geoffrey Hoops,[1] Peiming Zhang,[2] Jie Zhou, Donald E. Bergstrom,[1,2] and V. Jo Davisson[1,*]**

[1]Department of Medicinal Chemistry and Molecular Pharmacology
Purdue University
West Lafayette, Indiana 47907

[2]Walther Cancer Institute
Indianapolis, Indiana 46208

## Summary

**The development of novel artificial nucleobases and detailed X-ray crystal structures for primer/template/DNA polymerase complexes provide opportunities to assess DNA-protein interactions that dictate specificity. Recent results have shown that base pair shape recognition in the context of DNA polymerase must be considered a significant component. The isosteric azole carboxamide nucleobases (compounds 1–5; Figure 1) differ only in the number and placement of nitrogen atoms within a common shape and therefore present unique electronic distributions that are shown to dictate the selectivity of template-directed nucleotide incorporation by DNA polymerases. The results demonstrate how nucleoside triphosphate substrate selection by DNA polymerase is a complex phenomenon involving electrostatic interactions in addition to hydrogen bonding and shape recognition. These azole nucleobase analogs offer unique molecular tools for probing nonbonded interactions dictating substrate selection and fidelity of DNA polymerases.**

## Introduction

A property of major importance to the fidelity of the DNA replication process involves the substrate specificity of the DNA polymerase. The bonding and nonbonding interactions that provide a mechanism for recognition of the correct base pairs by DNA polymerases are features that have been probed using a variety of artificial nucleobases [1]. In addition, the recent analyses of X-ray crystal structures have inspired several hypotheses regarding the molecular basis for DNA polymerase proofreading and fidelity. X-ray crystal structures of several Pol I DNA polymerases including *Thermus aquaticus* (*Taq*) [2, 3], *Bacillus stearothermophilus* [4], and T7 [5] DNA polymerases complexed with DNA show unifying features [6, 7]. The protein induces an unwinding of the DNA helix more characteristic of A form over the first four to five base pairs of the primer-template duplex. The resulting shallow minor groove is accessible for direct interactions to specific protein side chains in the finger domain that alternately share (donate) hydrogen bonds with purine N3 and pyrimidine O2 positions on both primer and template strands. What emerges is an overall hypothesis that misincorporation leads to improper orientation of the purine N3 and pyrimidine O2 atoms. The reduced affinity of the protein-DNA complex could result in a kinetic "stalling" of the extension reaction. In the case for enzymes containing an exonuclease domain, a "switching" mechanism is then enabled that translocates the primer strand to the 3′ to 5′ exonuclease active site for excision of the incorrect base pairs.

While these minor groove recognition features elucidate important aspects of the proofreading mechanisms of the Pol I DNA polymerases, there are still unanswered questions regarding substrate selection. *Taq* DNA polymerase (exo−) is an example of a highly processive enzyme of the Pol I polymerase family that has an error rate on the average of 1 in every $10^5$ incorporation events [8]. The error frequency is not impressive on a biological time scale, but remarkable in the context of substrate selectivity in the extension activity. Several point mutants of the *Taq* DNA polymerase have identified potential roles of specific amino acids on the relative incorporation rates of dNTP and ddNTPs [9–12]. However, decoupling the fundamental nonbonding interactions between the DNA template and dNTP substrates that control the selectivity of the enzyme and the related members of the Pol I family remains a challenge.

The interactions within the DNA polymerase active site that relate structural features of the template-nucleoside to the triphosphate substrate are of central importance to understanding function. Isosteric base analogs with reconfigured hydrogen bonding patterns or alternate tautomeric forms have been designed and synthesized to assess hydrogen bond base pairing rules [13–20]. These efforts have provided clear examples of purine-pyrimidine base recognition mimics that can be utilized by DNA polymerases. In addition, shape recognition of base pairs by DNA polymerase has been proposed as a factor in substrate selection. The nonpolar nucleobase difluorotoluene, a thymine isostere lacking hydrogen bonding functionality, can effectively substitute for thymine in both the template strand and as the incoming nucleoside triphosphate, indicating the importance of shape in the absence of hydrogen bonding [21–25]. Similarly, 4-methylbenzimidazole is an effective surrogate for adenine when matched opposite the difluorotoluene nucleobase [26]. Other unnatural hydrophobic base pairs have been more recently described [27–29]. In addition, hydrophobic bases that have the capacity to form self-pairs in DNA polymerase extension reactions have been developed [30–32].

Our interest in assessing the type of nonbonding interactions that mediate DNA polymerase specificity has been defined by the use of degenerate recognition to probe the influence of template DNA bases on *Taq* DNA polymerase substrate selection. We previously proposed that isosteric azole carboxamide nucleobases would provide a platform for evaluating the effects of structure on base pair recognition by DNA polymerase

*Correspondence: vjd@pharmacy.purdue.edu

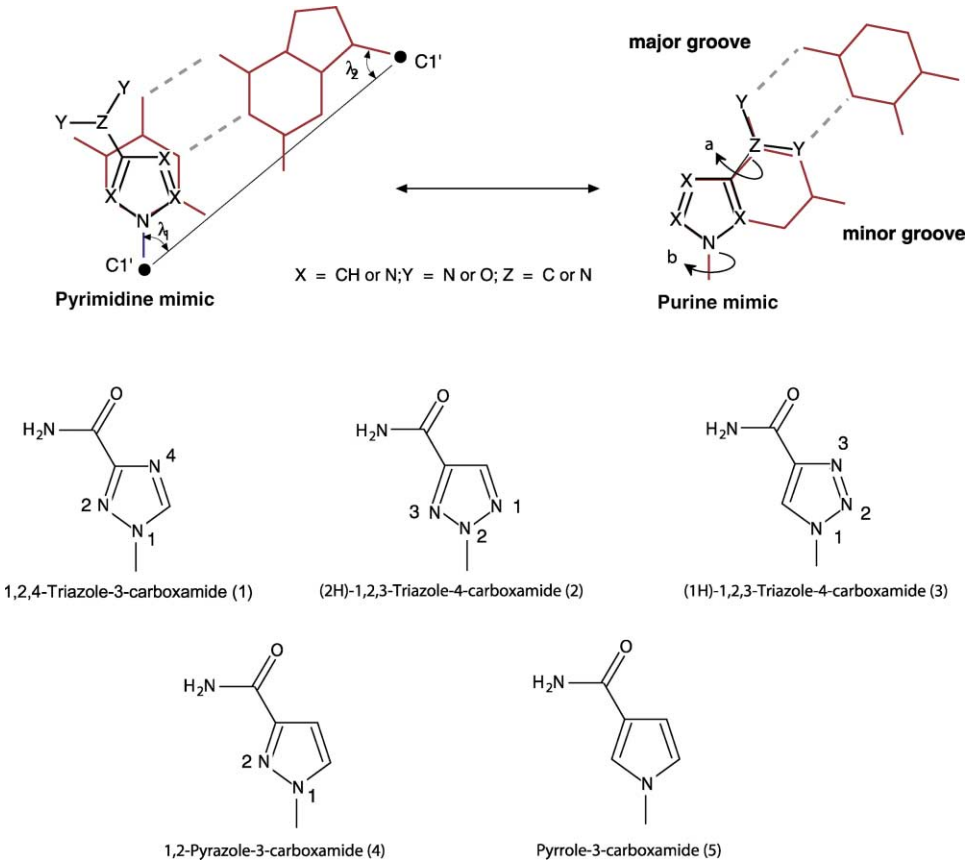



Figure 1. Model of the Azole Heterocycles Carboxamides as Nucleobase Pair Mimics
The key structural parameters in the context of duplex DNA are highlighted.

[33]. These template recognition studies of the azole nucleobases by *Taq* DNA polymerase indicate that other factors, in addition to shape, can influence the selectivity of base incorporation. Here we describe a focused analysis of the impact of template azole nucleobase ring electronics on the substrate incorporation step and offer a refined model of nonbonding interactions in the *Taq* DNA polymerase extension site.

