Structures of Mismatch Replication Errors Observed in a DNA Polymerase

Sean J. Johnson and Lorena S. Beese*
Department of Biochemistry
Duke University Medical Center
Durham, North Carolina 27710

Summary

Accurate DNA replication is essential for genomic stability. One mechanism by which high-fidelity DNA polymerases maintain replication accuracy involves stalling of the polymerase in response to covalent incorporation of mismatched base pairs, thereby favoring subsequent mismatch excision. Some polymerases retain a "short-term memory" of replication errors, responding to mismatches up to four base pairs in from the primer terminus. Here we a present a structural characterization of all 12 possible mismatches captured at the growing primer terminus in the active site of a polymerase. Our observations suggest four mechanisms that lead to mismatch-induced stalling of the polymerase. Furthermore, we have observed the effects of extending a mismatch up to six base pairs from the primer terminus and find that long-range distortions in the DNA transmit the presence of the mismatch back to the enzyme active site, suggesting the structural basis for the short-term memory of replication errors.

Introduction

Uncorrected DNA replication errors result in mutations that are essential for evolutionary processes and may lead to human diseases. DNA replication is catalyzed by high-fidelity polymerases that maintain replication accuracy through two major mechanisms. First, prior to covalent incorporation, the polymerases select for correct, Watson-Crick base pairing, while strongly discriminating against each of the 12 possible mismatched bases (Carroll et al., 1991; Echols and Goodman, 1991). Second, once incorporated, misinserted bases compromise the rate of DNA extension. Such "stalling" alters the balance between extension by the polymerase and excision by mismatch-editing exonucleases located either on the polymerase itself or on separate enzymes (Johnson, 1993; Kunkel and Bebenek, 2000). Analysis of the structural mechanisms by which mismatch incorporation leads to stalling is therefore important for understanding the fidelity of DNA replication.

Extensive kinetic studies of mismatch incorporation and extension have shown that the presence of a newly incorporated mismatch reduces the efficiency of subsequent nucleotide insertions and extensions by a hundred- to a million-fold (Echols and Goodman, 1991; Goodman et al., 1993; Kunkel and Bebenek, 2000). The magnitude of this effect depends on the identities of both the mismatch and the polymerase. Mismatch-induced stalling is not limited to the point of incorporation. For instance, mismatches located up to four base pairs away from the primer terminus show a significantly increased partitioning of the growing primer terminus between the polymerase and 3′-5′ exonuclease active sites in *E. coli* Klenow Fragment (KF) (Carver et al., 1994). KF polymerase activity is inhibited when DNA mismatches or lesions are placed at the *n-1, n-2,* or *n-3* positions (*n* is the point of new base pair incorporation; *n-1*, the double-stranded primer terminus), but full activity is recovered at the *n-5* position (Miller and Grollman, 1997). In HIV-reverse transcriptase, stalling of the polymerase reaction is also observed when mismatches are placed in the *n-1* to *n-3* positions (Johnson, 1993). Inhibition of activity by mismatches and damaged DNA is observed in the *n-1* to *n-4* positions in pol α (Ng et al., 1989; Weiss and Fisher, 1992). Extension of mismatches therefore retains a kinetically observable "short-term memory" of the misincorporation event that promotes subsequent error detection and correction. Mismatch-induced stalling is presumably the consequence of disruption of the structural mechanisms by which heteroduplex DNA interacts with the polymerase that has been the subject of much speculation in the absence of structural information (Echols and Goodman, 1991; Kunkel and Bebenek, 2000).

Structural studies of DNA polymerase complexes (reviewed in Brautigam and Steitz, 1998; Doublié et al., 1999; Friedberg et al., 2001; Joyce and Steitz, 1994; Steitz, 1999) in combination with extensive enzyme kinetic studies (Beckman and Loeb, 1993; Eger et al., 1991; Goodman et al., 1993; Johnson, 1993; Kuchta et al., 1988; reviewed in Kunkel and Bebenek, 2000) have revealed the dominant mechanistic and structural features that contribute to accurate DNA replication, which are shared in a large part by all polymerases. In the study presented here, we use the thermophilic *Bacillus* DNA polymerase I fragment (BF) as our model system (Johnson et al., 2003; Kiefer et al., 1997, 1998). Processive DNA replication can take place in crystals of this polymerase. BF is a high-fidelity Family A polymerase with extensive sequence and structural homology to the Klenow Fragments of *E. coli* (KF) and *T. aquaticus* (*Taq*) polymerases. The current understanding of the replication mechanism as it pertains to BF can be summarized as follows. Five important sites have been identified in the BF polymerase (Figures 1A and 1B; Johnson et al., 2003): (1) the "insertion site," in which the cognate nucleotide pairs with the template base (*n* position); (2) the "catalytic site" directly adjacent to the insertion site in which the 3′ hydroxyl of the primer strand and the coordination sphere for two Mg^{2+} ions are located, forming the catalytic center; (3) the "preinsertion site," which houses the template base in a step prior to incorporation; (4) the "postinsertion site" in which the growing 3′ end of the duplex DNA is located (*n-1* position); and (5) the "DNA duplex binding region" in which a four base pair duplex DNA segment is bound (*n-2*, …, *n-5* positions). Replication proceeds by threading DNA through these sites: the template base moves from the preinser-

*Correspondence: lsb@biochem.duke.edu

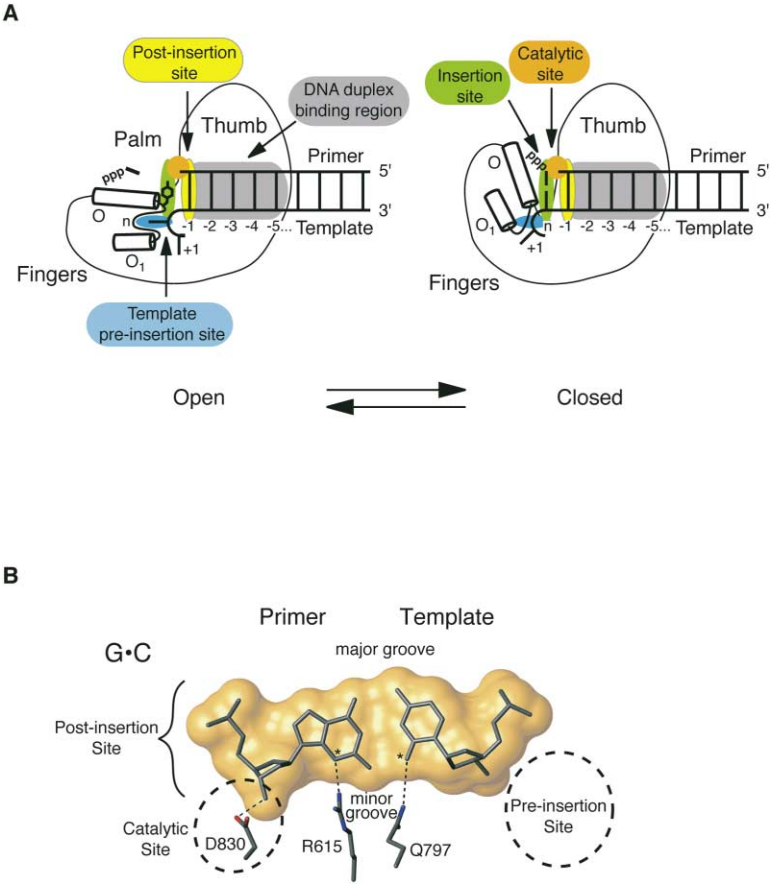


Figure 1. DNA Bound at the Polymerase Active Site

(A) Five regions that contribute to high-fidelity DNA replication are shaded: preinsertion (blue), insertion (green), catalytic (orange), and postinsertion sites (yellow), and the DNA duplex binding region (gray). "Fingers," "palm," and "thumb" subdomains are indicated. Transfer of the *n* template base from the template preinsertion site to the insertion site is accompanied by a transition from the open to a closed conformation, mediated by a hinge-bending motion centered around the O and O1 helices. At the insertion site, a complementary dNTP (indicated by PPP−) is selected prior to catalysis. The insertion site is occluded in the open conformation by a conserved tyrosine (Tyr714 in BF) located at the C terminus of the O helix that stacks with the DNA template. In the closed conformation, the tyrosine is displaced, allowing the new base pair to bind. Correct base pairing is detected by interactions at the postinsertion site and along the entire DNA duplex binding region.

(B) Cognate G•C base pair bound at the postinsertion (*n-1*) site. The molecular van der Waals surface of the base pair is shown in yellow. A hydrogen bond between Asp830 and the 3′ primer terminus is observed in the open conformation. Arg615 and Gln797 interact with hydrogen bond acceptor atom (*) located in the DNA minor groove, providing minor groove readout.

tion site to the insertion site and pairs with an incoming base; covalent incorporation takes place; the newly synthesized base moves into the postinsertion site; the DNA in the duplex binding region translocates by one base pair, releasing a base pair from the *n-5* position; the next template base moves into the preinsertion site.

