Structural Mechanisms of Accurate and Mutagenic DNA Replication

by

Weina Wang

Department of Biochemistry
Duke University

Date:_____

Approved:

_____
Lorena S. Beese, Supervisor

_____
Richard G. Brennan

_____
Charles W. Carter Jr.

_____
Leonard D. Spicer

_____
Pei Zhou

Dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor of Philosophy in the Department of
Biochemistry in the Graduate School
of Duke University

2012

ABSTRACT

Structural Mechanisms of Accurate and Mutagenic DNA Replication

by

Weina Wang

Department of Biochemistry
Duke University

Date:_____

Approved:

_____
Lorena S. Beese, Supervisor

_____
Richard G. Brennan

_____
Charles W. Carter Jr.

_____
Leonard D. Spicer

_____
Pei Zhou

An abstract of a dissertation submitted in partial
fulfillment of the requirements for the degree
of Doctor of Philosophy in the Department of
Biochemistry in the Graduate School
of Duke University

2012

Abstract

High-fidelity DNA polymerases achieve remarkable accuracy by exhibiting exquisite nucleotide substrate selection mechanisms. DNA polymerases not only discriminate highly against base-pair mismatches, but also show a high degree of selectivity for the correct sugar moiety deoxyribonucleotide over ribo- and dideoxy-nucleotides. Although DNA polymerase is highly accurate, some base-pair mismatches are still incorporated at low frequency, leading to spontaneous mutagenesis. Mechanisms for both accurate and mutagenic DNA replication have been subjected to intense solution studies and speculations over half a decade. However, structural understanding of both processes are still very limited due to the lack of crystal structures of DNA polymerase bound with incorrect nucleotide substrates especially at the insertion step prior to chemistry which accounts for the majority of the intrinsic specificity of the polymerase.

In this dissertation, X-ray crystallographic analyses were performed using a model system for studying replication fidelity, a thermostable strain of *Bacillus* DNA polymerase I large fragment (*Bacillus* fragment, BF) that catalyzes replication in crystals. A series of high-resolution crystal structures of BF polymerase variants and DNA duplex with mismatches, incorrect sugar substrates, and cognate base pairs bound at various fidelity filter sites on the polymerase surface were determined. By comparing structures of non-cognate base pairs with those of cognate base pairs captured under the same experimental condition, a unified picture of substrate selectivity by DNA polymerase for both mutagenic and accurate replication process has emerged.

Contents

List of Tables

List of Figures

Acknowledgements

I would like to thank my advisor Dr. Lorena S. Beese for a great project to work on and her guidance throughout my time in her lab. Among the many things I have learned from her, critical thinking, problem framing and solving, and effective communication both verbally and written are especially beneficial. I also want to thank Lorena for the opportunities for me to go to national and international conferences to present my work. She has shaped my scientific career more than anyone else has.

I am also grateful to my committee members for their encouragement and great suggestions along the way. Their doors were always open whenever I had a question.

I want to thank my collaborators on the YFP project, Homme Hellinga and Joshua Grimly. It was a great pleasure to work with them to turn a wealth of experimental data into advancement in understanding how this important sensor works. I also want to thank Homme for his help on my DNA polymerase project. I have learned a lot from his passion for science, approach to solving complicated problems, and outstanding writing skills.

I want to thank all Beese lab members, past and present, for their help and support, especially: Eugene Wu, who mentored me through my first three years in the lab and was always very generous with his time; James Phillips, who made an easy access to home-source X-ray machine possible; and Stephanie Armstrong, who helped me with mutagenesis and protein preparation related to the sugar recognition project.

I would also like to acknowledge the staff at SIBYLS and SER-CAT beamlines for their assistance with data collection.

Dedication

To Mom and Dad – my source of love and encouragement

To Yi − my best friend and sweetheart

Chapter 1 Introduction

Accurate DNA replication is a fundamental requirement for the survival of an organism and the passing of genetic information from one generation to the next [1]. Errors made during this process, if left uncorrected, lead to mutations, the basis for all genetic diseases and driving force for evolution. DNA polymerase carries out template-directed DNA replication by chemically incorporating its substrate, 2'-deoxyribonucleotide triphosphate (dNTP), onto the 3' terminus of the primer strand [2-4].