## Results and Discussion

### Azole Base Design

The azole carboxamide class of nucleobases offer a high degree of variation in electronic density and hydrogen bonding donor and acceptor sites within a single framework. Members of this molecular set can mimic more than one natural purine or pyrimidine nucleobase by virtue of rotation about bonds **a** and **b** (Figure 1). Modeling studies with azole carboxamides paired with the four natural bases in a B-DNA duplex show that two important parameters of base-pair geometry are maintained. To be isosteric with A:T or G:C base pairs, the C1′ to C1′ distance must be in the range of 10.8–11.0 Å, and $\lambda_1$ & $\lambda_2$ around 50° [34, 35]. Models of azole carboxamide bases paired with a natural base have predicted C1′-C1′ distances that fall within 0.2 Å of the natural base pairs. In addition, the $\lambda_1$ & $\lambda_2$ angles fall within 3° of the geometry of Watson-Crick pairs [36–38].

The electrostatic surface potentials were calculated (AM1) and mapped for the azole carboxamides containing from one to three nitrogen atoms in the heterocycle (Figure 2). Only those conformations that are of sufficiently low potential energy are included in the figure. For comparison, the electrostatic surface potentials for the four natural DNA bases are shown at the bottom of the Figure 2. Inspection of the figure reveals that the family includes a broad range of electronic surface profiles within a common steric framework. Since the size and shape of the azole carboxamide framework closely matches the natural bases, this family of analogs constitutes a particularly useful set of molecules for probing electrostatic interactions within protein sites that bind nucleic acid bases.

### Azole Nucleobase Effects on DNA Duplex Stability

Comparison of the thermodynamic stability of deoxyoligonucleotides containing nucleosides **1–5** reveals a consistent pattern of base pairing preferences that can be traced to the nitrogen atom placement in the azole ring. There are significant differences in the base pairing preferences of the five azole nucleobases, which can be attributed to the β ring nitrogen (Table 1). When a β

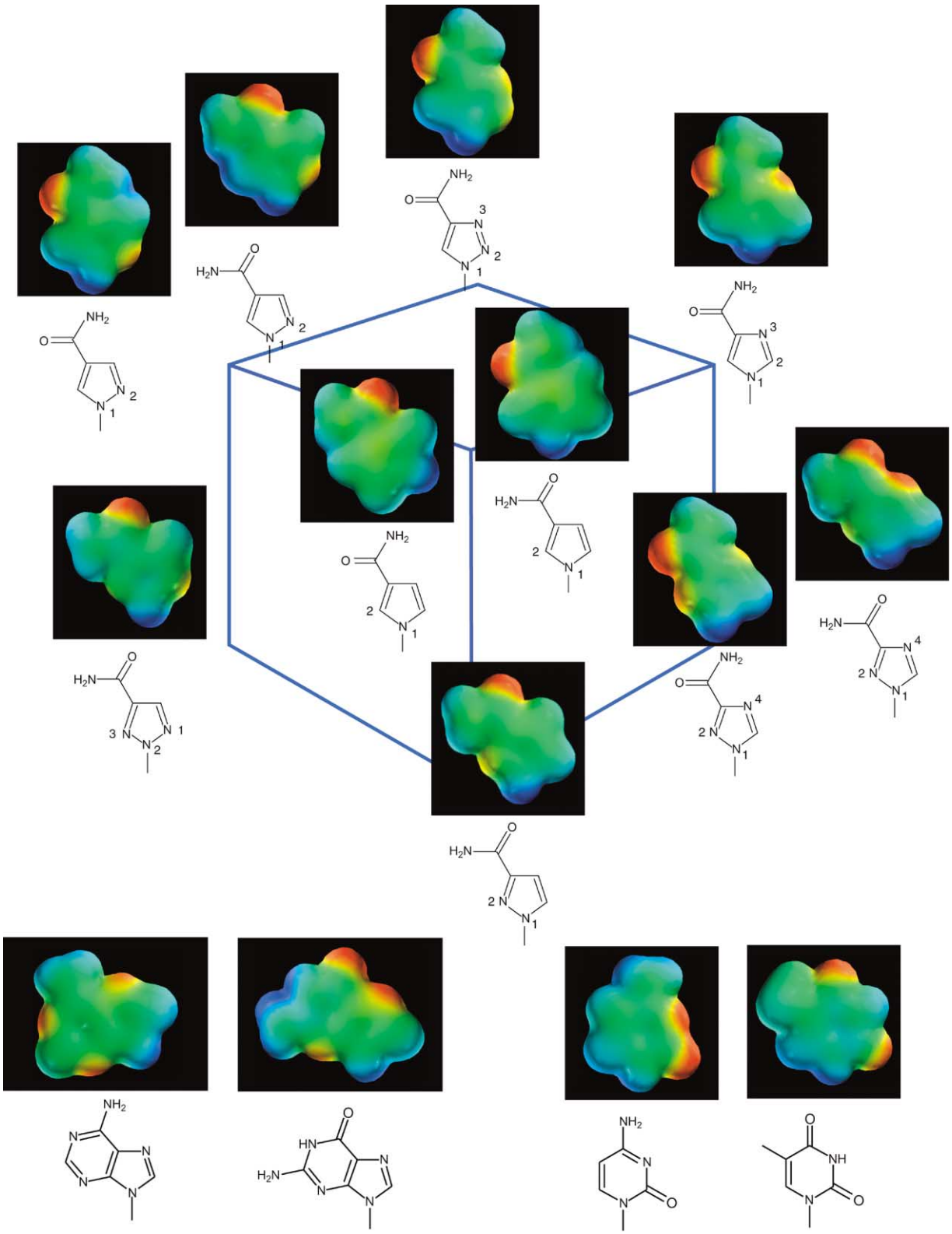


Figure 2. Electronic Profiles for All Azole Carboxamide Nucleobases in the Preferred Amide Conformations

Red designates the highest electronegative regions, green is neutral, and blue is net electropositive. The electrostatic contour diagrams shown in this figure were generated in the program MacSpartan Plus (Wavefunction, Inc., Irvine, CA), version 2.0. For display purposes the electron density potential was set for the range −64.000 (red) to 54.000 electrons/au3 (blue), and the isovalue was set at 0.002 (green). The partial atomic charges were calculated in MacSpartan Plus using the AM1 model [58].

Table 1. Thermal Melting Studies with Azole Base-Containing Oligonucleotides

| Azole Base | dA | dC | dG | dT |
|---|---|---|---|---|
| 1 | 29 | 21 | 40 | 40 |
| 2 | 27 | 28 | 28 | 28 |
| 3 | 31 | 14 | 51 | 39 |
| 4 | 20 | 25 | 24 | 31 |
| 5 | 19 | 15 | 25 | 34 |

$T_m$ values (°C) for the oligonucleotide d(CGCAzAATTNGCG), where Az = azole nucleobase **1**, **2**, **3**, **4**, or **5**, and N = A, C, G, or T. The $T_m$ values for sequences containing **5** have already been published elsewhere [36].

ring nitrogen is absent (nucleosides **2**, **4**, and **5**), the $T_m$ values opposite each of the four bases fall into a relatively narrow range [36]. Notably, nucleoside **2** showed no discrimination; the $T_m$ values of the four different 12-mers containing nucleoside **2** opposite each of the natural bases were virtually identical (28°C ± 1°C). In contrast, nucleoside **3**, which contains a β ring nitrogen, showed high discrimination. The duplexes with **3** opposite dG ($T_m$ = 51°C) and T ($T_m$ = 39°C) are stable when compared to the pair with dC ($T_m$ = 14°C). These results are consistent with a preferred conformation where the carboxamide NH is syn to the β ring nitrogen [38]. Nucleoside **1** (1,2,4-triazole-3-carboxamide) and imidazole-4-carboxamide show trends similar to **3**, although the $T_m$ values for the deoxyoligonucleotides with the nucleoside paired opposite dG and T are lower [37]. The base-pairing preferences for the azole carboxamide nucleobases reveal that the β ring nitrogen is a major determinant for duplex stability and base pairing discrimination.