Here we present the first high-resolution structures of mismatches captured at the active site of a DNA polymerase, by taking advantage of the catalytic activity of BF crystals (Johnson et al., 2003; Kiefer et al., 1998). All 12 possible covalently incorporated mismatches were obtained, positioned at the primer terminus, poised for subsequent extension reactions. Each of these mismatches disrupts the active site of the DNA polymerase in various ways that are the consequence of the specific interactions between the mismatch and the polymerase. Although all 12 DNA heteroduplexes are structurally distinct, we have been able to classify the mismatch-induced disruptions into four broad categories based on the nature of the changes in the active site. Additionally, we have been able to translocate several mismatches through the duplex binding region by using successive rounds of DNA replication in the crystal to extend the heteroduplex, which provide new insight into the mechanism for misincorporation memory. Together, these results provide a firm basis for understanding these early events in polymerase-mediated incorporation or evasion of mutations.

Results

Crystallographic Capture of Mismatches

Crystals with mismatches captured at the BF polymerase active site were obtained either by catalysis in the crystals, using mutagenic reaction conditions, or by cocrystallization of DNA duplexes that contain a mismatch at the primer terminus (Tables 1–3). In the presence of Mg^{2+}, accurate DNA replication is obtained in BF•DNA crystals, and synthesis stalls in preference to incorporation of DNA mismatches (Johnson et al., 2003). In the presence of Mn^{2+}, the specificity of the enzyme is relaxed, allowing enzymatic incorporation of mismatches. For those mismatches that were captured both by catalytic incorporation and by cocrystallization of preformed duplexes, both structures adopt the same conformation.

In solution, the efficiency of incorporation varies depending on the identity of the mismatch (Kunkel and Bebenek, 2000). DNA replication catalyzed in the BF•DNA crystals also exhibits differential mismatch incorporation and extension efficiencies: some mismatches are readily incorporated in the presence of Mn^{2+}, while others were captured only by cocrystallization. Once incorporated, the ease of mismatch extension catalyzed in the crystal parallels that reported for solution studies reported for other, related polymerases. Previous studies have demonstrated that accurate replication of Watson-Crick base pairs is faithfully repro-

Table 1. DNA Mismatches Captured at the BF Polymerase Active Site

Mismatch *(primer•template)*	Sequence[a]	Method[b]	Mispair Location	Mispair Conformation	Active Site Regions Disrupted by Mispair Binding[c]	Extension Products
Purine-Pyrimidine						
G•T	3′- GGCGACTAGCG 5′- ACG**T**CGCTGATCGCA	E,C	post-insertion site	wobble	Pre-IS, IS, Post-IS, DBR	*n-2,n-3,n-4,n-6*
T•G	3′- TTCGTAGTACG 5′- CCCG**A**GCATCATGCA	E	post-insertion site	wobble, water-stabilized	Pre-IS, IS, CS, Post-IS, DBR	no extension
A•C	3′- ACGACTAGCG 5′- ACGT**C**GCTGATCGCA	C	post-insertion site	disordered template	Pre-IS, IS, Post-IS, DBR	*n-2*
C•A	3′- CCGTAGTACG 5′- CCCG**A**GCATCATGCA	C	undetermined	disordered	Pre-IS, IS, Post-IS, DBR	not tested
Pyrimidine-Pyrimidine						
T•T	3′- TTGCACTAGCG 5′- GACG**T**ACGTGATCGCA	E,C	post-insertion site	wobble	IS, CS, Post-IS	no extension
C•T	3′- CCGACTAGCG 5′- GTACG**T**GCTGATCGCA	C	post-insertion site	open, water-stabilized	IS, CS, Post-IS	*n-2*
T•C	3′- TCGACTAGCG 5′- ACGT**C**GCTGATCGCA	C	post-insertion site	disordered template	Pre-IS, IS, Post-IS, DBR	position unclear (disordered)
C•C	3′- CCTAGTACG 5′- GATCG**C**GATCATGCA	E,C	insertion/pre-insertion sites	frayed	Pre-IS, IS, CS, Post-IS, DBR	no extension
Purine-Purine						
A•G	3′- AACGACTAGCG 5′- GTACG**T**GCTGATCGC	C	post-insertion site	*anti-anti*	Pre-IS, IS, CS, Post-IS, DBR	no extension
G•G	3′- GGACTAGCG 5′- GTACGT**G**CTGATCGCA	C	post-insertion site	*syn-anti*	Pre-IS, IS, Post-IS, DBR	*n-2*
A•A	3′- AATAGTTACGGTACG 5′- GATATCAATGCCATGC	E	insertion/pre-insertion sites	frayed	IS, CS	not tested
G•A	3′- GCGTACTACG 5′- CCCG**A**GCATGATGCA	C	insertion/pre-insertion sites	frayed	Pre-IS, IS, CS	not tested

[a] The mispair is shown in bold letters. Underlined bases indicate nucleotides that were incorporated into the BF-DNA cocrystals in the presence of Mn^{2+}.

[b] Denotes whether the mispair was captured by enzymatic incorporation in the crystal (E), cocrystallization (C), or both methods (E,C).

[c] Active site regions are abbreviated: preinsertion site (Pre-IS); insertion site (IS); catalytic site (CS); postinsertion site (Post-IS); DNA duplex binding region (DBR).

duced in BF crystals (Johnson et al., 2003). The mismatch replication process as observed in the crystal also appears to be a reasonable approximation of that observed in solution.

We have examined all 12 possible covalently incorporated DNA mismatches (Table 1). Six are placed at the postinsertion site and are well ordered (Figure 2) (G•T, T•G, T•T, C•T, A•G, G•G; the first and second letters refer to the primer and template strands, respectively); three mismatches are placed at the postinsertion site but are too disordered for interpretation as a unique molecular structure (T•C, A•C, C•A); and three (A•A, C•C, G•A) are located at the insertion site (primer strand) and preinsertion site (template strand), instead of the postinsertion site, and do not pair (i.e., are frayed).

The structure of each mismatch is distinct (Table 3). Although some of the mismatch structures adopt conformations that had been previously observed in heteroduplex DNA in the absence of protein (Kennard and Hunter, 1989; Patel et al., 1987), others do not. The polymerase, therefore, plays a significant role in defining the local and the global conformation of each mismatch structure. The type and degree of disruptions at the polymerase active site vary depending on the identity of the mismatch. We present our observations categorized by the type of base pairing.

Purine-Pyrimidine Mismatch: G•T, T•G

The G•T mismatch was obtained by enzymatic extension in the crystal in the presence of Mn^{2+} and is bound at the postinsertion site. The protein adopts a distorted open conformation (Figure 3). The newly incorporated, mismatched primer base (G) adopts a conformation similar to that of a cognate base at that position. Consequently, positioning of the 3′ hydroxyl and the assembly of the catalytic site remains essentially intact. In contrast, the positioning of the template base is significantly distorted (Figure 2). The two bases are paired in a wobble conformation (shear, Table 3), in which the thymine is positioned toward the DNA major groove. The minor groove interaction with Gln797 is lost and the template strand is displaced by as much as 3.6 Å from the surface of the protein, relative to the cognate binding site. Additionally, instead of stacking with the template strand, Tyr714 shifts by up to 3 Å and stacks with the primer base of the mismatch. This shift is accomplished by a hinge motion in the O, O1, and O2 helices in the finger domain, which results in a dramatic rearrangement of the loop between the O and O1 helices containing the preinsertion site; the protein backbone at position 716 is displaced by 6 Å, thereby preventing binding of the acceptor template base at the preinsertion site. Placement of a mismatch at the postinsertion site also alters

Table 2. Data Collection and Refinement Statistics for Mismatch Structures

	G•T	T•G	T•T	C•T	A•G	G•G	A•A	C•C	G•A
Data Collection (all data)									
Resolution, Å	50.0–1.9	30.0–1.65	50.0–2.0	50.0–2.0	50.0–1.7	30.0–1.6	35.0–2.1	50.0–2.1	50.0–1.9
Outer shell, Å	1.97–1.90	1.71–1.65	2.07–2.00	2.07–2.00	1.76–1.70	1.66–1.60	2.18–2.10	2.18–2.10	1.97–1.90
No. reflections									
Unique	63,846	100,253	58,937	59,392	94,629	113,762	50,872	51,291	63,331
Total	311,125	603,440	259,656	320,048	292,563	664,723	491,438	245,582	396,272
Mean I/σ_I[a]	23.4 (3.8)	22.1 (1.7)	20.6 (3.1)	16.9 (2.7)	20.2 (3.3)	19.3 (1.9)	17.5 (2.8)	15.5 (2.6)	21.9 (3.2)
Completeness, %	92.5	95.2	99.4	99.8	98.2	98.4	99.6	99.7	91.3
R_{sym}[a]	5.9 (36.6)	7.1 (46.8)	7.6 (56.4)	8.9 (41.4)	4.7 (26.0)	6.4 (66.7)	13.6 (55.8)	10.2 (53.1)	8.0 (42.6)
Refinement									
Completeness, %[a]	90.9 (56.3)	90.9 (58.2)	95.7 (90.0)	96.7 (88.9)	96.0 (86.3)	90.3 (62.4)	94.6 (79.1)	95.9 (88.4)	89.4 (51.4)
R_{free}, %[a]	24.3 (33.0)	23.9 (36.1)	25.0 (36.1)	25.4 (32.1)	21.3 (24.7)	25.4 (38.2)	22.8 (27.3)	25.9 (29.4)	23.6 (30.4)
R_{cryst}, %[a]	21.3 (27.1)	21.8 (33.4)	21.7 (30.7)	22.4 (28.9)	19.3 (24.6)	22.8 (35.1)	20.0 (27.3)	22.1 (27.5)	20.6 (27.9)
Nonhydrogen atoms									
Total	5610	5689	5593	5530	5829	5590	5758	5382	5632
Solvent	423	569	386	368	658	485	500	306	510
Rmsd from ideal geometry									
Bond lengths, Å	0.006	0.005	0.006	0.006	0.008	0.005	0.005	0.006	0.005
Bond angles, °	1.2	1.2	1.2	1.1	1.2	1.2	1.1	1.2	1.1
Average isotropic B value, Å^2	30.6	23.4	30.6	30.3	20.8	27.1	31.6	35.3	25.7
PDB accession code	1NJW	1NJX	1NJY	1NJZ	1NK0	1NK4	1NK5	1NK6	1NK7

