Enzyme Fidelity

DOI: 10.1002/anie.200504296

Remarkable Sensitivity to DNA Base Shape in the DNA Polymerase Active Site**

*Herman O. Sintim and Eric T. Kool**

The mechanisms of DNA synthesis by DNA polymerase enzymes have come under intensive study recently because of their close biological connections to cancer.[1] Replicative polymerases, such as Pol I and Pol III in bacteria, must synthesize DNA with high fidelity to avoid deleterious mutations.[2] Repair polymerases, on the other hand, operate with lower fidelity and serve to replicate DNA beyond damaged and mismatched sites, keeping cells alive while other enzymes act to correct the damage and prevent oncogenic mutations from proliferating.[3] Because of the importance to human health, a focus of research in many labs has been on the factors that govern fidelity in these enzymes.[4]

Recent studies have suggested that steric effects can have strong influences in DNA polymerase enzymes. In early studies with nonpolar nucleoside isosteres,[5] it was observed that purinelike molecular shapes can be selectively paired with pyrimidine shapes,[6] and a large pyrene nucleobase analogue was shown to be selectively replicated against a very small (abasic) nucleoside.[7] Experiments by Hirao and co-workers[8] have lent support to a strong contribution from sterics in DNA replication, and studies by Shultz, Romesberg, and co-workers[9] with methylated and fluorinated synthetic bases have shown differing efficiency with changes in substitution. Furthermore, steric effects in polymerase active sites have also been observed on the level of the sugar–phosphate backbone in experiments by Marx and co-workers.[10]

However, the influence of steric effects in polymerases has remained in question, in part owing to recent experiments by Berdis and co-workers,[11] and by Engels, Kuchta, and co-workers.[12] Berdis and co-workers studied a set of 5-substituted indoles as deoxynucleoside triphosphate derivatives and observed low selectivity for incorporation of the analogues by T4 DNA polymerase opposite the four natural DNA bases. Engels, Kuchta, and co-workers tested compounds with the related benzimidazole skeleton by using the Klenow polymerase and reported little relationship between different substitution patterns and selectivity between natural DNA bases. In both studies, it was concluded that sterics do not play an important role in the activities observed. Thus size and shape effects and their contribution to DNA polymerase fidelity remain in active debate.

We have undertaken a series of studies to begin to address the steric hypothesis in a systematic way. An earlier report evaluated the effects of increasing DNA base size while keeping the shape constant by using nonpolar thymine shape mimics (toluene, difluoro-, dichloro-, dibromo- and diiodotoluene) paired opposite adenine.[13] These compounds varied in size over a small 1.0 Å range, yet experiments showed that there were large differences in efficiency and fidelity across the series.[13c] However, steric effects depend not only on size, but also shape—that is, size at specific positions in a molecule. Herein, we investigate the effects of nucleobase shape in the DNA polymerase I active site through variation in the positions of substitution of H, F, Cl, and Br atoms on the toluene skeleton. We find surprisingly large kinetic effects resulting from subtle changes in substituent size and location.

Design of this new series started with the known 2,4-difluorotoluene deoxyriboside,[5] a nearly perfect isostere of thymidine, and in particular, the recent variant 2,4-dichlorotoluene deoxyriboside[13a] (**6**, Figure 1), which has virtually the same shape but is slightly larger. Both of these molecules are highly active with replicative DNA polymerases and encode replication opposite adenine.[13c,14] To test whether it is shape, size, or other properties of such thymidine-mimicking compounds (and by inference, thymidine itself) that make them active, we have now altered the shapes in regular increments through the substitution of larger and smaller groups along the periphery of the toluene ring. The sets of varied molecular shapes are shown in Figure 1 and Figure 2. Comparison of compounds **1**–**3** preserves the size but varies the position of a single chlorine substituent, whereas compounds **4**–**6**, with varied double substitution, helps to determine the degree to which shape effects are important for the remarkable thymine-like behavior of the 2,4-isomer in pairing and in replication.[13c] Finally, comparisons of four-compound sets allows for slow variation in size at the hypothetically important 2- and 4-positions separately (Figure 2). For example, the series **1**→**9**→**6**→**10** incrementally increases the size of the 4-position with H, F, Cl, and Br atoms,[13a,15] and the **3**→**7**→**6**→**8** series addresses effects at the 2-position in the same way.