### Triazole Nucleobase Template Properties with *Taq* DNA Polymerase

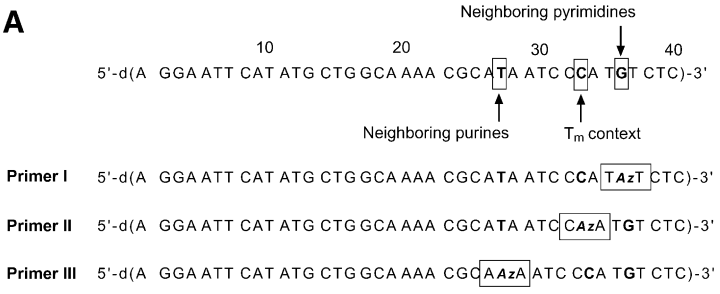
*Taq* DNA polymerase recognition of the azole carboxamide nucleobases can be assessed by measuring the relative incorporations of natural bases across from a DNA template containing an azole carboxamide at a specified position. In order to assess the rates of incorporation of natural bases opposite the azole nucleobases in the template, we performed steady-state kinetic studies using the gel-based Goodman assay [39]. The assay involved extension of the Cy5-labeled primer (23-mer) hybridized with the template strand (28-mer) containing the appropriate azole nucleobase (**1–5**). These single nucleotide incorporation studies revealed an interesting spectrum of selectivities toward the natural nucleotides. For natural purine/pyrimidine nucleotide substrates and templates, a 1000-fold distinction between matched and mismatched pairs is observed in $V_{max}/K_m$ values. When the azole nucleobases are in the template, the distinctions between $V_{max}/K_m$ values for each of the bases is relaxed, ranging from 1- to 50-fold (Table 2). In all the compounds, **1–5**, the rate of incorporation showed a different level of degeneracy toward the natural bases. For example, nucleobase **1** is a good purine mimic and shows a 25-fold higher selectivity for dC and T in comparison to dA and dG, which are similar to the mismatched natural base pair formation. In general, we observe similar discrimination (as reflected in the differences of $V_{max}/K_m$ values) toward the incorporation of natural bases opposite compounds **1–5** despite the greater hydrophobicity of compounds **4** and **5**.

Table 2. Steady-State Incorporation of Natural Nucleobases Opposite the Azole Carboxamide Nucleobases in the Template Strand

| Template Base | dNTP | $K_m$ | $V_{max}$ | $V_{max}/K_m$ | % PCR |
|---|---|---|---|---|---|
| dG | dATP | 261 (62) | 0.016 (0.002) | 0.06 | ND |
| | dCTP | 3 (0) | 0.456 (0.035) | 152.07 | ND |
| | dGTP | 129 (23) | 0.018 (0.001) | 0.14 | ND |
| | dTTP | 367 (63) | 0.070 (0.004) | 0.19 | ND |
| dA | dTTP | 5 (2) | 0.781 (0.096) | 156.14 | ND |
| **1** | dATP | NA | NA | NA | 1 |
| | dCTP | 156 (40) | 0.421 (0.035) | 2.70 | 47 |
| | dGTP | 15 (2) | 0.002 (0.000) | 0.15 | 0 |
| | dTTP | 194 (35) | 0.625 (0.043) | 3.22 | 52 |
| **2** | dATP | 180 (20) | 0.473 (0.020) | 2.63 | 40 |
| | dCTP | 191 (43) | 0.450 (0.038) | 2.36 | 39 |
| | dGTP | 55 (34) | 0.002 (0.000) | 0.04 | 0 |
| | dTTP | 247 (51) | 0.421 (0.035) | 1.70 | 20 |
| **3** | dATP | 162 (46) | 0.386 (0.035) | 2.38 | 11 |
| | dCTP | NA | NA | NA | 0 |
| | dGTP | 22 (6) | 0.518 (0.018) | 23.52 | 84 |
| | dTTP | 16 (5) | 0.008 (0.000) | 0.51 | 6 |
| **4** | dATP | 222 (51) | 0.156 (0.018) | 0.70 | 2 |
| | dCTP | 38 (8) | 0.569 (0.024) | 14.98 | 51 |
| | dGTP | 175 (58) | 0.092 (0.011) | 0.53 | 1 |
| | dTTP | 81 (12) | 0.773 (0.032) | 9.54 | 45 |
| **5** | dATP | 219 (78) | 0.377 (0.053) | 1.72 | 33 |
| | dCTP | 83 (34) | 0.022 (0.001) | 0.26 | 0 |
| | dGTP | 74 (42) | 0.012 (0.001) | 0.17 | 2 |
| | dTTP | 86 (12) | 0.719 (0.026) | 8.36 | 65 |

Here, $K_m$ is expressed in μM, $V_{max}$ is in μmol/min·mg, and $V_{max}/K_m$ is in μmol/min·mg·μM × $10^{-3}$. NA indicates that the data for these particular base pairs could not be obtained because the turnover was below the detection limit of the protocol used for estimation. For comparison, the percent incorporation values for the PCR assay (from Figure 3) are included for clarity. ND indicates values that were not determined using the PCR-based assay. Values in parentheses indicate the standard error. The protein concentrations in the enzyme assay were determined using Sypro Ruby staining in SDS-PAGE. The results for the PCR assay for **4** and **5** have already been published elsewhere [33].

**A**




**B**

| Nucleotide incorporated | Primer-Azole | I-1 | I-2 | I-3 | II-3 | III-3 |
|---|---|---|---|---|---|---|
| | dA | 1(1) | 40 (10) | 11 (4) | 8 (2) | 10 (1) |
| | dC | 47 (8) | 39 (7) | 0 (0) | 0 (0) | 0 (0) |
| | dG | 0 (0) | (0) | 84 (4) | 85 (7) | 88 (1) |
| | dT | 52 (7) | 20 (3) | 4 (3) | 7 (6) | 1 (0) |

Figure 3. Incorporation of Natural Substrate Opposite the Azole Nucleobase in PCR Assay

(A) Azole nucleobase-containing PCR primers and their sequence context used to amplify the *HisF* gene. For **3**, primers II and III were also synthesized and used in the same assay.

(B) The percent natural substrate incorporation using templates that contain the azole nucleobases. Azole carboxamide nucleotides **1–3** were incorporated into the *HisF* gene by PCR using primers I–III. PCR, dideoxy cycle sequencing reactions, and PAGE were performed for each azole carboxamide-containing oligonucleotide. The percentage incorporation of each deoxyribonucleotide by *Taq* DNA polymerase opposite the sites containing the azole carboxamide nucleotide analogs represents the average of between four and eight runs, with the standard deviation in parentheses.

In order to probe the recognition of the azole nucleobases in larger DNA templates, we prepared the primers shown in Figure 3 for a PCR-based assay [33]. These primers are complementary to the 5′-end of the *hisF* gene and allowed direct comparisons to previous results. PCR amplification of the gene using the sense primers I–III and an antisense primer yielded full-length gene products in which different amounts of natural nucleotides are incorporated opposite positions 35 (primer I), 33 (primer II), and 27 (primer III). DNA sequence analysis and base quantification at the specified position allowed for assessment of the template properties of the azole nucleobases. Each of the three azoles nucleobases **1–3** provided distinct distributions of purine and pyrimidine base incorporations. Most strikingly, the results from the PCR incorporation of the natural bases opposite the azole nucleobases correlates well with the $V_{max}/K_m$ values obtained in the kinetic studies (Table 2).

One of the bases, 1,2,3-triazole-4-carboxamide **3**, was examined in three different sequence contexts to assess if neighboring base effects in the template dictate the distribution of base incorporations and, hence, the substrate preferences. The results of these incorporation studies are also summarized in Figure 3. Only small differences were observed in comparing primer sequence context for azole carboxamide **3** at positions 35, 33, and 27 (primers I–III). At least in these template sequences, the position of azole carboxamide does not play a significant role in recognition of the azole-substrate nucleobase pairs during the incorporation step by *Taq* DNA polymerase.