DNA polymerase achieves remarkably high replication fidelity by selecting the correct nucleotide for incorporation while discriminating against incorrect substrates [5]. The error rate for mismatch incorporation is determined by the intrinsic specificity of the DNA polymerase, the identity of the base, and the sequence flanking the incorporation site [5-8]. In addition to base selectivity, DNA polymerase also chooses the correct sugar moiety (2'-deoxyribose) to incorporate and discriminates highly against the ribose sugar [9, 10].

Extensive biochemical, kinetic, and structural studies on a variety of DNA polymerases have provided valuable insights into the replication fidelity mechanisms employed by this important molecular machinery. This chapter will review the current understanding of DNA replication fidelity mechanisms with a special focus on the staggering amount of structural studies on various DNA polymerases that have become available in recent years.

Classification of DNA polymerases

Since the discovery of the first DNA polymerase *E. coli* DNA polymerase I (Pol I) more than half a century ago [1], many more polymerases have been identified and characterized.

Accurate DNA replication is achieved by cooperative efforts from DNA polymerases involved in genome replication, DNA repair, and translesion synthesis [11]. Based on primary sequence homology [12, 13], polygenetic relationships [14], and structural analysis [15], DNA polymerases have been classified into seven families: A, B, C, D, X, Y, and reverse transcriptase (RT) [16].

Families A, B, and C are mostly high-fidelity DNA polymerases mainly responsible for genome replication and DNA repair. Family A DNA polymerases are further divided into two groups based on their functionality: replicative (e.g., bacteriophages T5 and T7 DNA polymerases; the eukaryotic mitochondrial DNA polymerase γ) and repairing and gap filling (e.g., *E. coli*, *T. aquaticus*, and *Bacillus* DNA polymerase I) [13]. Family B polymerases include major replicative polymerases for various species (e.g. *E. coli* DNA polymerase II; bacteriophage φ29 and RB69 DNA polymerases; human DNA polymerase α) [13]. Family C polymerases are bacterial replication DNA polymerase (e.g. *E. coli* DNA polymerase III α subunit) [13]. Crystal structures have been determined for DNA polymerases from all three families.

Family X polymerases are involved in DNA repair pathways such as base excision repair (BER) by human Pol β. Other members include template-independent terminal deoxynucleotidyl transferase (TdT); Polλ; Polμ and so on [17]. Family Y polymerases carry out translesion synthesis past damaged template which stalls high-fidelity DNA polymerases and exhibit low fidelity when copying normal DNA [11, 18]. They are widespread from bacteria to human (e.g. Dpo4, Rev1, dINb, UmuC, and human Pol η) [19]. RT family members are from retroviruses (e.g. human immunedeficiency virus (HIV) 1 RT and Moloney murine leukemia virus (MMLV) RT) and eukaryotic telomerases [20]. They are very important drug targets to treat retroviral diseases [21]. Crystal structures of a number of X, Y, and RT family members have been determined. Family D polymerases are archeal and are at an early stage of characterization [22]. There has been no crystal structure determined for this family yet.

Structural features of DNA polymerases

Over the past two decades, a large number of crystal structures of DNA polymerases from six families (A, B, C, X, Y, and RT) have been determined, which greatly enhanced our understanding of DNA replication mechanisms.

Overall architecture of DNA polymerase

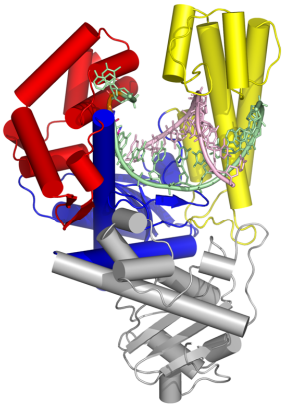
DNA polymerases share the same polymerase domain, while exhibiting notably different features with regard to the presence of additional domains related to the specific function of the polymerase [16]. Ternary complexes of representative DNA polymerases from six families are shown in Figure 1-1.

The polymerase domains of DNA polymerases are similar in three aspects [23]. First, they are all composed of fingers, palm, and thumb subdomains, following the naming convention originated from the first structure determined for *E. coli* DNA polymerase I large fragment (Klenow fragment, KF) [24] (Figure 1-1B). DNA binds in a cleft formed by the three subdomains (Figure 1-1B). Second, the function of each subdomain is similar across different DNA polymerases. The fingers domain binds the incoming nucleotide triphosphate and contacts the 5' overhang of the template. The palm domain is where nucleotidyl transfer reaction occurs. The thumb domain is involved in the binding of DNA duplex region [25]. Third, all DNA polymerases share the same two-metal-ion mechanism for catalysis [15, 26]. In ternary complexes, α-phosphate of the incoming nucleotide, 3' hydroxyl group of the primer terminus, two universally conserved aspartate residues, and two divalent metal ions, are aligned for chemistry. The active site assembly is remarkably similar across all families (Figure 1-2). Despite the similarities, the polymerase domains of various DNA polymerases exhibit notable differences in their sizes, secondary structures, and detailed interactions [23] (Figure 1-1B).