A generalized strategy for the preparation of the new C-glycosides involved reacting the appropriate lithiated halotol-

[*] Dr. H. O. Sintim, Prof. Dr. E. T. Kool
Department of Chemistry
Stanford University
Stanford, CA 94305-5080 (USA)
Fax: (+1) 650-725-0259
E-mail: kool@stanford.edu

[**] This work was supported by the US National Institutes of Health (GM072705).

Supporting information for this article is available on the WWW under http://www.angewandte.org or from the author.

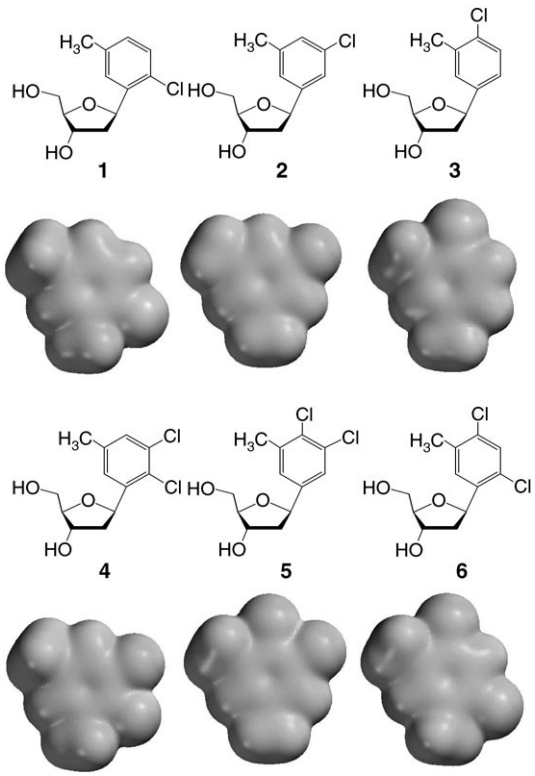
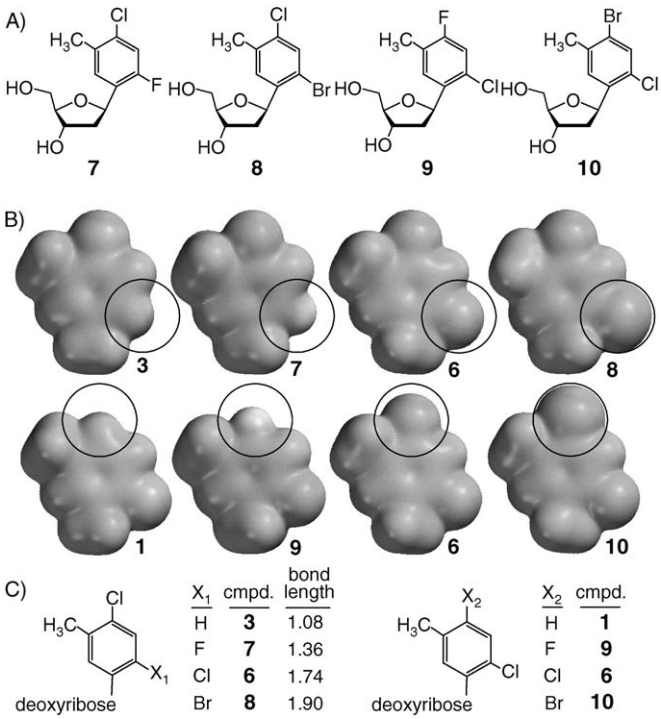


Figure 1. Structures of nonpolar thymidine analogues with systematically varied shape and constant size. Compounds **1**–**3** have a single chlorine substituent at varied positions along the Watson–Crick pairing edge, while **4**–**6** have two substituents. Space-filling models are shown with methyl group replacing deoxyribose.