A question that arises with azole carboxamide nucleobases is whether they resemble matched or mismatched base pairs. While the overall effects of the azole carboxamides are destabilizing, there are defined cases of preferred pairings (Table 1). If base-pairing preferences are contributing factors in dictating *Taq* DNA polymerase specificity, natural base insertion opposite a template azole nucleotide should be reflected in the thermodynamic preferences. However, triazole **1**, which pairs well with dG and T in the $T_m$ experiments, fails to incorporate any dG (Tables 1 and 2). Moreover, *Taq* DNA polymerase directs the insertion of dC opposite triazole **1** despite the low stability of this base pair. Triazole **2** shows no preference in base pairing, but *Taq* DNA polymerase directs the incorporation of dA, dC, and T but not dG. In contrast, the base pair between triazole **3** and dG is particularly stable when compared with dA, dC, or T, and *Taq* DNA polymerase inserts dG opposite **3** at a higher rate than the other three nucleotides. Finally, the stability of the base pairs of T with triazole **1** and **3** are equivalent; however, *Taq* DNA polymerase does not effectively incorporate T when paired with **3**.

These results indicate that there is a significant contribution to *Taq* DNA polymerase-directed base insertion stemming from factors other than shape and hydrogen bonding between complementary bases. Inspection of possible base pairs between the triazoles and natural nucleic acid bases (Figure 4) reveals a striking correspondence between the relative position of the α $sp^2$ hybridized nitrogen (N2 in triazoles **1** and **3**, and N1 and N3 in triazole **2**) and the selected substrate nucleotides. The model in which azole carboxamide-natural base pairs most closely mimic purine-pyrimidine base pairs precludes any direct involvement of the α nitrogen in hydrogen bonding to the incoming nucleotide substrate. When the α nitrogen is on the same side of the azole as the carboxamide group (Figure 4A), the base is able to function as a universal purine. In this case, the bond between the amide carbonyl carbon and azole heterocycle is assumed to rotate 180° so that the amide-$NH_2$ mimics either the exocyclic amino group of adenine or the N(1)-H of guanine (Figure 4A). When the α nitrogen is on the opposite side (from the carboxamide substituent) of the heterocycle and the heterocycle contains a

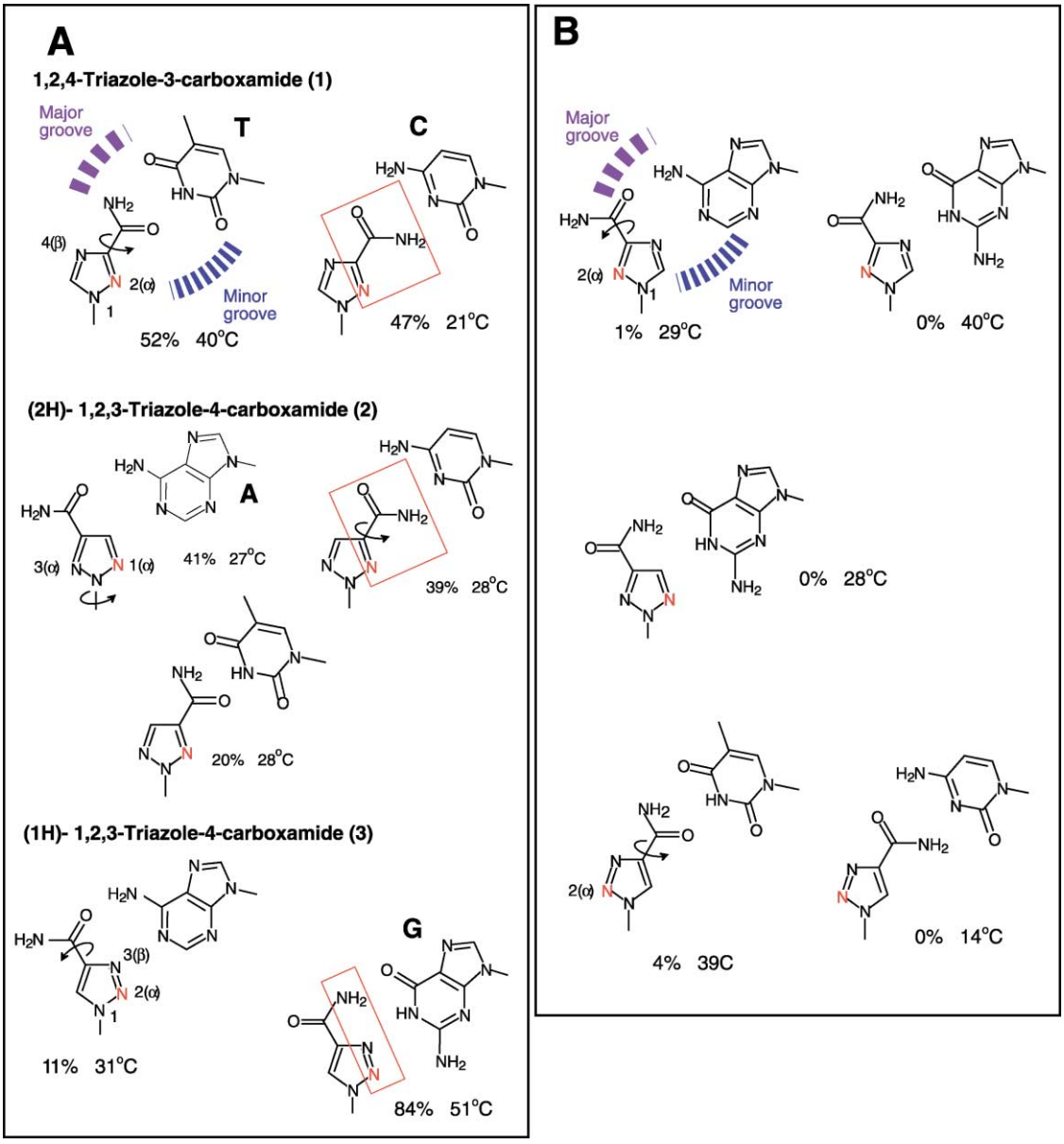



Figure 4. The Model for Orientation of the Template Azole Nucleobases with Respect to the Incoming dNTP Substrate

(A) These base conformations are predicted to be those that lead to incorporation of the specific dNTP and extension of the primer strand DNA. The percentages are stated for the relative amount of the specific dNTP in the PCR product (Figure 3) and the $T_m$ value from Table 1.

(B) Shown are the conformations of the template azole nucleobases that do not lead to minimal or no incorporation of the specific dNTP and the corresponding $T_m$.

β nitrogen (azole **3**, Figure 4A), the molecule appears to adopt a conformation that enables it to function as a universal pyrimidine. Triazole **2** contains a nitrogen atom in both α positions. Consequently, it assumes either a pyrimidine- or a purine-like configuration during extension. Indeed, this base efficiently directs the incorporation of three nucleotides, dA, dC, and T.

### Empirical Rule for Minor Groove Orientation

Among the three triazole carboxamides included in this study and the four pyrazole, imidazole, and pyrrole carboxamides previously investigated as substrates for *Taq* DNA polymerase [33], only triazole **3** directs the insertion of dG. Triazole **3** is the only azole nucleobase that is configured to orient the α nitrogen toward the minor groove while positioning both the β nitrogen and the amide-$NH_2$ for hydrogen bonding to dG (Figure 4A versus 4B). In contrast, azole carboxamides **1** and **2**, as well as pyrazole 3-carboxamide (**4**) [33], are configured so that when the α nitrogen orients toward the minor groove then the amide group is in position for hydrogen bonding to dC (Figure 4A). For those azole carboxamide nucleobases that direct the incorporation of dA or T, there are no strict requirements for complementary hy-

drogen bonding, but the position of the α nitrogen strongly influences the preference for dA or T (Figure 4B). To direct the insertion of T, the carboxamide must be on the β carbon directly adjacent to the α nitrogen (e.g., **1** and **4**), while specificity for dA is observed when the α nitrogen and β carboxamide are on opposite sides of the ring (pyrazole-4-carboxamide) [33].