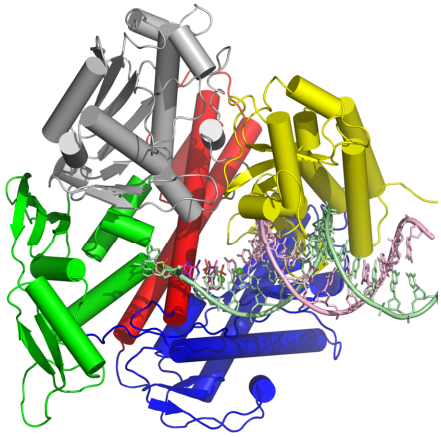
Figure 1-1 Ternary complexes of representative DNA polymerases from A, B, C, X, Y, and RT families.

A, DNA polymerases with all domains shown. B, DNA polymerases with only the polymerase domain shown. The polymerase domain is composed of three subdomains, fingers (red), thumb (yellow), and palm (blue). Other domains of the polymerases are colored differently: family A *Bacillus* Fragment (BF) (4DQI) [27], a vestigial exonuclease domain (gray) [28]. B-family, *Enterobacteria* phage RB69 DNA polymerase (pdb code: 3NCI) [29], NH_2-terminal domain (green) and Exo domain (gray) [30]; C-family, *Geobacillus kaustophilus* DNA polymerase PolC (3F2B), OB domain (cyan), PHP domain (hot pink), and DB domain (orange) [31]; X-family, human DNA polymerase beta (2FMQ) [32], lyase domain (teal) [33]; Y-family, human DNA polymerase eta (3MR2), little finger (salmon) [34]. Reverse Transcriptase (RT) family, HIV-1 RT (3KK2) [35], connection domain (salmon) and RNaseH domain (light cyan) [36]. Selection of the structures for polymerase families was based on availability and resolution of the ternary complexes. Domain naming was cited from the papers where the structures were first described.