X_1	cmpd.	bond length
H	**3**	1.08
F	**7**	1.36
Cl	**6**	1.74
Br	**8**	1.90

X_2	cmpd.
H	**1**
F	**9**
Cl	**6**
Br	**10**

Figure 2. Two series of nonpolar thymidine analogues with gradually varied shape. Circles highlight the position with increasing steric demand. A) Structures of deoxyribosides **7**–**10**. B) Space-filling models showing gradual increases in size at the 2-position (**3**→**7**→**6**→**8**) and separately at the 4-position (**1**→**9**→**6**→**10**). Models are shown with a methyl group replacing deoxyribose. C) Expected bond lengths at the variable positions for the two series.[28]

uene (generated through lithium–halogen exchange) with disiloxanediyl-protected deoxyribonolactone[16] **11** followed by triethylsilane reduction under Lewis acid catalysis (see the Supporting Information). We prepared the 5′-dimethoxytrityl-3′ phosphoramidite derivatives (see the Supporting Information) and incorporated them into synthetic 12 mer and 28 mer oligonucleotides by using an automated synthesizer. The 5′-triphosphate derivatives were also prepared following published procedures.[17]

To investigate the small shape effects for pairing in DNA, we evaluated the relative pairing stabilities of compounds **1**–**10** placed opposite each of the four natural DNA bases in a 12 mer duplex context[18] (see the Supporting Information). Optically monitored melting studies were used to evaluate stabilities at pH 7.0 in a buffer solution containing Mg^{2+} (10 mM) and Na^{+} (100 mM). The results showed that, despite their differences in shape, these ten molecules have quite similar pairing properties in this sequence context. With natural bases as partners in the 12 mer duplex, compounds **1**–**10** were all destabilizing relative to a natural T–A pair (also observed previously for **6**[13b] and other low-polarity compounds[18–20]) and generally displayed little pairing selectivity among the natural partners. With the exception of the 3-substituted analogues (**2**, **4**, and **5**) and the 4-chloro compound **3**, some of the new analogues showed a small but significant preference (by approximately 2–4 °C in T_m) for adenine as partner. In our previous studies with 2,4-dihalotoluene base analogues[13b] and studies of others,[9] a similar small preference for adenine was also observed; this may be due to the stronger stacking of adenine relative to other DNA bases.[21] Interestingly, two cases, the 2,4-dichloro compound **6** and the 2-chloro-4-bromo compound **7**, showed a stronger preferences for adenine over a mismatch (5–7 °C); this may be owing to a small degree of thymine-like shape preference opposite adenine conferred by the backbone. However, the selectivity was lower than that of thymine (8 °C) and the analogues were both considerably destabilizing overall (by 12 °C and greater), consistent with their low polarity and lack of significant hydrogen bonding ability. Overall, despite their differences in shape, these ten molecules have quite similar pairing properties in this sequence context, which suggests that the DNA backbone confers relatively low shape selectivity.

To determine the DNA polymerase shape selectivity, the properties of the shape-varied compounds in the active site of KF exonuclease-deficient (*exo*$^-$) polymerase were then tested. Preliminary gel analyses of single-nucleotide insertion reactions against each of the four DNA bases showed that all ten compounds were enzymatically paired with adenine and thymine in varied amounts but not significantly with guanine or cytosine (see the Supporting Information). Thus, we proceeded to measure the steady-state kinetic efficiencies when these nonpolar thymine analogues are paired with A or

T, wherein pairing with A defines absolute efficiency and pairing with T, the most efficient mismatch, defines selectivity or fidelity.

The kinetics data (shown in Figure 3, Figure 4, and tabulated in the Supporting Information) as a whole showed that the analogues were enzymatically paired with widely varying efficiencies and selectivities. As template-base analogues, the enzyme incorporated dATP opposite the compounds with efficiencies that varied by three orders of magnitude. The most efficient analogues were compounds **6**, **7**, and **10**, which gave efficiencies within error limits of each other and were only 2–3-fold less efficient than natural thymidine as a template. As incoming nucleotide (dNTP) derivatives, the compounds varied in efficiency by an even larger 3500-fold for insertion opposite a template adenine. The most efficient was compound **10**, which was tenfold below natural dTTP in efficiency. Thus a general inspection of the data shows that shape effects are clearly large in magnitude with this enzyme and that a number of the new analogues may be even more efficient as thymidine mimics than any previously reported.[13c,14]

Data for analogues in the monochlorinated series **1**–**3** showed that they were generally more efficient in a template strand than as dNTP derivatives (Figure 3A,B). The 2- and 4-substituted compounds were enzymatically paired opposite adenine preferentially, with efficiencies that were moderate to high. The 2-chloro compound **1** showed greater selectivity for adenine than did the 4-chloro case **3**. The selectivity of **1** was moderate, with a 34-fold to 470-fold preference for adenine over the most efficient mismatch (thymine). In contrast to this, the 3-substituted compound **2** was paired very poorly with adenine, 3–4 orders of magnitude below thymidine, and the selectivity underwent a switch in the template strand with a small preference for being paired with T over A.