$R_{sym} = (\Sigma|(I - \langle I \rangle)|)/(\Sigma I)$, where $\langle I \rangle$ is the average intensity of multiple measurements. R_{cryst} and $R_{free} = (\Sigma||F_{obs}| - |F_{calc}||)/(\Sigma|F_{obs}|)$. R_{free} was calculated over 5% of the amplitudes not used in refinement. Rms deviations reported include both protein and DNA residues.

[a]Values in parentheses correspond to those in the outer resolution shell.

Table 3. DNA Base Pair Parameters at the Postinsertion Site of BF

Base Pair	$d_{C1'\text{-}C1'}$ (Å)	λ_{primer} (°)	$\lambda_{template}$ (°)	Shear (Å)	Stretch (Å)	Stagger (Å)	Buckle (°)	Propeller (°)	Opening (°)
Watson-Crick[a]	10.3 (0.2)	57.6 (1.8)	57.6 (1.1)	0.04 (0.20)	−0.09 (0.08)	−0.37 (0.26)	23.0 (5.2)	−6.4 (2.7)	9.6 (3.0)
G•T	10.4	46.2	62.2	−1.59	−0.38	−0.06	25.1	1.5	1.7
T•G	10.8	59.7	40.9	1.67	−0.31	−0.66	22.9	1.8	−3.6
T•T	9.1	72.1	42.0	2.60	−1.64	0.21	19.9	−17.2	−8.4
C•T	8.7	60.8	55.4	0.46	−1.55	−0.28	25.4	−7.3	−7.7
A•G	12.3	48.6	50.5	0.09	1.58	−0.18	37.9	−16.2	7.7
G•G	11.0	37.0	60.2	−1.10	−3.72	0.11	13.5	−23.0	27.6

$d_{C1'\text{-}C1'}$ is the distance between the C1′ atoms of the base pair. λ is the angle between the glycosidic bond of the primer or template base and the line drawn between the C1′ atoms. All other parameters are defined in Dickerson et al. (1989). All values were calculated using 3DNA (Lu et al., 2000). The CEHS reference frame was used to calculate the opening parameter.

[a] Average values for all four Watson-Crick base pairs observed at the *n-1* position of BF polymerase (Johnson et al., 2003). Standard deviations are shown in parenthesis.

the geometry of the adjacent insertion site, which may alter the base pairing specificity at that site. Distortions extend into the DNA duplex binding region as well, affecting base pairs as far away as the *n-3* base position, by changing the conformation of the DNA along the duplex binding region from the normally observed A form to a B form conformation.

The T•G mismatch was captured in the polymerase crystal by enzymatic extension and also adopts a wobble conformation (shear, Table 3), but an additional water-mediated hydrogen bond is observed between the mismatched bases in DNA minor groove (Figure 2). As with the G•T structure, the DNA template is displaced from its normal binding site and the polymerase adopts a distorted open conformation. However, some differences are observed in binding of the primer strand. In the T•G structure, the minor groove interaction between the primer base and Arg615 in the postinsertion site is lost, and Arg615 drops to the floor of the polymerase into an apo-like conformation. In addition, some disorder is observed in the position of the deoxyribose ring of the thymine nucleotide, suggesting that the primer terminus is predominantly displaced, although the normal interaction between the primer 3′ hydroxyl and Asp830 in the catalytic site may be retained to some degree.

Pyrimidine-Pyrimidine Mismatch: T•T, C•T

The T•T mismatch was obtained by catalysis in the crystal. The mismatch is bound at the postinsertion site. In contrast to the G•T and T•G mismatches, the distortions in the T•T mismatch are confined primarily to the primer strand, and the protein conformation is undisturbed (Figures 2 and 3). The T•T mismatch adopts a wobble conformation (shear, Table 3) in which the primer base lifts up into the DNA major groove and the template base rotates slightly toward the minor groove, similar to the wobble conformation reported in an analogous U•U RNA heteroduplex structure obtained in the absence of protein (Baeyens et al., 1995). The displacement of the primer base is accompanied by a ~1.4 Å shift of the 3′ hydroxyl, thereby breaking the hydrogen bond that normally exists between the primer 3′ hydroxyl and Asp830, thus disrupting assembly of the catalytic site (Figure 3).

The C•T mismatch is also bound in the postinsertion site. This mismatch does not adopt a wobble conformation, but rather the bases pair directly opposite each other (Figure 2). The mismatch shows significant opening in the minor groove (Table 3) where the two opposing O2 atoms interact via a bridging water molecule. As with the T•T structure, the primer terminus is displaced, but the DNA template and protein conformation is essentially undisturbed.

Purine-Purine Mismatch: A•G, G•G

The A•G mismatch is bound at the postinsertion site, with the polymerase adopting a distorted open conformation (Figure 3). Both bases in the mismatch maintain an *anti* conformation about the glycosidic bond (Figure 2), which is one of the three conformations that have been observed in heteroduplex DNA in the absence of protein (Kennard and Hunter, 1989). Consequently, accommodation of the two large purine bases results in a 2.0 Å increase of the helical width (Table 3) and causes extensive disruptions and movement of both template and primer strands (Figure 3). Similar to the G•T mismatch, the template strand lifts away from the surface of the protein, the finger domain rearranges to block the preinsertion site, and the duplex binding region is distorted. At the postinsertion site, the disrupted minor groove contact by Gln797 is replaced with water-mediated interactions to both the N2 and N3 positions of the template guanine base. Along the primer strand, the minor groove interaction between Arg615 and the primer base is retained. However, the widening of the base pair as well as a change in sugar pucker displaces the 3′ hydroxyl by 2.6 Å, thereby significantly altering the interaction with Asp830 at the catalytic site, but without breaking it.

Not all pur•pur mismatches retain an *anti-anti* conformation in the polymerase active site. In the G•G mismatch, the primer base of the mismatch rotates 180° with respect to the sugar moiety into a *syn* conformation while the template base remains in an *anti* conformation (Figure 2). This rearrangement, which also has been observed in heteroduplex DNA in the absence of protein (Skelly et al., 1993), allows the mismatch to form two hydrogen bond interactions and results in a helical width that is much closer to a Watson-Crick base pair than is observed in the *anti-anti* A•G structure (Table 3). Consequently, disruptions to the active site are more limited than in the A•G structure and are confined to the template strand. In this regard, the G•G structure more closely resembles the G•T complex. The template