The conformational relationships for the ring nitrogens and carboxamide groups in these azoles (Figures 4A and 4B) represent an empirical observation based upon the principles of canonical Watson-Crick base pairs. However, the inferior stability of the azole hydrogen bonding relationships with natural bases reveals that simple base pairing rules cannot explain the nucleotide substrate selection by *Taq* DNA polymerase (Table 2) [36–38, 40]. Explanations such as "enforced" Watson-Crick interactions (through induced fit) that have been used to explain the basis for how DNA polymerases select a dNTP substrate [7] are not sufficient to explain the results observed for the azole carboxamide templates.

### Base Pair Contacts in *Taq* DNA Polymerase

The template effects of the azole nucleobases with *Taq* DNA polymerase are dependent upon the nitrogen substitution pattern. In turn, these results indicate that additional nonbonding interactions are important in the enzyme-template-substrate complexes. The azole bases offer a different perspective from other nucleobases used to study Pol I polymerases. Previously studied artificial nucleobase systems have recognition features based upon hydrogen bonding capacity [13, 18–20] and/or shape (hydrophobics) as a strategy of maintaining the overall base pair stability [21, 23–25, 41]. In contrast, the series of azole carboxamide nucleobases studied to date have similar shapes, reduced hydrogen bonding potential, and generally unstable duplex base pairs [36, 37, 40, 42, 43]. The results with nucleobases **1–3** and those previously studied [33] represent an example of how *Taq* DNA polymerase can detect specific electronic features of the template-substrate pair during the substrate selection process. It is likely that the electronic distribution plays a significant role in dictating the base-pairing property. Earlier reports have described hydrophobic analogs that are excellent substrates for the Pol I family of DNA polymerases, when pairing with dA. As suggested in these studies with nonpolar nucleoside analogs, this could be due to the better base stacking ability and solvation effects [1, 24, 25, 44].

Important insights regarding the common structural and functional features of DNA polymerases have emerged [6, 7, 45]. Several key features derived from the X-ray crystal structures for the Pol I family of DNA polymerases have been revealed [3, 4, 42, 43]. Of particular significance are the recent advances in the structural data for the *Taq* DNA polymerase derivative Klentaq I. Two important ternary complexes of primer-template-ddNTP have been solved and reveal the conformational distinctions between the "open" and "closed" forms [2]. A series of primer/template and ddNTP structures in conformational states that represent the "closed complex" of Klentaq have also been solved [3]. Figure 5A is a model with the template azole nucleobase that highlights the extensive conformational changes in the template base and several active site residues that must occur to render the initial substrate-enzyme-template open complex to the closed state. The open forms do not involve any base-base recognition between the dNTP and the template for either the natural or azole nucleobases. In these closed complexes, the ddNTP and primer strand are poised for formation of the phosphodiester bond and offer another vantage for understanding the features of the enzyme that are critical in substrate recognition.

The molecular environment that exists in the closed form would be sensitive to the nitrogen substitution patterns. A close examination of amino acid contacts in the closed conformation with the template/NTP base pair in Klentaq [2, 3, 46, 47] and the related Pol I enzyme from T7 [4] reveals a major feature of the closed conformation that involves an ordered water molecule on the incipient minor groove edge of the base pair in the closed conformation [2]. Amino acid residues Glu615, Asn750, and Gln754 all participate in hydrogen bonding with the ordered water molecule (Figure 5B). Together with the conserved residues that line the top of the bases (mentioned above), the tightly held water molecule resides within 3 Å of the substrate base (N2 or O3) and forms an electrostatic "wall" on the minor groove edge of the incipient base pairs. In addition, Tyr671 is also positioned in a region that forms a "cap" and completes a cage around the minor groove side of the incipient base pair. Not pictured is an additional Arg573 residue that is located within hydrogen bonding distance to the water molecule that also contributes to the electronic environment on the minor groove side [3]. These intimate relationships between the amino acid side chains and the template-substrate complex are completely conserved within the T7 DNA polymerase structure [4]. Mutations of the corresponding residues in the Klenow fragment enzyme indicate functional roles for these residues in the extension step [48]. As proposed in Figure 4, a model consistent with our observations regarding the template-substrate specificity indicates the importance of a nitrogen atom oriented toward the minor groove side of the base pair. The structural data are consistent with a functional role for this nitrogen atom to reside in a region close to the electrostatic pocket.

These closed complexes show an additional type of nonbonding interaction on the top of the base pairs. Two amino acid methylene groups are positioned directly above the aromatic π system of the base pair. Among the Pol I DNA polymerases, Phe667 and Gly668 are conserved residues that undergo major conformational changes upon forming the closed ternary complex, as shown in Figures 5C and 5D. The Klentaq conformation positions the methylene group of Gly668 in a CH/π interaction with the imidazole ring of the template base, guanine. The methylene group of Phe667 lies over the hydrogen bonding region, while the aromatic ring forms a wall over the substrate-template base pair. In the T7 DNA polymerase complex with cytosine as the template site, the corresponding methylene groups of Tyr526 and Gly527 maintain a similar relationship to the

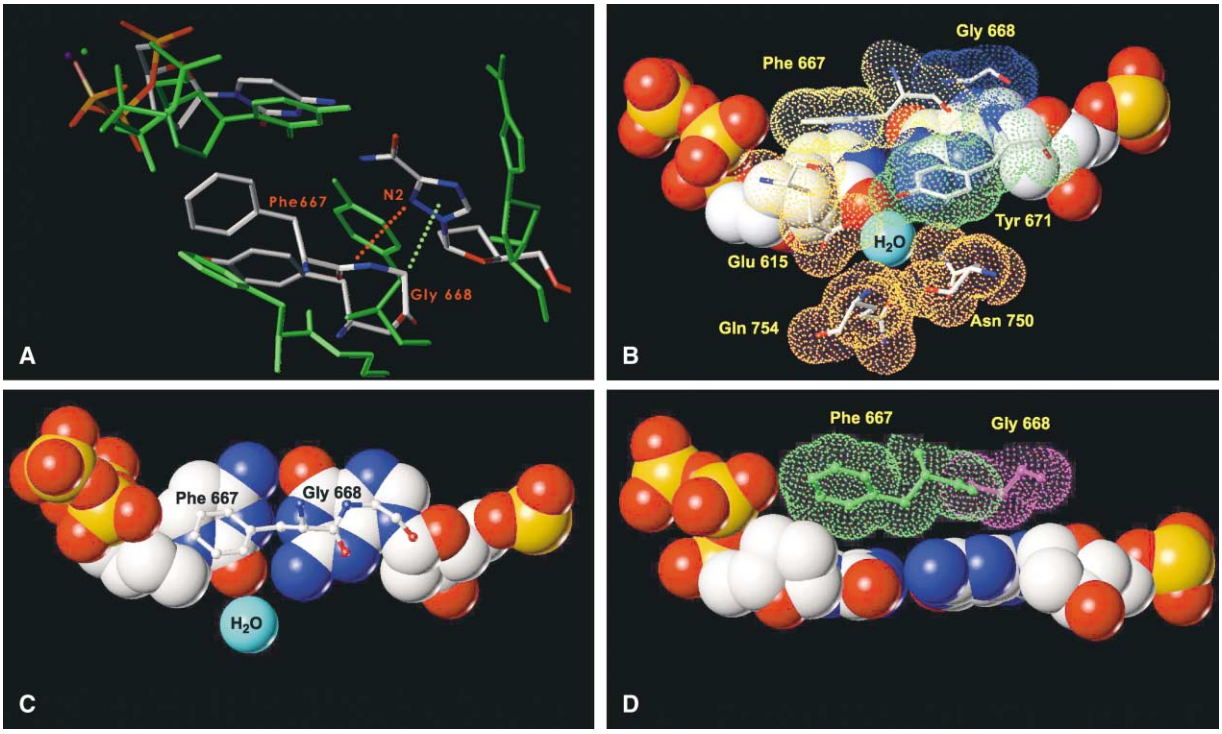



Figure 5. Models of the Key Residues Interacting with the Substrate-Template Pair in the Ternary Complex of *Taq* DNA Polymerase

(A) Stick figure illustrating the open (green) to closed (multicolor) transition in the ternary complex of *Taq* DNA polymerase, primer-template, and dCTP. The template G base has been removed and replaced with the azole nucleobase **1** and is shown along with the amino acids Phe668, Gly667, and Tyr671.