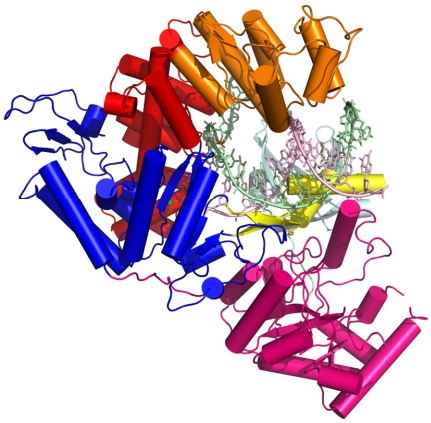
A



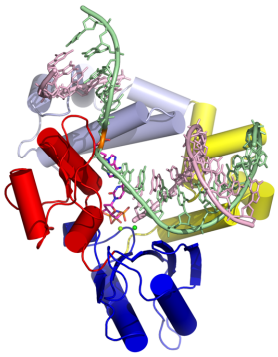
Family A
BF Pol I



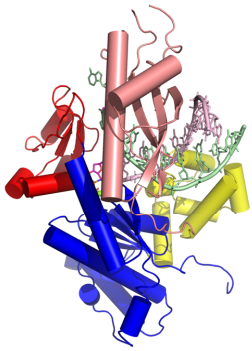
Family B
RB69 Pol



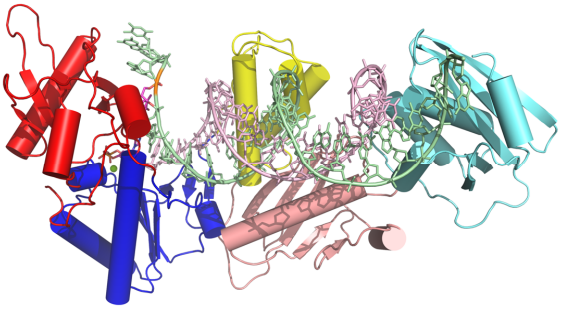
Family C
GkaPolC



Family X
human Pol β



Family Y
human Pol η



RT family
HIV-1 RT

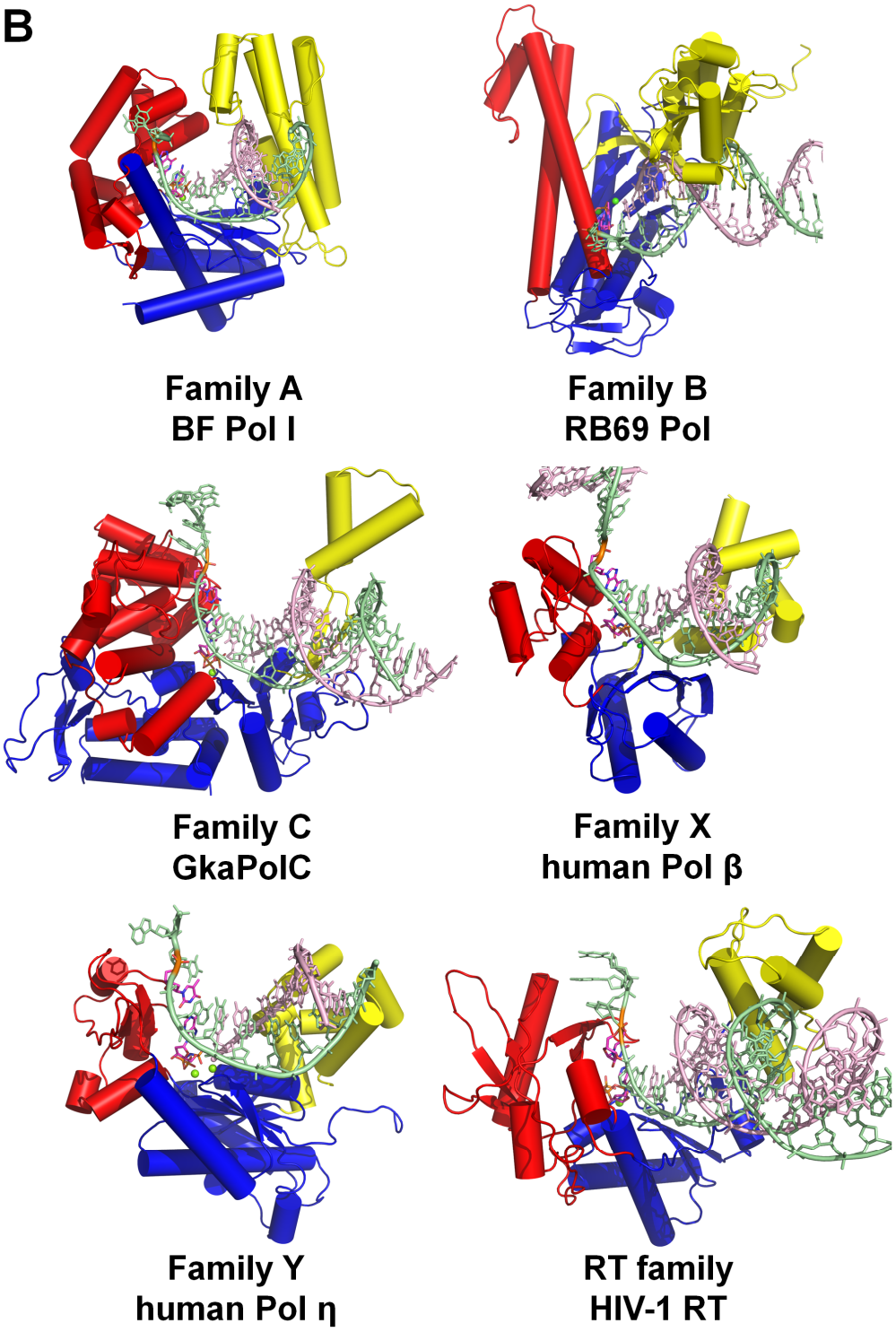
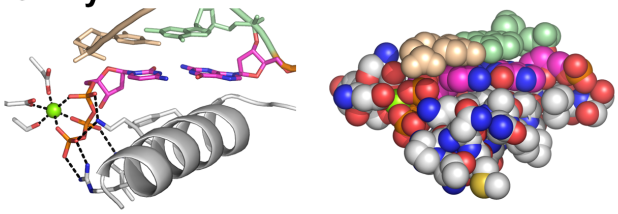