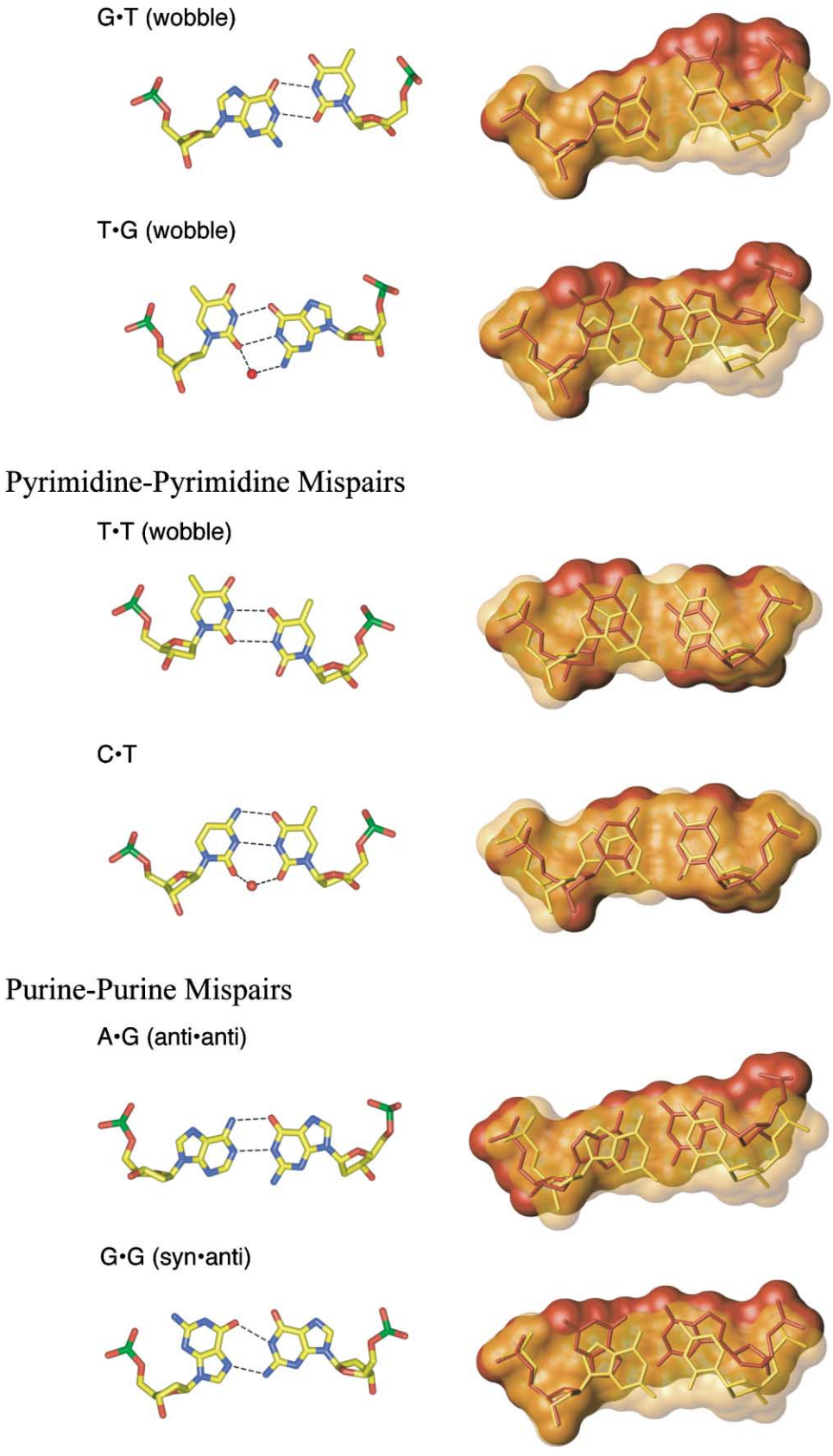
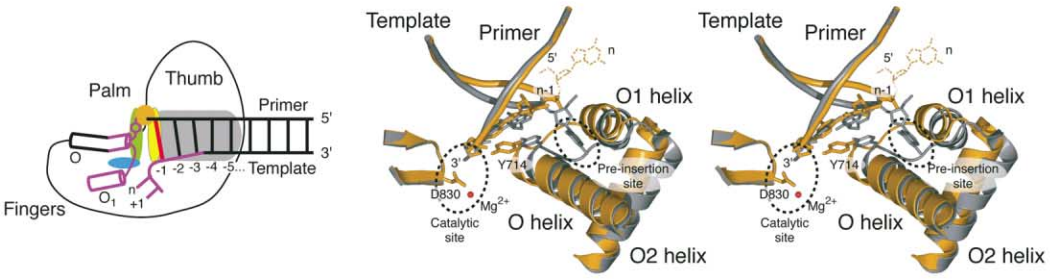


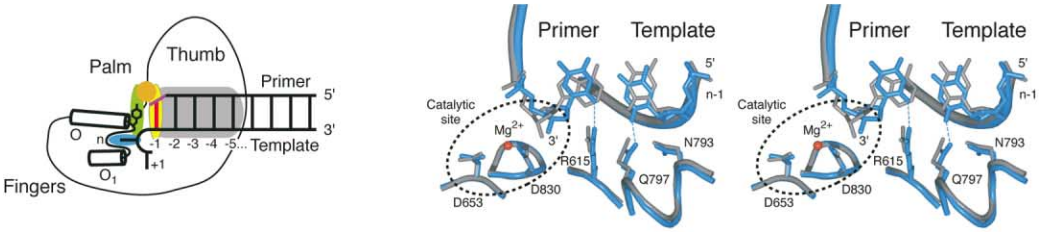
Figure 2. DNA Mismatches Bound at the Polymerase Postinsertion Site

The bases are shown in the same orientation and location as the G•C base pair in Figure 1B. Left, hydrogen bonding pattern. Right, superimposition of the molecular surface of the mismatch (red) and cognate G•C base pair (yellow, PDB ID 1L3S) bound at the postinsertion site, highlighting differences in shape and location of the primer terminus.

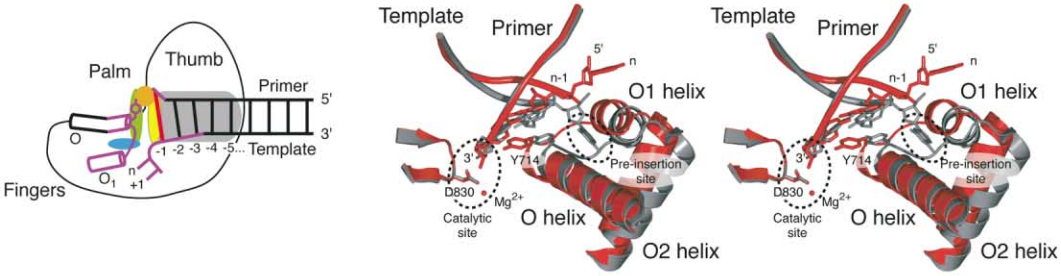
1. Disruption of template strand and pre-insertion site (**G•T**, G•G, A•C, T•C)



2. Disruption of primer strand and assembly of catalytic site (**T•T**, C•T)



3. Disruption of template and primer strands (**A•G**, T•G)



4. Fraying of DNA at insertion site (**A•A**, G•A, C•C)

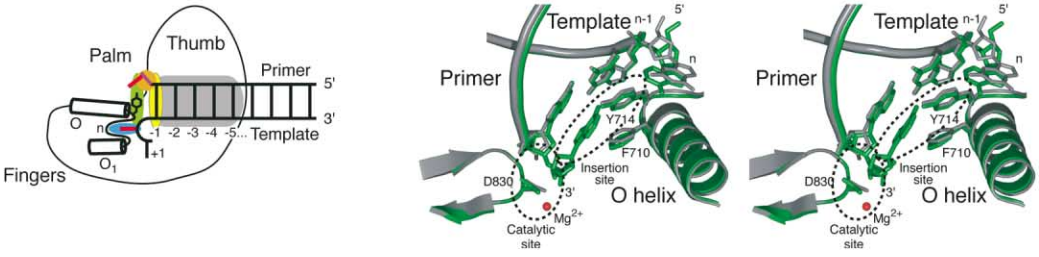


Figure 3. Mismatch-Induced Disruptions at the Polymerase Active Site

Four categories of active site disruptions are observed in the presence of a replication error. For each category, a representative mismatch structure (bold text) is illustrated. Right, schematic of the polymerase•mismatch active site (color coded as in Figure 1A) indicating the regions (magenta) disrupted by the mismatch (red line). Left, mismatch complex (color) superimposed on a cognate A•T base pair structure (gray, PDB ID 1L3U).

Mechanism 1. A G•T mismatch bound at the postinsertion site results in displacement of the template strand, repositioning of Tyr714, and a rearrangement of the O, O1, and O2 helices that blocks the template preinsertion site, resulting in a distorted open conformation that differs from that seen in the presence of a cognate base pair. The DNA helix adopts a B form conformation at the active site rather than the more A form conformation observed in cognate base pairs. The catalytic site is undisturbed.

Mechanism 2. In the T•T mismatch structure, the primer terminus is displaced, and the interaction between residue Asp830 and the primer 3′ hydroxyl at the catalytic site is disrupted. The template base of the mismatch rotates toward the floor of the polymerase active site, but the DNA backbone of the template strand is undisturbed.

Mechanism 3. The A•G mismatch structure induces displacement of the template strand and blocking of the preinsertion site (similar to mechanism 1) and also displacement of the primer strand, resulting in disruption of the catalytic site (similar to mechanism 2).

Mechanism 4. The A•A mismatch structure is frayed, with the primer base of the mismatch bound at the insertion site and the template base at the preinsertion site.

strand is displaced and the preinsertion site is blocked, but the position of the primer terminus is undisturbed. However, rotation of the primer guanine base to a *syn* conformation removes the N3 hydrogen bond acceptor atom from the minor groove and prevents Arg615 from coordinating the nucleotide. This lost interaction may have particularly disruptive consequences in the closed conformation where Arg615 plays a dual role in anchoring both the primer base at the postinsertion site and the incoming dNTP at the insertion site (Johnson et al., 2003).

Disordered Mismatches: T•C, A•C, C•A

T•C and A•C mismatches were also captured at the postinsertion site, but the mismatches are sufficiently disordered to preclude deduction of a detailed molecular model for the DNA. The disorder is confined primarily to the template strand, extending to the *n-3* position of the duplex binding region. A mixture of A and B forms of DNA can be discerned in the duplex binding region. No density for the template strand is observed in either pre- or postinsertion site. In both cases, the protein adopts a distorted open conformation in which the preinsertion site is blocked. Disorder on the primer strand is essentially localized to the primer base. The C•A mismatch structure is too disordered throughout the DNA to permit assignment of the mismatch location.