The second series (Figure 3C,D) compared enzymatic processing of doubly chlorinated toluenes (**4**–**6**). Of the three, only the 2,4-dichloro compound **6** showed selectivity for adenine (selectivity was high, 64- to 810-fold), as previously reported.[13c] In contrast to this, the other two compounds (**4** and **5**) showed selectivity for thymine over adenine. Thus these two 3-substituted compounds demonstrated a selectivity switch as had the other 3-substituted compound **3**, above.

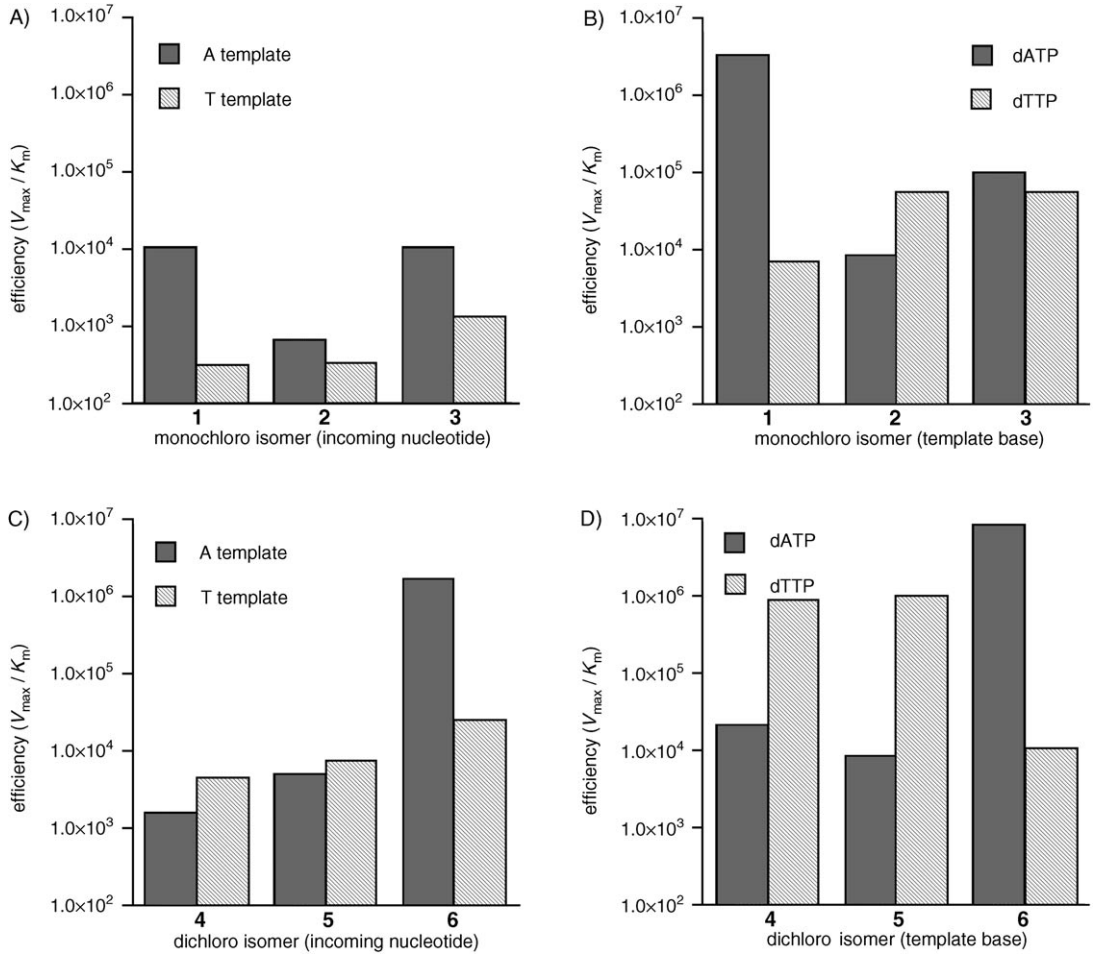


Figure 3. Shape effects on DNA replication, as measured with thymidine analogues having the same size but altered shapes. A) and B) Data for monochlorinated analogues **1**–**3**. C) and D) Data for dichlorinated analogues **4**–**6**. Kinetic efficiencies with the Kf enzyme are plotted on a $\log_{10}$ scale; black columns show data for replication opposite deoxyadenosine (as expected for a thymidine analogue), and striped columns show data for replication against thymidine (the most efficient mismatch). Data are shown both for analogues as incoming nucleotides (A,C) and as template bases (B,D).