(B) Minor groove view of the G:C base pair with the order water molecule as discussed in the text. The amino acids residues (dot surfaces) around the water molecule include Glu615, Asn750, and Gln754. Tyr671 is also shown along with Phe667 and Gly668 over the top of the G:C base pair. For clarity, this graphic omits residue R573.

(C) Top view of the G:C base pair (space fill) at the active site in the closed ternary complex of *Taq* DNA polymerase with Phe667 and Gly668 in stick. The substrate in this case is dCTP.

(D) Side view of the same G:C base pair with Phe667 and Gly668 in space fill (dot surfaces) to show the close proximity of the substrate β-phosphate oxygen to the aromatic side chain.

template bases. Tyr526 of T7 DNA polymerase is in a different position relative to that observed in Klentaq, but remains stacked above the purine π system of the incoming substrate analog, ddGTP (data not shown). For every case mentioned, the distances from the amino acid methylene carbon to the specific atom of the nucleic acid bases are in the range of 3.3 to 3.8 Å. The carbonyl amide of Phe667 also lies close to N2 and O3 of the respective bases in the template site (Figure 5C and 5D). In the case of Klentaq, these interactions are present in all four template/substrate pairs that have been solved to date [3].

The CH/π interaction is one that is orientation dependent and entropically favored. These interactions have been described as behaving like weak hydrogen bonds between soft acids and soft bases [49]. CH/π interactions involving adenine and guanine ring nitrogen atoms (aromatic $sp^2$ hybridized) and a methylene side chain of other protein amino acid show the same distance relationships in the Klentaq models [50–52]. The optimal point of contact with the amino acid methylene groups is expected to shift for analogs **1–5** relative to the natural bases. Likewise, the position of the amide carbonyl of Phe667 would also reposition dependent upon the location of the minor groove "N2- or O3-like" atoms in the nucleobase analog. The conformational flexibility of these azoles through simple "flips" and amide side chain rotations allows each analog to reposition its electronic surface to optimize these nonbonding interactions.

The positioning of the amino acid side chains over the top of the substrate-template base pair could affect the stability of the complex and the rate of the incipient bond-forming step. As shown in Figure 5D, Phe667 is in close contact with the β-phosphate oxygen in the dNTP substrate; a small translation of the side chain would result in a repulsive van der Waals interaction. The net effect would be minor alterations of the phosphate-oxygen bridge conformation, which is known to have a direct effect on the transition state for reaction with the incoming 3′-hydroxyl group of the template [7]. Mutations in this position in both *Taq* and T7 DNA polymerases have been observed to affect the substrate selectivity, consistent with these structural observations [9, 10]. Reduced fidelity effects for mutations at this position in the *E. coli* Klenow fragment have long been known [53].

From previous studies of Pol I polymerases, distinct conformational changes have been indicated in the cat-

alytic cycle [54, 55]. The formation of the closed conformation has been postulated as the key kinetic barrier prior to the chemical bond forming step. The enzyme-substrate-template structures now indicate extensive nonbonding interactions for nucleobase recognition in the closed conformation. However, their impact on the overall kinetic mechanism for selection of the dNTP remains ambiguous. From our modeling studies, it is clear that the *Taq* DNA polymerase active site accommodates the azole nucleobases and allows for different conformations that would arise from reversible transitions from the open and closed states. An important aspect of the substrate selection process could be the relative stabilization of the closed ternary complex [41, 56]. Each azole conformation offers differing degrees of stabilizing and destabilizing interactions in the closed ternary complexes. The details of the mechanism and relative contributions will require a better understanding of the kinetic consequences of azole nucleobase substitutions in *Taq* DNA polymerase turnover.

## Significance

**A set of isosteric azole nucleobase analogs with unique electrostatic properties has been developed that provide a means to probe features of DNA polymerase-template-substrate recognition distinct from base-base hydrogen bonding and shape. When these azole nucleobases are used as part of DNA templates, they offer identical overall spatial configurations, and only the distinctive electronic features of each ring system influences the substrate selection process at the DNA polymerase extension site. Based upon our analysis of the *Taq* DNA polymerase with 1–5 and the current structural data for the ternary complexes in the Pol I family, there is a specific recognition surface that is created on the minor groove side of the incipient substrate-template pair. In addition, the top face of the substrate-template base pair constitutes an important recognition feature for specific interactions with the DNA polymerase. The $\pi-$ facial and electrostatic recognition features in addition to base pair shape recognition occur only in the closed conformation. Together, they constitute a more complete perspective on the molecular basis for fidelity of nucleic acid synthesis and indicate that weak, cumulative nonbonding interactions are important in substrate selection. These recognition features are distinct from minor groove contacts recently identified distal to the active site between the DNA polymerase, the purine N3, and pyrimidine O2, which have a likely role in primer extension efficiency. As a set of general probes, the azole nucleobase will open the door to a more detailed study of the nonbonding interactions and dynamics in other polymerase families. The role of the specific amino acid residues in other conserved families of DNA polymerases is not expected to be the same as in Pol I polymerases. However, the distinctions at this level of molecular analysis are likely to reveal important evolutionary paths that relate ultimately to the context of their biological functions.**

## Experimental Procedures

Synthesis of the three deoxyribonucleosides has been reported [40, 57]. For incorporation into oligonucleotides, the three nucleosides were transformed into 5′-dimethoxytrityl-3′-phosphoramidites as described for the structurally related imidazole carboxamide [37]. The self-complementary oligonucleotides used for $T_m$ experiments were synthesized and purified as previously described for related azole nucleoside analogs [36]. Sense strand 40-mer primers for PCR corresponding to positions −10 to 30 of the *E. coli hisF* gene in the expression vector p*hisF-tac* were prepared with the analogs incorporated in position 36 [33]. A 30-mer antisense primer was composed entirely of natural bases. All oligonucleotide primers were prepared either (1) by using an ABI 380B or 392 DNA synthesizer in the Laboratory for Macromolecular Structure at Purdue University or (2) at Midland Certified Reagent Company (Midland, TX) using phosphoramidites described above. All primers were purified by urea-PAGE before use. Template DNA for PCR was produced by PvuII digestion of the expression vector p*hisF-tac*.

The 40-mer oligonucleotides containing modified bases served as sense strand primers for polymerase chain reactions (PCRs) catalyzed by the *Taq* DNA polymerase (Amplitaq from Perkin-Elmer). PCR, sequencing, and quantitation of the sequencing gels were accomplished as previously described [33]. Area integration of all four lanes of a sequencing gel imaged by a Storm 860 Phosphorimager (Molecular Dynamics) yielded data on the relative amounts of nucleotide incorporation for a given nucleobase. Averaging between four and eight sequencing gel images assessed the recognition of each azole nucleobase in a DNA template. The sequencing data in Table 2 represent values for nucleobase incorporation that showed a standard deviation of <8%.