Frayed Mismatches: A•A, C•C, G•A

In contrast to the other mismatches, these three mismatches do not bind in the postinsertion site. Instead, the primer base of each mismatch is located at the insertion site and the template base is located at the preinsertion site or displaced completely from the active site. The bases, therefore, do not pair with each other, and the 3′ primer terminus has a "frayed" appearance.

In the A•A mismatch, obtained by catalysis in the crystal in the presence of Mn^{2+}, the primer base is located at the insertion site and the template base at the preinsertion site (Figure 3). The two bases cannot pair because Tyr714 blocks access of the template base to the insertion site. Tyr714 also prevents the primer base from stacking adjacent to the DNA helix; the primer base is displaced from the floor of the insertion site and stacks against Phe710. Thus, although the primer base is located at the insertion site, it does not superimpose with the dNTP position observed in a closed BF ternary complex (Johnson et al., 2003). Additionally, since the 3′ primer terminus is located at the insertion site rather than the postinsertion site, the catalytic site is completely disrupted. Density consistent with a single metal ion is observed at the catalytic site (B); the second catalytic metal (A) is not observed.

The G•A and C•C mismatch structures were obtained by cocrystallization with a preformed heteroduplex. Although they differ in detail, the disruption of the catalytic site resembles that observed in the A•A structure.

Mismatch Extensions

Mismatches diminish but do not block the rate of extension. We therefore have been able to probe the mechanism by which processive replication continues upon incorporation of a mismatch by further taking advantage of the catalytic properties of the BF crystals. Extension of a number of mismatches was examined starting with cocrystallized mismatch substrates (Table 1) and directing progression of heteroduplexes through the polymerase in successive extension steps by soaking in different combinations of nucleotides in the presence of Mg^{2+} to control the final location of the mismatch (one of the *n-2*, …, *n-6* positions of the duplex binding region). Under these conditions, the A•G, T•T, T•G, and C•C mismatches failed to extend, whereas the G•T, C•T, and G•G mismatches were all successfully extended (Tables 1, 4, and 5). For the T•C mismatch, the DNA is too disordered to permit interpretation of the product structure. We note that in one case (T•G), we have also successfully extended the mismatch in the crystal under a variety of other reaction conditions (unpublished data).

G•T Extension

The G•T mismatch was captured at the *n-2, n-3*, *n-4*, and *n-6* positions in separate rounds of replication in the presence of different nucleotide mixtures (Figure 4). As we observed at the *n-1* position, the mismatch adopts a wobble conformation at the *n-2* position, although water is bound in both the major and minor grooves. At positions *n-3* and *n-4*, the wobble inverts; the thymine base moves into the minor groove and the guanine base moves toward the major groove. This conformation is stabilized by direct contacts between the polymerase and the DNA backbone and solvent-mediated contacts in the minor groove. At position *n-6*, the G•T mismatch has exited the polymerase and adopts a wobble conformation resembling the crystal structure of a B form G•T heteroduplex obtained in the absence of protein (Hunter et al., 1987).

The G•T base pairing observed at the *n-3* and *n-4* positions is inconsistent with the interbase hydrogen bonding geometry associated with the major tautomeric form of the nucleotides. This suggests that a tautomeric shift, or ionization of the mismatch, has taken place. It is not possible to deduce the position of the hydrogens in these bases, however, since there are several, structurally similar possibilities, and the 2.0 Å resolution of the X-ray structures does not permit clear distinction between the alternatives.

Mismatches at the *n-2* and *n-3* locations disrupt the normal A to B form conformational transition in this region. This disruption results in partial release of the DNA template, repositioning of Tyr714, and closing of the template preinsertion site, similar to the disruptions observed when the G•T mismatch is positioned at the primer terminus. Consequently, the presence of this mismatch within the duplex binding region continues to disrupt to the active site. Positioning of the mismatch in the *n-4* location partially restores the normal DNA structure in the duplex binding region, and a mixture of both a disrupted and an undisrupted active site is observed. Positioning at *n-6* fully restores the DNA conformation at all other sites on the polymerase (the 3′ primer terminus in this complex has a blunt end, so there is no template base to occupy the preinsertion site).

C•T Extension

The C•T mismatch was extended in a single round of replication, placing it in the *n-2* location of the duplex

Table 4. Data Collection and Refinement Statistics for G•T and C•T Extension Experiments

	G•T (*n-2*)	G•T (*n-3*)	G•T (*n-4*)	G•T (*n-6*)	C•T (*n-2*)
Data Collection (All Data)					
Resolution, Å	50.0–1.9	50.0–1.9	50.0–2.0	50.0–1.8	50.0–1.8
Outer shell, Å	1.97–1.90	1.97–1.90	2.07–2.00	1.86–1.80	1.86–1.80
No. reflections					
Unique	64,853	60,333	52,699	62,356	80,698
Total	276,640	283,443	185,081	209,950	446,710
Mean I/σ_I[a]	18.1 (2.8)	28.7 (3.5)	18.9 (2.0)	31.1 (4.6)	21.9 (2.5)
Completeness, %	94.2	87.1	88.3	76.7	97.8
R_{sym}*	6.6 (25.7)	4.7 (21.7)	6.9 (32.6)	4.2 (13.6)	6.4 (47.4)
Refinement					
Completeness, %[a]	91.6 (57.8)	86.3 (42.0)	86.6 (47.9)	76.5 (16.8)	93.8 (81.7)
R_{free}, %[a]	21.9 (26.4)	23.7 (27.7)	24.3 (25.5)	21.8 (26.5)	24.8 (27.2)
R_{cryst}, %[a]	19.0 (23.6)	20.4 (24.6)	20.5 (23.1)	18.6 (23.0)	21.6 (25.8)
Nonhydrogen atoms					
Total	5732	5622	5595	5912	5655
Solvent	532	423	352	613	513
Rmsd from ideal geometry					
Bond lengths, Å	0.005	0.005	0.006	0.005	0.006
Bond angles, °	1.2	1.2	1.2	1.1	1.2
Average isotropic B value, Å^2	23.9	30.3	30.8	20.6	29.1
PDB accession code	1NK8	1NK9	1NKB	1NKC	1NKE

$R_{sym} = (\Sigma|(I - \langle I \rangle)|)/(\Sigma I)$, where $\langle I \rangle$ is the average intensity of multiple measurements.
R_{cryst} and $R_{free} = (\Sigma|F_{obs} - F_{calc}|)/(\Sigma|F_{obs}|)$. R_{free} was calculated over 5% of the amplitudes not used in refinement.
Rms deviations reported include both protein and DNA residues.
[a]Values in parentheses correspond to those in the outer resolution shell.

binding region. Although positioning of the C•T mismatch at the postinsertion site disrupts the catalytic site, translocation of the C•T mismatch to the *n-2* position fully restores the catalytic site. The mismatch at the *n-2* position shows a 2 Å increase in the separation between the mismatched bases (stretch) as compared to the original structure (Table 5). Consequently, the DNA adopts a normal helical width, but interbase distance in the mismatch is ~5 Å, preventing direct contact between the two opposing bases (Figure 5). This conformation is stabilized by a bridging water molecule that forms interbase hydrogen bonds, as well as direct contacts between the DNA phosphate backbone and the protein, stacking with the adjacent base pairs, and a water-mediated minor groove interaction between the protein and the minor groove base—interactions that are all present in normal Watson-Crick structures. No distortion is observed in the DNA helix outside the mismatch or in the conformation of the active site.

Discussion

The high-resolution structures of covalently incorporated DNA mismatches at the active site of the BF polymerase allow us to suggest mechanisms by which replication errors stall a polymerase, enhancing subsequent dissociation and exonucleolytic excision. Such stalling is the consequence of mismatch-induced disruptions of the polymerase active site. We observe four broad categories of mismatch-induced disruptions at the active site. Furthermore, we find that mismatches need not be positioned at the primer terminus to induce these

Table 5. Base Pair Parameters for Extended G•T and C•T Mismatches

Base Pair	$d_{C1'\text{-}C1'}$ (Å)	λ_{primer} (°)	$\lambda_{template}$ (°)	Shear (Å)	Stretch (Å)	Stagger (Å)	Buckle (°)	Propeller (°)	Opening (°)
G•T(*n-1*)	10.4	46.2	62.2	−1.59	−0.38	−0.06	25.1	1.5	−4.8
G•T(*n-2*)	10.3	41.9	69.2	−2.44	−0.57	0.07	19.5	0.7	−1.8
G•T(*n-3*)	10.9	63.7	44.9	1.91	0.00	−0.70	−17.7	−3.6	−2.7
G•T(*n-4*)	11.2	62.9	41.4	2.06	0.03	−0.11	2.9	7.9	−6.4
G•T(*n-6*)	10.5	43.3	66.1	−2.17	−0.43	0.02	−4.4	−21.9	1.2
G•T (B-DNA)*	10.4	42.5	69.9	−2.49	0.64	−0.04	13.4	−8.3	−3.0
G•T (B-DNA)[a]	10.3	44.6	72.0	−2.72	0.67	0.10	9.9	10.2	−2.3
C•T (*n-1*)	8.7	60.8	55.4	0.46	−1.55	−0.28	25.4	−7.3	6.0
C•T (*n-2*)	10.6	60.2	58.3	0.39	0.58	0.25	26.2	−16.3	10.7

$d_{C1'\text{-}C1'}$ is the distance between the C1′ atoms of the base pair. λ is the angle between the glycosidic bond of the primer or template base and the line drawn between the C1′ atoms. All other parameters are defined in Dickerson et al. (1989). All values were calculated using 3DNA (Lu et al., 2000).
[a]Two G•T crystal structures observed in heteroduplex DNA (Hunter et al., 1987).