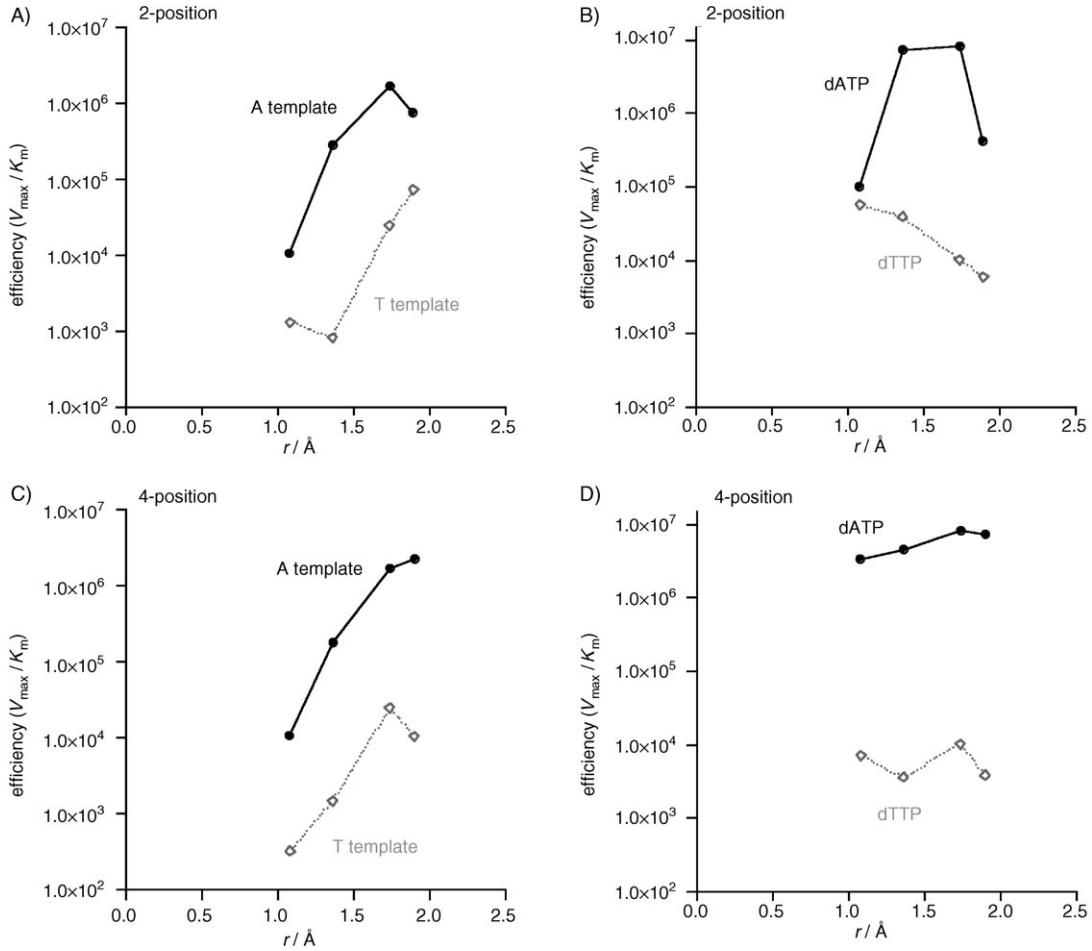


Figure 4. Effects of a gradually altering shape on DNA replication as measured with a series of thymidine analogues with one substituent of varied size (over an 0.8 Å bond length range, Figure 2). Kinetic efficiencies with the Kf enzyme are plotted on a log scale. A) and B) Data for varied size at the 2-position (analogues **3**, **7**, **6**, and **8**). C) and D) Data for varied size at the 4-position (analogues **1**, **9**, **6**, **10**). Black lines show data for replication against deoxyadenosine and gray dashed lines show data for mismatched replication opposite the most efficient mismatch (thymidine). Data are shown both for analogues as incoming nucleotides and as template bases. r = bond length.