A single-nucleotide incorporation assay was used to measure the steady-state extension efficiency against the natural and the artificial base in the template DNA [39]. The primer sequence was 5′-[Cy5]-TAATACGACTCACTATAGGGAGA (23-mer) for all the experiments. The template sequence was 3′-ATTATGCTGAGTGATAT CCCTCTXGTCA (28-mer), where X represents the natural or artificial base. The labeled primer and unlabeled template were mixed at a ratio of 1:2 and used at a final concentration of 125 nM in each 50 μL reaction containing 10 mM Tris-HCl (pH 8.3) (at 25°C), 50 mM KCl, 1.5 mM $MgCl_2$, 0.001% gelatin, and different concentrations of dNTP (Promega) in each reaction. The concentration of the dNTPs ranged from 10 to 1000 μM for the artificial bases and from 0.5 to 50 μM for the natural bases. Incubations were carried out at 70°C on mixing with Amplitaq DNA polymerase (Perkin-Elmer, 2.1–8.5 ng per reaction) for ten different time intervals ranging from 4 to 90 min for the artificial bases and from 6 to 60 min for the natural bases. In cases of lower reaction rates, the incubations were carried out for 1–4 hr, with suitable enzyme concentration to attain 5%–20% turnover. The reactions were quenched using the buffer containing 80% formamide, 2 mM EDTA, and Tris Borate EDTA buffer. The parameters were chosen so that the turnover lies below 20%. The reactions were analyzed on 20% acrylamide (19:1 mono:bis) denaturing (7 M urea) gels (30 cm × 40 cm × 0.4 mm). The products were quantified using red fluorescence on a laser-based imager, Storm 860 (Molecular Dynamics). The $V_{max}$ and $K_m$ values were determined by plotting the hyperbola using Origin 6.0.

### Supplemental Data

General materials and procedures for the synthesis of phosphoramidites of **1**, **2**, and **3**, analysis of oligonucleotides by enzymatic digestion and HPLC, thermal melting studies, PCR, and DNA sequencing and analysis are available at http://www.chembiol.com/cgi/content/full/10/9/815/DC1.

### Acknowledgments

Financial support from the National Institutes of Health GM 53155 is gratefully acknowledged. We acknowledge the fellowship to V.C.N. from the Indiana Instrumentation Institute (III) supported by the Indiana 21st Century Fund. Many helpful discussions with Alan Friedman are gratefully acknowledged. We are also grateful to Xiaoming Chen for technical assistance.

Received: April 9, 2003
Revised: June 27, 2003
Accepted: July 21, 2003
Published: September 19, 2003

**References**

1. Kool, E.T. (2002). Active site tightness and substrate fit in DNA replication. Annu. Rev. Biochem. *71*, 191–219.
2. Li, Y., Korolev, S., and Waksman, G. (1998). Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of *Thermus aquaticus* DNA polymerase I: structural basis for nucleotide incorporation. EMBO J. *17*, 7514–7525.
3. Li, Y., and Waksman, G. (2001). Crystal structures of a ddATP-, ddTTP-, ddCTP-, and ddGTP-trapped ternary complex of Klentaq I: insights into nucleotide incorporation and selectivity. Protein Sci. *10*, 1225–1233.
4. Kiefer, J.R., Mao, C., Bramen, J.C., and Beese, L.S. (1998). Visualizing DNA replication in a catalytically active *Bacillus* DNA polymerase crystal. Nature *391*, 304–307.
5. Doublié, S., Tabor, S., Long, A.M., Richardson, C.C., and Ellenberger, T. (1998). Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution. Nature *391*, 251–258.
6. Steitz, T.A. (1999). DNA polymerases: structural diversity and common mechanisms. J. Biol. Chem. *25*, 17395–17398.
7. Brautigam, C.A., and Steitz, T.A. (1998). Structural and functional insights provided by crystal structures of DNA polymerases and their substrate complexes. Curr. Opin. Struct. Biol. *8*, 54–63.
8. Eckert, K.A., and Kunkel, T.A. (1990). High fidelity DNA synthesis by the *Thermus aquaticus* DNA polymerase. Nucleic Acids Res. *18*, 3739–3744.
9. Suzuki, M., Yoshida, S., Adman, E.T., Blank, A., and Loeb, L. (2000). *Thermus aquaticus* DNA polymerase I mutants with altered fidelity. J. Biol. Chem. *275*, 32728–32735.
10. Tabor, S., and Richardson, C.C. (1995). A single residue in DNA-polymerases of the *Escherichia coli* DNA polymerase I family is critical for distinguishing between deoxyribonucleotides and dideoxyribonucleotides. Proc. Natl. Acad. Sci. USA *92*, 6339–6343.
11. Tosaka, A., Ogawa, M., Yoshida, S., and Suzuki, M. (2001). O-helix mutant T664P of *Thermus aquaticus* DNA polymerase I. J. Biol. Chem. *276*, 27562–27567.
12. Li, Y., Mitaxov, V., and Waksman, G. (1999). Structure-based design of *Taq* DNA polymerases with improved properties of dideoxynucleotide incorporation. Proc. Natl. Acad. Sci. USA *96*, 9491–9496.
13. Horlacher, J., Hottiger, M., Podust, V.N., Hubscher, U., and Benner, S.A. (1995). Recognition by viral and cellular DNA polymerases of nucleosides bearing bases with nonstandard hydrogen bonding patterns. Proc. Natl. Acad. Sci. USA *92*, 6329–6333.
14. Brown, D.M., and Lin, P.K.T. (1991). Synthesis and duplex stability of oligonucleotides containing adenine-guanine analogues. Carbohydr. Res. *216*, 129–139.
15. Brown, D.M., and Lin, P.K.T. (1991). The structure and application of oligodeoxyribonucleotides containing modified, degenerate bases. Nucleic Acids Research Symposium Series *24*, 209–212.
16. Hill, F., Loakes, D., and Brown, D.M. (1998). Polymerase recognition of synthetic oligodeoxyribonucleotides incorporating degenerate pyrimidine and purine bases. Proc. Natl. Acad. Sci. USA *95*, 4258–4263.
17. Bergstrom, D.E. (2001). Unnatural nucleosides with unusual base pairing properties. In Current Protocols in Nucleic Acid Chemistry, D.E.B.S.L. Beaucage, G.D. Glick, and R.A. Jones, eds. (New York: John Wiley and Sons), pp. 1.4.1–1.4.13.
18. Piccirilli, J.A., Krauch, T., Moroney, S.E., and Benner, S.A. (1990). Enzymatic incorporation of a new base pair into DNA and RNA extends the genetic alphabet. Nature *343*, 33–37.
19. Switzer, C.Y., Moroney, S.E., and Benner, S.A. (1993). Enzymatic recognition of the base pair between isocytidine and isoguanosine. Biochemistry *32*, 10489–10496.
20. Voegel, J.J., and Benner, S.A. (1994). Nonstandard hydrogen bonding in duplex oligonucleotides. The base pair between an acceptor-donor-donor pyrimdine analog and a donor-acceptor-acceptor analog. J. Am. Chem. Soc. *116*, 6929–6930.
21. Kool, E.T. (1998). Replication of non-hydrogen bonded bases by DNA polymerases: a mechanism for steric matching. Biopolymers *48*, 3–17.
22. Guckian, K.M., and Kool, E.T. (1997). Highly precise shape mimicry by a difluorotoluene deoxynucleoside, a replication-competent substitute for thymidine. Angew. Chem. Int. Ed. Engl. *36*, 2825–2828.
23. Guckian, K.M., Morales, J.C., and Kool, E.T. (1998). Structure and base pairing properties of a replicable nonpolar isostere for deoxyadenosine. J. Org. Chem. *63*, 9652–9656.
24. Moran, S., Ren, R.X.F., and Kool, E.T. (1997). A thymidine triphosphate shape analog lacking Watson-Crick pairing ability is replicated with high sequence selectivity. Proc. Natl. Acad. Sci. USA *94*, 10506–10511.
25. Moran, S., Ren, R.X.F., Rumney, S.I., and Kool, E.T. (1997). Difluorotoluene, a nonpolar isostere for thymine, codes specifically and efficiently for adenine in DNA replication. J. Am. Chem. Soc. *119*, 2056–2057.
26. Morales, J.C., and Kool, E.T. (2000). Varied molecular interactions at the active sites of several DNA Polymerases: nonpolar nucleoside isosteres as probes. J. Am. Chem. Soc. *122*, 1001–1007.
27. Fujiwara, T., Kimoto, M., Sugiyama, H., Hirao, I., and Yokoyama, S. (2001). Synthesis of 6-(2-thienyl)purine nucleoside derivatives that form unnatural base pairs with pyridinone nucleosides. Bioorg. Med. Chem. Lett. *11*, 2221–2223.
28. Ogawa, A.K., Wu, Y., Berger, M., Schultz, P.G., and Romesberg, F.E. (2000). Rational design of an unnatural base pair with increased kinetic selectivity. J. Am. Chem. Soc. *122*, 8803–8804.
29. Ohtsuki, T., Kimoto, M., Ishikawa, M., Mitsui, T., Hirao, I., and Yokoyama, S. (2001). Unnatural base pairs for specific transcription. Proc. Natl. Acad. Sci. USA *98*, 4922–4925.
30. McMinn, D.L., Ogawa, A.K., Wu, Y., Liu, J., Schultz, P.G., and Romesberg, F.E. (1999). Efforts toward expansion of the genetic alphabet: DNA polymerase recognition of a highly stable, self-pairing hydrophobic base. J. Am. Chem. Soc. *121*, 11585–11586.
31. Tae, E.L., Wu, Y., Xia, G., Schultz, P.G., and Romesberg, F.E. (2001). Efforts toward expansion of the genetic alphabet: replication of DNA with three base pairs. J. Am. Chem. Soc. *123*, 7439–7440.
32. Wu, Y.O., A.K. Berger, M., McMinn, D.L., Schultz, P.G., and Romesberg, F.E. (2000). Efforts toward expansion of the genetic alphabet: optimization of interbase hydrophobic interactions. J. Am. Chem. Soc. *122*, 7621–7632.
33. Hoops, G.C., Zhang, P., Johnson, W.T., Paul, N., Bergstrom, D.E., and Davisson, V.J. (1997). Template directed incorporation of nucleotide mixtures using azole-nucleobase analogs. Nucleic Acids Res. *25*, 4866–4871.
34. Hunter, W.N. (1992). Crystallographic studies of DNA containing mismatches, modified and unpaired bases. In Methods in Enzymology: DNA Structures Part A: Synthesis and Physical Analysis of DNA, D.M.J. Lilley, and J.E. Dahlberg, eds. (New York: Academic Press), pp. 221–231.
35. Kennard, O. (1987). The molecular structure of base pair mismatches. In Nucleic Acids and Molecular Biology, F. Eckstein, and D.M.J. Lilley, eds. (Berlin: Springer-Verlag), pp. 25–52.
36. Zhang, P., Johnson, W.T., Klewer, D., Paul, N., Hoops, G., Davisson, V.J., and Bergstrom, D.E. (1998). Exploratory studies on azole carboxamides as nucleobase analogs: thermal denaturation studies on oligodeoxyribonucleotide duplexes containing pyrrole-3-carboxamide. Nucleic Acids Res. *26*, 2208–2215.
37. Johnson, W.T., Zhang, P., and Bergstrom, D.E. (1997). The synthesis and stability of oligodeoxyribonucleotides containing the deoxyadenosine mimic 1-(2′-deoxy-β-D-ribofuranosyl)imidazole-4-carboxamide. Nucleic Acids Res. *25*, 559–567.
38. Bergstrom, D.E., Zhang, P., and Johnson, W.T. (1996). Design and synthesis of heterocyclic carboxamides as natural nucleic acid mimics. Nucleosides Nucleotides *15*, 59–68.
39. Boosalis, M.S., Petruska, J., and Goodman, M.F. (1987). DNA