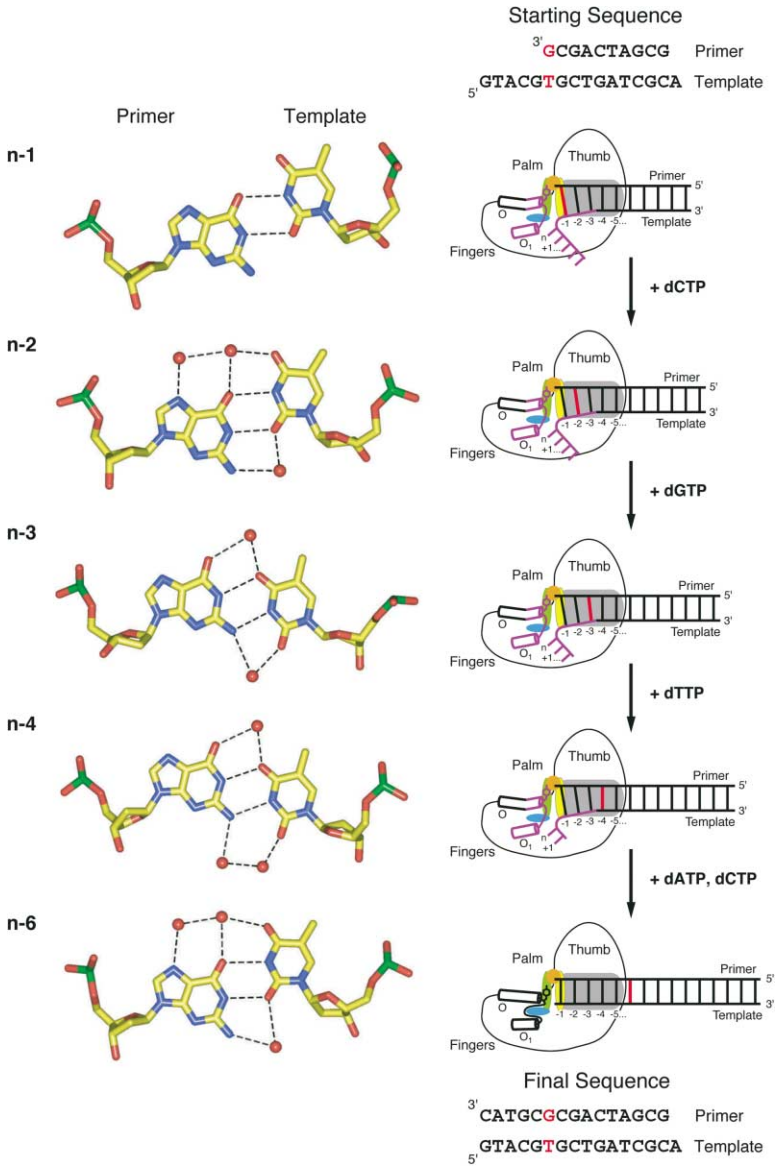


Figure 4. Extension of a G•T Mismatch by Successive Rounds of Replication

The conformation of the G•T mismatch is shown at each position (left), including interacting water molecules (red spheres). Dashed lines indicate potential hydrogen bonds. At the *n-3* and *n-4* positions, hydrogen bonds are shown between groups within the appropriate distance (≤3.2 Å) and correspond to tautomerization or ionization of one of the bases (see text). A schematic representation (right) of the mismatch complex, drawn and color coded as described in Figures 1–3, indicates regions of the polymerase active site that are disrupted upon binding of the mismatch (red line). Mismatch binding at positions *n-1* to *n-4* along the DNA duplex binding region (gray) results in a distorted open conformation at the polymerase active site as described by mechanism 1 (Figure 3). The normal open conformation observed with homoduplexes is fully restored when the mismatch is bound at the *n-6* position.

disruptions. Mismatches positioned up to four base pairs in from the point incorporation can still disrupt the active site, thereby providing a mechanism whereby the polymerase retains a short-term memory of the mismatch incorporation event.

Four Broad Categories of Mismatch-Induced Disruptions at the Active Site

Although each of the 12 mismatches is structurally unique, we find that mismatch-induced disruptions of the polymerization mechanism can be divided into four broad categories (Figure 3): (1), displacement of the template strand and disruption of the preinsertion site; (2), disruption of primer strand and the assembly of the catalytic site; (3), disruption of both the template and primer strands; and (4), fraying of the DNA at the insertion site. A consequence of each of these mechanisms is disruption of the insertion site.

All four mechanisms are observed when a mismatch is positioned at the primer terminus: (1), G•T, G•G, A•C, T•C; (2), T•T, C•T; (3), A•G, T•G; (4), A•A, G•A, C•C. The disruptions induced by the G•T mismatch as it is extended through the duplex binding region (the memory mechanism) are confined to mechanism 1.

Mismatch Extension and the Mechanism of Replication Error Memory

Extension of a number of mismatches through the duplex binding region of a polymerase retains a "memory" of the misincorporation event that results in further stalling, even though the mismatch is located up to four base pairs from the point of incorporation (Carver et al., 1994; Miller and Grollman, 1997). Of particular relevance to understanding mismatch extensions, therefore, are the structural contortions that allow a heteroduplex to translocate through the duplex binding region, which is stereochemically complementary to correctly formed base pairs (Johnson et al., 2003; Li and Waksman, 2001),

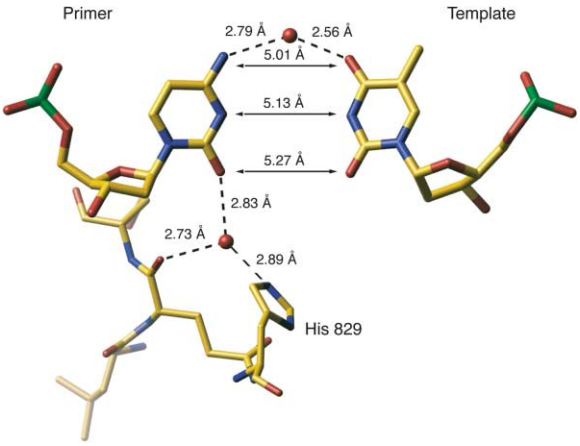


Figure 5. The C•T Mismatch Extension Product

The C•T mismatch located at the *n-2* position provides a dramatic example of DNA conformations that have not been observed in heteroduplex structures determined in the absence of protein. The bases are separated by >5.0 Å and are stabilized by an intervening water molecule, the protein, and DNA stacking interactions.

and the mechanism by which a heteroduplex disrupts the polymerase active site at a distance to provide a short-term memory of the mismatch incorporation event.

In one case (G•T), we have carried out multiple extension reactions (Figure 4) and captured the progression of this mismatch through the entire duplex binding region (*n-1*, …, *n-6*). The structural adaptations to heteroduplex are confined primarily to the DNA, with some local protein side chain motions. There are three striking features of the adaptations in the DNA structure: (1) at each position the conformation of the mismatch is different; (2) at several positions base pair geometry indicates that the tautomeric or ionization state of the bases has changed; and (3) distortions are not localized to the vicinity of the mismatch, but can also affect the DNA and protein conformation at the active site. The long-range distortions are transmitted primarily through the template strand and result in disruption of the single-stranded DNA template and the preinsertion site at the active site (mechanism 1), suggesting a mechanism whereby the polymerase retains a short-term memory of the mismatch after incorporation.

The template-mediated distortions are the result of significant disruptions in the transition from the A form (at the active site) to the B form (at *n-6*) of the DNA helix. This transition in the DNA conformation is a structural feature that is common to many polymerase families (Doublié et al., 1998; Huang et al., 1998; Jacobo-Molina and Arnold, 1993; Kiefer et al., 1998; Li et al., 1998; Pelletier et al., 1994). A notable exception is the "error prone" Y family, in which no distortion of the DNA conformation is observed along the polymerase-DNA interface (Ling et al., 2001). The absence of a conformational transition in the DNA duplex binding region may contribute to the insensitivity of these polymerases to downstream replication errors or damaged DNA.

Mismatch Extension Efficiencies

It is tempting to speculate that the category of mismatch-induced disruptions can be correlated with differences in mismatch extension efficiencies observed with other Family A DNA polymerases (Huang et al., 1992; Joyce et al., 1992). However, such correlations are not straightforward, because the inherent flexibility of DNA permits access to many states (see below). Interference with the catalytic site is the most critical, yet even in severely disrupted catalytic sites a catalytically competent conformation can remain accessible by virtue of this flexibility. Conversely, in cases where the catalytic site is undisturbed, observed decreases in the rates of mismatch extension must arise from distortions in one or more of the other sites.