The above data suggests strongly that the 2- and 4-positions were important in the determination of thymidine-like pairing behavior. Thus two series of compounds were evaluated to investigate the effects of gradual shape changes at these positions individually, by slowly increasing size (Figure 4 A–D). The data for the 2-position size variation are plotted in Figure 4 A, B and those for 4-position variation are given in Figure 4 C,D.

The incoming nucleotide analogues had gradually varied shapes (Figure 4 A, C), and the results showed that the substituent size preference was strong at both the 2- and 4-positions, with a 110-fold and 210-fold difference over the size range. The maximum efficiency at the 4-position occurred with bromo substitution, and this size preference appeared to be slightly larger than that at the 2-position, wherein the maximum appeared upon chloro substitution. The effects in the template strand (Figure 4 B, D) were somewhat similar in that the 4-position had a larger size preference (chloro/bromo) as compared with the 2-position (fluoro/chloro). However, there was one substantial difference, the 4-position selectivity overall was quite modest with a high efficiency regardless of substituent size (Figure 4 D).

Data were also plotted for the most efficient mismatch (replication opposite thymidine; Figure 4, gray dashed lines), which allowed for evaluation of the effects of gradually altered nucleobase shape on pairing selectivity with this enzyme. As a whole, the analogue with highest fidelity in a template strand was the 2-chloro-4-bromo compound **10**; as incoming nucleotides, fidelities were the same for compound **10** and for the 2-fluoro-4-chloro compound **7**. The overall fidelity of compound **10** as an incoming nucleotide analogue was approximately threefold lower than natural thymidine (230-fold preference for adenine versus 790-fold). In the template strand, however, compound **10** was processed with higher fidelity than thymidine (1900-fold versus 1100-fold) although identity of the most common mismatch was different.

Taken together, the results demonstrate that the Klenow DNA polymerase is highly sensitive to even small differences in nucleobase shape. Overall, the shape analogues in this

series displayed a remarkable 3500-fold range in enzymatic efficiency despite the fact that they are all closely related chlorinated toluenes. We showed previously that this same enzyme is responsive to nucleobase analogues with variable size but that maintain thymine-like shape.[13c] The present results are distinct in that they shed light on the importance of the location of steric substitutions rather than on the size of the nucleobase as a whole. Changing a hydrogen substituent to a chlorine (in dNTP analogues) at positions 2 or 4 has a strong positive effect (>100-fold) on the efficiency of insertion opposite adenine (compare **1** with **6**, and **3** with **6**). In marked contrast, making the same steric change at the 3-position has a negative effect on insertion opposite adenine (by a factor of 2–5; compare **1** with **4** and **3** with **5**).

We hypothesize that these local steric preferences are defined by the shape of the base opposite these analogues,[22] and that the magnitude of the shape preferences is enforced by the active-site tightness of the enzyme around the incipient base pair.[23,13c] The kinetic optimization that occurred here with varying shape may reflect the optimal filling of space in the active site; local steric protrusions that clash with adenine (such as at the thymine 3-position) or with the enzyme at other positions would be energetically costly at the closed transition state for phosphodiester bond formation. On the other hand, groups that are smaller than needed to fill the space leave a void in the protein, which is also energetically costly.[24]

We considered whether other factors apart from sterics, such as polarity or electrostatic effects, might explain these enzymatic results. For example, nucleobase dipoles might conceivably influence base pairing or active site binding.[25] However, comparison of members of the analogue series **1**–**3** and **4**–**6** reveals widely varying replication efficiencies with essentially constant dipole strengths. In addition, local electrostatic charges might conceivably contribute. However, examination of data for the increasing size series (**3**→**7**→**6**→**8** and **1**→**9**→**6**→**10**), showed that the maximum efficiencies occur with the largest halogens (bromine and chlorine), whereas the strongest negative charge is associated with fluorine. Moreover, local electrostatic charges associated with chlorine would remain virtually the same in compounds **1**–**4** and in **4**–**6**, but the replication efficiencies vary by orders of magnitude. Thus we conclude that it is the size and location of space-filling substituents, rather than polar factors, that best explains these results.