polymerase insertion fidelity. Gel assay for site-specific kinetics. J. Biol. Chem. *262*, 14689–14696.
40. Pochet, S., and Dugue, L. (1998). Imidazole-4-carboxamide and 1,2,4-triazole-3-carboxamide deoxyribonucleotides as simplified DNA building blocks with ambiguous pairing capacity. Nucleosides Nucleotides *17*, 2003–2009.
41. Dzantiev, L., Alekseyev, Y.O., Morales, J.C., Kool, E.T., and Romano, L.J. (2001). Significance of nucleobase shape complementarity and hydrogen bonding in the formation and stability of the closed polymerase-DNA complex. Biochemistry *40*, 3215–3221.
42. Pochet, S., Dugue, L., Sala, M., Pezo, V., and Wain-Hobson, S. (1997). Ambiguous base pairing of 1-(2-deoxy-β-D-ribofuranosyl)imidazole-4-carboxamide during PCR. Nucleosides Nucleotides *16*, 1749–1752.
43. Sala, M., Pezo, V., Pochet, S., and Wain-Hobson, S. (1996). Ambiguous base pairing of the purine analogue 1-(2-deoxy-β-D-ribofuranosyl)imidazole-4-carboxamide. Nucleic Acids Res. *24*, 3302–3306.
44. Berger, M., Wu, Y., Ogawa, A.K., McMinn, D.L., Schultz, P.G., and Romesberg, F.E. (2000). Universal bases for hybridization, replication and chain termination. Nucleic Acids Res. *28*, 2911–2914.
45. Patel, P.H., Suzuki, M., Adman, E., Shinkai, A., and Loeb, L.A. (2001). Prokaryotic DNA polymerase I: evolution, structure, and "base flipping" mechanism for nucleotide selection. J. Mol. Biol. *308*, 823–837.
46. Li, Y., Kong, Y., Korolev, S., and Waksman, G. (1998). Crystal structures of the Klenow fragment of *Thermus aquaticus* DNA polymerase I complexed with deoxyribonucleoside triphosphates. Protein Sci. *7*, 1116–1123.
47. Li, Y., and Waksman, G. (2001). Structural studies of the Klentaq I DNA polymerase. Curr. Org. Chem. *5*, 871–883.
48. Spratt, T.E. (2001). Identification of hydrogen bonds between Escherichia coli DNA polymerase I (Klenow fragment) and the minor groove of DNA by amino acid substitution of the polymerase and atomic substitution of the DNA. Biochemistry *40*, 2647–2652.
49. Nishio, M., Hirota, M., and Umezawa, Y. (1998). The CH/π Interaction. Evidence, Nature, and Consequences (New York: Wiley-VCH).
50. Chakrabarti, P., and Samanta, U. (1995). CH/π interaction in the packing of the adenine ring in protein structures. J. Mol. Biol. *251*, 9–14.
51. Moodie, S.L., Mitchell, J.B.O., and Thorton, J.M. (1996). Protein recognition of adenylate: an example of a fuzzy recognition template. J. Mol. Biol. *263*, 486–500.
52. Umezawa, Y., and Nishio, M. (1998). CH/π interactions as demonstrated in the crystal structure of guanine-nucleotide binding proteins, Src homology-2 domains and human growth hormone in complex with specific ligands. Bioorg. Med. Chem. *6*, 493–504.
53. Carroll, S.S., Cowart, M., and Benkovic, S.J. (1991). A mutant of DNA polymerase I (Klenow fragment) with reduced fidelity. Biochemistry *30*, 804–813.
54. Brandis, J.W., Edwards, S.G., and Johnson, K.A. (1996). Slow rate of phosphodiester bond formation accounts for the strong bias that *Taq* DNA polymerase shows against 2′,3′-dideoxynucleotide terminators. Biochemistry *35*, 2189–2200.
55. Johnson, K.A. (1993). Conformational coupling in DNA polymerase fidelity. Annu. Rev. Biochem. *62*, 685–713.
56. Showalter, A.K., and Tsai, M.D. (2002). A reexamination of the nucleotide incorporation fidelity of DNA polymerases. Biochemistry *41*, 10571–10576.
57. Makabe, O., Suzuki, H., and Umezawa, S. (1977). Synthesis of D-arabinofuranosyl and 2′-deoxy-D-ribofuranosyl 1,2,3-Triazolecarboxamides. Bull. Chem. Soc. Jpn. *50*, 2689–2693.
58. Dewar, M.J.S., Zoebisch, E.G., Healy, E.F., and Stewart, J.J.P. (1985). AM1: a new general-purpose quantum-mechanical molecular-model. J. Am. Chem. Soc. *107*, 3902–3909.