For example, the G•T mismatch, which is one of the easiest mismatches to extend in solution and is associated with a high frequency of mutagenesis (Huang et al., 1992; Joyce et al., 1992; Kunkel and Bebenek, 2000), shows no structural disruption of the catalytic site and can be extended readily in the polymerase crystals. In this case, the observed displacement of the template is predicted to interfere with transfer from the preinsertion site to the insertion site. This disruption may account for the reduced extension efficiency of G•T mismatch of ten-fold to a thousand-fold depending on the polymerase (Echols and Goodman, 1991; Huang et al., 1992; Joyce et al., 1992). Although the structural disruptions observed in the G•G mismatch are similar to G•T, the additional loss of a hydrogen bond with Arg615, which plays a dual role in anchoring the primer base at the postinsertion site and the incoming dNTP at the insertion site (see Results), may account for the observation that G•G extension efficiencies are more than 100-fold slower than that of G•T (Huang et al., 1992; Joyce et al., 1992).

In contrast, the C•T mismatch complex has an extensively disrupted active site, but nevertheless can be extended readily both in solution and in the BF crystal. This case illustrates that DNA must be conformationally dynamic and that competent conformers that are not observed in the crystals are kinetically accessible at the *n-1* position. Furthermore, a reasonable model for a catalytically competent C•T mismatch conformation is suggested by the structure of the C•T mismatch located at the *n-2* position, generated by extension in the crystal (Figure 5). In this structure, the pyrimidine bases are stretched apart in a manner that does not disrupt the DNA helix. If a similar conformation is adopted at the postinsertion site, the primer 3′OH would be appropriately positioned for catalysis. The loss of interbase hydrogen bonds in this conformation is compensated for by hydrogen bonds between protein residues and the minor groove. This model suggests a structural basis for the ease of C•T extension reported in both KF (Joyce et al., 1992) and *Taq* (Huang et al., 1992).

Another case is represented by the A•G mismatch, which is among the most difficult to extend in solution (Huang et al., 1992; Joyce et al., 1992). The A•G mismatch shows significant disruptions both in the catalytic site and template binding regions. Extension efficiencies of the A•G mismatch are reduced by up to a million-fold in related polymerases (Huang et al., 1992; Joyce et al., 1992). With extension efficiencies between that of the A•G and G•T mismatches (Huang et al., 1992; Joyce et al., 1992), the T•T mismatch has a significantly disrupted catalytic site but intact template binding re-

gion. The T•T and A•G mismatches have yet to be extended in the BF crystal.

Stalling and Dissociation of the Mismatch from the Polymerase

Exonucleolytic excision of mismatches requires dissociation of the heteroduplex from the polymerase active site, followed by cleavage by a proofreading exonuclease located either on the polymerase itself (not in the case of BF) or on a separate enzyme (Joyce, 1989; Kunkel and Bebenek, 2000; Perrino and Loeb, 1989). Furthermore, dissociation of a mismatch from the polymerase per se contributes to replication fidelity, since mismatch extension stops. A number of the structures that we have observed show significantly diminished interactions between the heteroduplex and the polymerase, indicating weakened binding of a heteroduplex by a polymerase. In the cases where the template strand distortions are observed (G•T, T•G, A•G, G•G, extension of G•T), the number of hydrogen bonds between the polymerase and DNA is reduced, and the template has lifted away from the polymerase surface. Additionally, ordered waters can be observed between the DNA and the polymerase, which are absent in homoduplex structure. In several cases (T•G, G•G, extensions of G•T), the amino acid side chains involved in minor groove readout (Arg615, Gln797) adopt a conformation observed in an apo-polymerase structure (Kiefer et al., 1997).

Furthermore, in three cases the mismatch is not paired, but frays instead (G•A, A•A, C•C). In the frayed structures, the bases are held apart by interactions with the protein (Figure 3). Factors that stabilize the formation of 3′ single-stranded ends such as observed in these frayed mismatches enhance excision (Kunkel and Bebenek, 2000). Here we observe that the polymerase itself can stabilize these single-stranded mismatched structures, which is likely to contribute directly to stalling prior to dissociation and subsequent excision editing steps.

Heteroduplex DNA Conformation of the Mismatches

We observe that mismatches form stable complexes with the DNA polymerase by exploiting the large number of conformational degrees of freedom that are thermodynamically and kinetically accessible within the DNA heteroduplex, resulting in noncanonical DNA structures, and heteroduplex-protein interactions that differ from the homoduplex interactions. Relatively small motions and changes are sufficient to avoid steric barriers and hydrogen bonding deficiencies in the interactions between the polymerase and the mismatch. The interactions between the protein and the DNA clearly alter the equilibria between accessible conformations and states. Furthermore, given the wide range of conformational flexibility and the potential stabilization of any of these states by protein-DNA interactions, there is no a priori expectation that the mismatch DNA structures in complex with the polymerase will be similar to those observed in the absence of protein.

Some mismatch conformations that we have observed have not been previously described in a DNA heteroduplex (C•T, T•T, and extensions of C•T and G•T). Binding of mismatches to the polymerase active site stabilizes structures that include stretched pairings (C•T extension, Figure 5), transition in base orientation from *anti* to *syn* (G•G), and water-bridged interbase hydrogen bonds (C•T, T•G, and extensions of C•T or G•T). Structural adaptations in mismatches are not confined to simple conformational rearrangements; the noncanonical hydrogen bonds observed in the G•T extension reaction products (Figure 4) involve either a tautomeric rearrangement or ionization of the bases.

Additionally, large-scale disruptions of the DNA helix is a common theme in our polymerase structures (Figures 3 and 4), in contrast to heteroduplex structures obtained in the absence of protein, which are dominated by local disruptions (Kennard and Hunter, 1989). Striking examples of such large-scale transitions are the effect of a mismatch on the long-range transition between A and B forms of DNA and distortions in the template strand.

Multiple Base Conformations at the Mismatch Site

Three of the mismatch structures are disordered, with the disorder being confined to the position of the mismatch, retaining a well-ordered structure in the other parts of the DNA and polymerase. As with the frayed A•A and C•C mismatches, the A•C and C•A mismatches are unable to form more than a single hydrogen bond in their dominant tautomeric state, although additional hydrogen bonding could occur upon formation of a different tautomer or ionized state. Indeed, an A•C mismatch formed in the absence of protein adopts a protonated form that forms two hydrogen bonds (Hunter et al., 1986). The observed disorder could therefore indicate the presence of several different species, arising from an equilibrium between several different states of the bases, again illustrating the dynamic nature of the protein-DNA interactions.

Nonequivalence of Mismatch Binding

The binding interactions of equivalent correct base pairs (e.g., A•T, T•A) to the polymerase are identical (Johnson et al., 2003; Li and Waksman, 2001), whereas each equivalent mismatch (e.g., G•T, T•G) interacts in a unique manner with the protein. For instance, the T•G mismatch breaks contact with Arg615 in the minor groove and shows some disruption in the catalytic site due to a shifting of the primer thymine into the major groove. On the other hand, the G•T mismatch shows no disruption in either position along the primer strand but has disruptions associated with the template instead. The nonequivalence of interactions continues to be observed at the level of chemical classes. For instance, purine•purine mismatches interact differently. In the case of the A•G mismatch, both bases retain an *anti* conformation, which results in significant disruption of the catalytic site. Conversely, the primer base of the G•G mismatch switches to the *syn* conformation, which narrows the width of the base pair and preserves the assembly of the catalytic site.

Outlook

The data presented here are expected to form the basis for future biochemical and structural studies, molecular

simulations, and protein engineering experiments for understanding the mechanism of mismatch and lesion incorporation. The structures that we have captured are unlikely to represent all possible interactions between a mismatch and the polymerase. For a given mismatch, different conditions (pH, sequence context, etc.) may lead to other relevant structural changes. However, it is likely that we have delineated some of the major mechanisms by which mismatches can disrupt extension and enhance the fidelity of replication. Many of the mechanisms described here for a Family A polymerase may be generally applicable to other DNA polymerases and may even extend to RNA polymerases (Yin and Steitz, 2002).

Experimental Procedures

Cocrystallization of BF with DNA Primer Template Substrates

Purification of the thermostable *Bacillus stearothermophilus* DNA polymerase large fragment (BF) and crystallization with DNA substrates was performed as described previously (Johnson et al., 2003; Kiefer et al., 1998). The DNA sequences used for cocrystallization are shown in Table 1.

Data Collection, Structure Determination, and Analysis

Diffraction data were collected at 98 K using an RAXIS-IV detector (Molecular Structure) on a Rigaku rotating anode X-ray generator, at the NSLS X12B and X25 beam lines, and at the APS 14-BMC and 14-BMD beam lines (Tables 2 and 4). Data were processed, phases calculated, and models refined and built as described previously (Kiefer et al., 1998). DNA structure analysis was performed with 3DNA (Lu et al., 2000). A summary of base pair and helical parameters for the mismatch structures is shown in Tables 3 and 5 and Supplemental Table S1 at http://www.cell.com/cgi/content/full/116/6/803/DC1.