Earlier experiments with larger nonpolar nucleobase analogues (indoles and benzimidazoles) led to the suggestion that steric effects may not be a dominant factor within this and other enzymes.[11,12] Herein, we offer a steric argument that may explain a number of those findings. The indole and closely related benzimidazole frameworks can exist in an alternative *syn* conformation, such as is found for 8-oxoGuanine in its mispairing opposite adenine.[26] We suggest that 5- and 6- substitution of these molecular skeletons renders them far too large to fit opposite natural bases in canonical base pair geometry. Thus, steric repulsion might cause them to occupy the *syn* conformation instead; in this conformation the 5- and 6-positions project into the open major groove where the enzyme has few steric constraints. This would explain the low selectivity and little difference among the analogues seen by Berdis, as all the molecules would present essentially the same steric environment if they were oriented *syn*. Engels, Kuchta, and co-workers studied closely related 5- and 6-substituted benzimidazole nucleotides, and they observed that nearly all of them gave pairing selectivity for guanine.[12] Once again this could be readily explained if those molecules flipped to the *syn* conformation in the active site, driven by the expected large steric clash of the 5- and 6-substituents with the template base. Importantly, the 9-nitrogen of benzimidazole, if in *syn* orientation, would be analogous to N3 of cytosine. Thus in this alternative conformation, one would expect the best steric complementarity (and also H-bonding complementarity) to guanine, which was the observed replication result. The energetic cost of flipping to the *syn* conformation might help explain the generally low activity of the benzimidazole analogues that were studied.

In conclusion, we have shown that *E. coli* DNA polymerase I (KF *exo*$^-$) can distinguish easily between nucleobases that have the same sizes but different shapes. This enzyme's sensitivity to shape is surpisingly high, with subangstrom changes being readily distinguished. Furthermore, we have shown that the steric requirements of the 2-, 3-, and 4-positions of thymine are all different. The 2- and 4-position steric protrusions play crucial roles in defining the adenine-encoding behavior of thymine, whereas steric substitutions at the 3-position eliminate thymine-like behavior. Taken together, our results suggest that nucleobase shape plays a more prominent role in base-pairing efficiency and selectivity than other factors in this enzyme. The results add insight into the basic mechanisms of DNA replication, and into the origins of mutations that arise during this process. More work is needed, however, to evaluate how other enzymes respond to differences in nucleobase shape and size.

Received: December 2, 2005

Keywords: base pairing · DNA replication · enzymes · steric effects

[1] a) J. F. Davidson, H. H. Guo, L. A. Loeb, *Mutat. Res.* **2002**, *509*, 17–21; b) A. Sarasin, *Mutat. Res.* **2003**, *544*, 99–106.
[2] L. A. Loeb, T. A. Kunkel, *Annu. Rev. Biochem.* **1982**, *51*, 429–457.
[3] a) M. F. Goodman, *Annu. Rev. Biochem.* **2002**, *71*, 17–50; b) L. Stojic, R. Brun, J. Jiricny, *DNA Repair* **2004**, *3*, 1091–1101; c) A. Sancar, L. A. Lindsey-Boltz, K. Unsal-Kacmaz, S. Linn, *Annu. Rev. Biochem.* **2004**, *73*, 39–85; d) C. H. McGowan, *Mutat. Res.* **2003**, *27*, 75–84; e) M. Christmann, M. T. Tomicic, W. P. Roos, B. Kaina, *Toxicology* **2003**, *15*, 3–34.
[4] a) H. Echols, M. F. Goodman, *Annu. Rev. Biochem.* **1991**, *60*, 477–511; b) T. A. Kunkel, K. Bebenek, *Annu. Rev. Biochem.* **2000**, *69*, 497–529; c) E. T. Kool, *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 1–22; d) W. A. Beard, S. H. Wilson, *Structure* **2003**, *11*, 489–496.
[5] a) B. A. Schweitzer, E. T. Kool, *J. Org. Chem.* **1994**, *59*, 7238–7242; b) N. C. Chaudhuri, R. X.-F. Ren, E. T. Kool, *Synlett* **1997**, 341–347.
[6] J. C. Morales, E. T. Kool, *Nat. Struct. Biol.* **1998**, *5*, 950–954.
[7] T. J. Matray, E. T. Kool, *Nature* **1999**, *399*, 704–708.

[8] T. Mitsui, M. Kimoto, A. Sato, S. Yokoyama, I. Hirao, *Bioorg. Med. Chem. Lett.* **2003**, *13*, 4515 – 4518.
[9] a) A. K. Ogawa, Y. Q. Wu, M. Berger, P. G. Schultz, F. E. Romesberg, *J. Am. Chem. Soc.* **2000**, *122*, 8803 – 8804; b) S. Matsuda, F. E. Romesberg, *J. Am. Chem. Soc.* **2004**, *126*, 14419 – 14427.
[10] M. Strerath, J. Cramer, T. Restle, A. Marx, *J. Am. Chem. Soc.* **2002**, *124*, 2056 – 2057.
[11] X. Zhang, I. Lee, A. Berdis, *J. Biochem.* **2005**, *44*, 13101 – 13110.
[12] K. Kincaid, J. Beckman, A. Zivkovic, R. L. Halcomb, J. W. Engels, R. D. Kuchta, *Nucleic Acids Res.* **2005**, *33*, 2620 – 2628.
[13] a) T. Kim, E. T. Kool, *Org. Lett.* **2004**, *6*, 3949 – 3952; b) T. Kim, E. T. Kool, *J. Org. Chem.* **2005**, *70*, 2048 – 2053; c) T. Kim, J. C. Delaney, J. M. Essigmann, E. T. Kool, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 15803 – 15808.
[14] a) S. Moran, R. X.-F. Ren, S. Rumney, E. T. Kool, *J. Am. Chem. Soc.* **1997**, *119*, 2056 – 2057; b) S. Moran, R. X.-F. Ren, E. T. Kool, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 10506 – 10511.
[15] M. L. Hamm, S. Rajguru, A. M. Downs, R. Cholera, *J. Am. Chem. Soc.* **2005**, *127*, 12220 – 12221.
[16] U. Wichai, S. A. Woski, *Org. Lett.* **1999**, *1*, 1173.
[17] a) M. Yoshikawa, T. Kato, T. Takenishi, *Tetrahedron Lett.* **1967**, *8*, 5065 – 5068; b) T. Kovács, L. Ötvös, *Tetrahedron Lett.* **1988**, *29*, 4525 – 4528.
[18] E. T. Kool, J. C. Morales, K. M. Guckian, *Angew. Chem.* **2000**, *112*, 1046 – 1068; *Angew. Chem. Int. Ed.* **2000**, *39*, 990 – 1009.
[19] Z. Sun, D. Chen, T. Lan, L. W. McLaughlin, *Biopolymers* **2002**, *65*, 211 – 217.
[20] J. Parsch, J. W. Engels, *Nucleosides Nucleotides Nucleic Acids* **2001**, *20*, 815 – 818.
[21] a) K. M. Guckian, R. X.-F. Ren, N. C. Chaudhuri, D. C. Tahmassebi, E. T. Kool, *J. Am. Chem. Soc.* **2000**, *122*, 2213 – 2222; b) S. Bommarito, N. Peyret, J. SantaLucia, *Nucleic Acids Res.* **2000**, *28*, 1929 – 1934.
[22] E. T. Kool, *Biopolymers* **1998**, *48*, 3 – 17.
[23] E. T. Kool, *Annu. Rev. Biochem.* **2002**, *71*, 191 – 219.
[24] A. E. Eriksson, W. A. Baase, X. J. Zhang, D. W. Heinz, M. Blaber, E. P. Baldwin, B. W. Matthews, *Science* **1992**, *255*, 178 – 183.
[25] K. Seio, H. Ukawa, K. Shohda, M. J. Sekine, *J. Biomol. Struct. Dyn.* **2005**, *23*, 735 – 746.
[26] a) L. G. Brieba, B. F. Eichman, R. J. Kokoska, S. Doublie, T. A. Kunkel, T. Ellenberger, *EMBO J.* **2004**, *23*, 3452 – 3461; b) G. W. Hsu, M. Ober, T. Carell, L. S. Beese, *Nature* **2004**, *431*, 217 – 221; c) S. J. Johnson, L. S. Beese, *Cell* **2004**, *116*, 803 – 816.
[27] M. F. Goodman, S. Creighton, L. B. Bloom, J. Petruska, *Crit. Rev. Biochem. Mol. Biol.* **1993**, *28*, 83 – 126.
[28] The values were obtained from the *CRC Handbook of Chemistry and Physics 81st ed.*, *9–2*, *5* (Ed.: David R. Lide), CRC, Florida, **2000–2001**.