Catalysis in the Crystal

The protocol for accurate DNA synthesis in the BF•DNA crystals has been described previously (Johnson et al., 2003). G•T, T•G, T•T, and A•A mismatches were incorporated into BF•DNA complexes by replacing $MgSO_4$ with $MnSO_4$ in the reaction buffer, using the DNA sequence and dNTP indicated in Table 1. Mismatch extension experiments were performed in the presence of $MgSO_4$ by transferring BF•DNA cocrystals containing a mismatch at the 3′ primer terminus into reaction buffers containing nucleotides complementary to the template strand.

Acknowledgments

We thank J.S. Taylor and L.J. Forsberg for assistance with the A•A and C•T extension structures, respectively. We thank H.W. Hellinga for extensive discussions and C.M. Joyce and P. Modrich for critical reading of the manuscript. Research was carried out in part at the National Synchrotron Light Source, Brookhaven National Laboratory, which is supported by the U.S. Department of Energy, Division of Materials Sciences and Division of Chemical Sciences, under Contract No. DE-AC02-98CH10886. Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Basic Energy Sciences, Office of Science, under Contract No. W-31-109-Eng-38. Use of the BioCARS Sector 14 was supported by the National Institutes of Health, National Center for Research Resources, under grant number RR07707. The work was supported by grants to L.S.B. from the Human Frontiers Science Program (RG0351/1998-M) and the National Cancer Institute (P01 CA92584).

Received: February 19, 2003
Revised: January 7, 2004
Accepted: January 23, 2004
Published: March 18, 2004

References

Baeyens, K.J., De Bondt, H.L., and Holbrook, S.R. (1995). Structure of an RNA double helix including uracil-uracil base pairs in an internal loop. Nat. Struct. Biol. *2*, 56–62.

Beckman, R.A., and Loeb, L.A. (1993). Multi-stage proofreading in DNA replication. Q. Rev. Biophys. *26*, 225–331.

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.

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.

Carver, T.J., Hochstrasser, R.A., and Millar, D.P. (1994). Proofreading DNA: recognition of aberrant DNA termini by the Klenow Fragment of DNA polymerase I. Proc. Natl. Acad. Sci. USA *91*, 10670–10674.

Dickerson, R.E., Bansal, M., Calladine, C.R., Diekmann, S., Hunter, W.N., Kennard, O., Lavery, R., Nelson, H.C.M., Olson, W.K., Saenger, W., et al. (1989). Definitions and nomenclature of nucleic acid structure parameters. J. Mol. Biol. *205*, 787–791.

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.

Doublié, S., Sawaya, M.R., and Ellenberger, T. (1999). An open and closed case for all polymerases. Struct. Fold. Des. *7*, R31–R35.

Echols, H., and Goodman, M.F. (1991). Fidelity mechanisms in DNA replication. Annu. Rev. Biochem. *60*, 477–511.

Eger, B.T., Kuchta, R.D., Carroll, S.S., Benkovic, P.A., Dahlberg, M.E., Joyce, C.M., and Benkovic, S.J. (1991). Mechanism of DNA replication fidelity for three mutants of DNA polymerase I: Klenow Fragment KF(exo+), KF(polA5), and KF(exo-). Biochemistry *30*, 1441–1448.

Friedberg, E.C., Fischhaber, P.L., and Kisker, C. (2001). Error-prone DNA polymerases: novel structures and the benefits of infidelity. Cell *107*, 9–12.

Goodman, M.F., Creighton, S., Bloom, L.B., and Petruska, J. (1993). Biochemical basis of DNA replication fidelity. Crit. Rev. Biochem. Mol. Biol. *28*, 83–126.

Huang, M.M., Arnheim, N., and Goodman, M.F. (1992). Extension of base mispairs by Taq DNA polymerase: implications for single nucleotide discrimination in PCR. Nucleic Acids Res. *20*, 4567–4573.

Huang, H., Chopra, R., Verdine, G.L., and Harrison, S.C. (1998). Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. Science *282*, 1669–1675.

Hunter, W.N., Brown, T., Anand, N.N., and Kennard, O. (1986). Structure of an adenine.cytosine base pair in DNA and its implications for mismatch repair. Nature *320*, 552–555.

Hunter, W.N., Brown, T., Kneale, G., Anand, N.N., Rabinovich, D., and Kennard, O. (1987). The structure of guanosine-thymidine mismatches in B-DNA at 2.5-Å resolution. J. Biol. Chem. *262*, 9962–9970.

Jacobo-Molina, A., and Arnold, E. (1993). Crystal structure of Human Immunodeficiency Virus Type I reverse transcriptase complexed with double-stranded DNA at 3.0 Å shows bent DNA. Proc. Natl. Acad. Sci. USA *90*, 6320–6324.

Johnson, K.A. (1993). Conformational coupling in DNA polymerase fidelity. Annu. Rev. Biochem. *62*, 685–713.

Johnson, S.J., Taylor, J.S., and Beese, L.S. (2003). Processive DNA synthesis observed in a polymerase crystal suggests a mechanism for the prevention of frameshift mutations. Proc. Natl. Acad. Sci. USA *100*, 3895–3900.

Joyce, C.M. (1989). How DNA travels between the separate polymerase and 3′-5′ exonuclease sites of DNA polymerase I (Klenow Fragment). J. Biol. Chem. *264*, 10858–10866.

Joyce, C.M., and Steitz, T.A. (1994). Function and structure relationships in DNA polymerases. Annu. Rev. Biochem. *63*, 777–822.

Joyce, C.M., Sun, X.C., and Grindley, N.D.F. (1992). Reactions at the polymerase active site that contribute to the fidelity of *Esche-*

richia coli DNA polymerase I (Klenow Fragment). J. Biol. Chem. *267*, 24485–24500.

Kennard, O., and Hunter, W.N. (1989). Oligonucleotide structure: a decade of results from single crystal X- ray diffraction studies. Q. Rev. Biophys. *22*, 327–379.

Kiefer, J.R., Mao, C., Hansen, C.J., Basehore, S.L., Hogrefe, H.H., Braman, J.C., and Beese, L.S. (1997). Crystal structure of a thermostable *Bacillus* DNA polymerase I large fragment at 2.1 Å resolution. Structure *5*, 95–108.

Kiefer, J.R., Mao, C., Braman, J.C., and Beese, L.S. (1998). Visualizing DNA replication in a catalytically active *Bacillus* DNA polymerase crystal. Nature *391*, 304–307.

Kuchta, R.D., Benkovic, P., and Benkovic, S.J. (1988). Kinetic mechanism whereby DNA polymerase I (Klenow) replicates DNA with high fidelity. Biochemistry *27*, 6716–6725.

Kunkel, T.A., and Bebenek, K. (2000). DNA replication fidelity. Annu. Rev. Biochem. *69*, 497–529.

Li, Y., and Waksman, G. (2001). Crystal structures of a ddATP-, ddTTP-, ddCTP, and ddGTP- trapped ternary complex of Klentaq1: insights into nucleotide incorporation and selectivity. Protein Sci. *10*, 1225–1233.

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.

Ling, H., Boudsocq, F., Woodgate, R., and Yang, W. (2001). Crystal structure of a Y-family DNA polymerase in action: a mechanism for error-prone and lesion-bypass replication. Cell *107*, 91–102.

Lu, X.J., Shakked, Z., and Olson, W.K. (2000). A-form conformational motifs in ligand-bound DNA structures. J. Mol. Biol. *300*, 819–840.

Miller, H., and Grollman, A.P. (1997). Kinetics of DNA polymerase I (Klenow fragment exo-) activity on damaged DNA templates: effect of proximal and distal template damage on DNA synthesis. Biochemistry *36*, 15336–15342.

Ng, L., Weiss, S.J., and Fisher, P.A. (1989). Recognition and binding of template-primers containing defined abasic sites by *Drosophila* DNA polymerase alpha holoenzyme. J. Biol. Chem. *264*, 13018–13023.

Patel, D.J., Shapiro, L., and Hare, D. eds. (1987). Conformation of DNA Base Pair Mismatches in Solution (Berlin: Springer-Verlag).

Pelletier, H., Sawaya, M.R., Kumar, A., Wilson, S.H., and Kraut, J. (1994). Structures of ternary complexes of rat DNA polymerase beta, a DNA template-primer, and ddCTP. Science *264*, 1891–1903.

Perrino, F.W., and Loeb, L.A. (1989). Proofreading by the epsilon subunit of *Escherichia coli* DNA polymerase III increases the fidelity of calf thymus DNA polymerase alpha. Proc. Natl. Acad. Sci. USA *86*, 3085–3088.

Skelly, J.V., Edwards, K.J., Jenkins, T.C., and Neidle, S. (1993). Crystal structure of an oligonucleotide duplex containing G.G base pairs: influence of mispairing on DNA backbone conformation. Proc. Natl. Acad. Sci. USA *90*, 804–808.

Steitz, T.A. (1999). DNA polymerases: structural diversity and common mechanisms. J. Biol. Chem. *274*, 17395–17398.

Weiss, S.J., and Fisher, P.A. (1992). Interaction of *Drosophila* DNA polymerase alpha holoenzyme with synthetic template-primers containing mismatched primer bases or propanodeoxyguanosine adducts at various positions in template and primer regions. J. Biol. Chem. *267*, 18520–18526.

Yin, Y.W., and Steitz, T.A. (2002). Structural basis for the transition from initiation to elongation transcription in T7 RNA polymerase. Science *298*, 1387–1395.