B
Family A
BF Pol I
Family B
RB69 Pol
Family C
GkaPolC
Family X
human Pol β
Family Y
human Pol η
RT family
HIV-1 RT

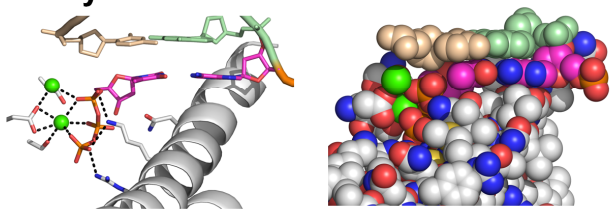
Figure 1-2 Active site of ternary complexes of DNA polymerases from A, B, C, X, Y, and RT families.

Left panel, cartoon representation of important features of the active site; right panel, same view with spheres shown. Catalytic site is assembled with the octahedral coordination of the catalytic metal ions (Mg^{2+} shown as green spheres; Na^{+} bound to Pol beta shown as purple sphere) by two conserved aspartate residues and the triphosphate. The polymerase interacts with the side of the triphosphate facing away from the metal ions via hydrogen bonds. In *Bacillus* fragment of family A (4DQI) [27] and RB69 DNA polymerase of family B (pdb code: 3NCI) [29], the trisphosphate interacts with side chains of the residues on an α-helix of the fingers domain. However, in GkaPolC of C-family (3F2B) [31], human Pol beta of X-family (2FMQ) [32], human Pol eta of Y-family, (3MR2) [34], and HIV-1 RT of RT family (3KK2) [35], residues interacting with the triphosphate are more diversely distributed through the fingers domain.

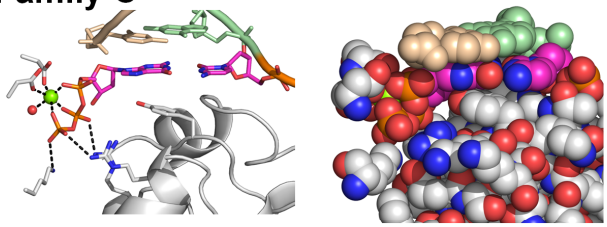
Family A



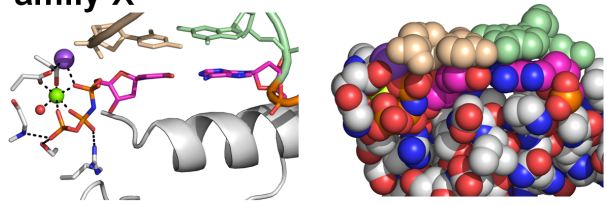
Family B



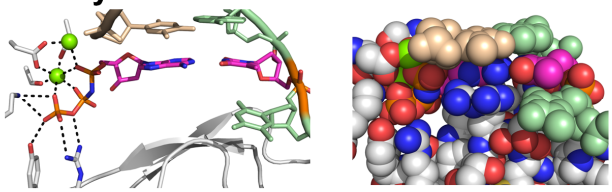
Family C



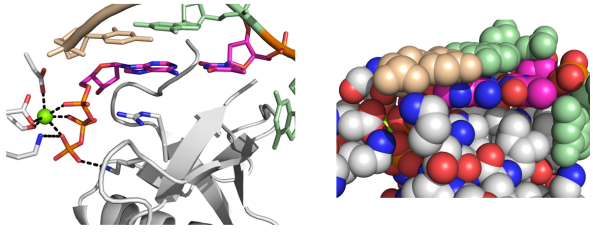
Family X



Family Y



RT family



DNA polymerase conformational changes

Comparisons between polymerase-DNA binary complex and polymerase-DNA-dNTP ternary complex show that A, B, X, and RT family DNA polymerases exhibit two distinct fingers domain conformations: open and closed (Figure 1-3). The polymerase fingers domain go through a large swinging motion from the open to the closed state upon the binding and pairing of correct nucleotide substrate and the nucleotide is aligned at the active site at the closed conformation ready for chemistry [37-42].

For family C, a comparison between binary and ternary complexes is not possible due to the lack of the former. However, by comparing ternary complex of *Taq* DNA Polymerase III or ternary complex of GkaPolC with apo structure of *Taq* DNA Polymerase III, it was suggested that Pol C also goes through major conformational changes upon the binding of DNA and nucleotide substrate [31, 43]. The low-fidelity family Y DNA polymerases show no large scale fingers domain movement that are observed for other families of DNA polymerases [44].

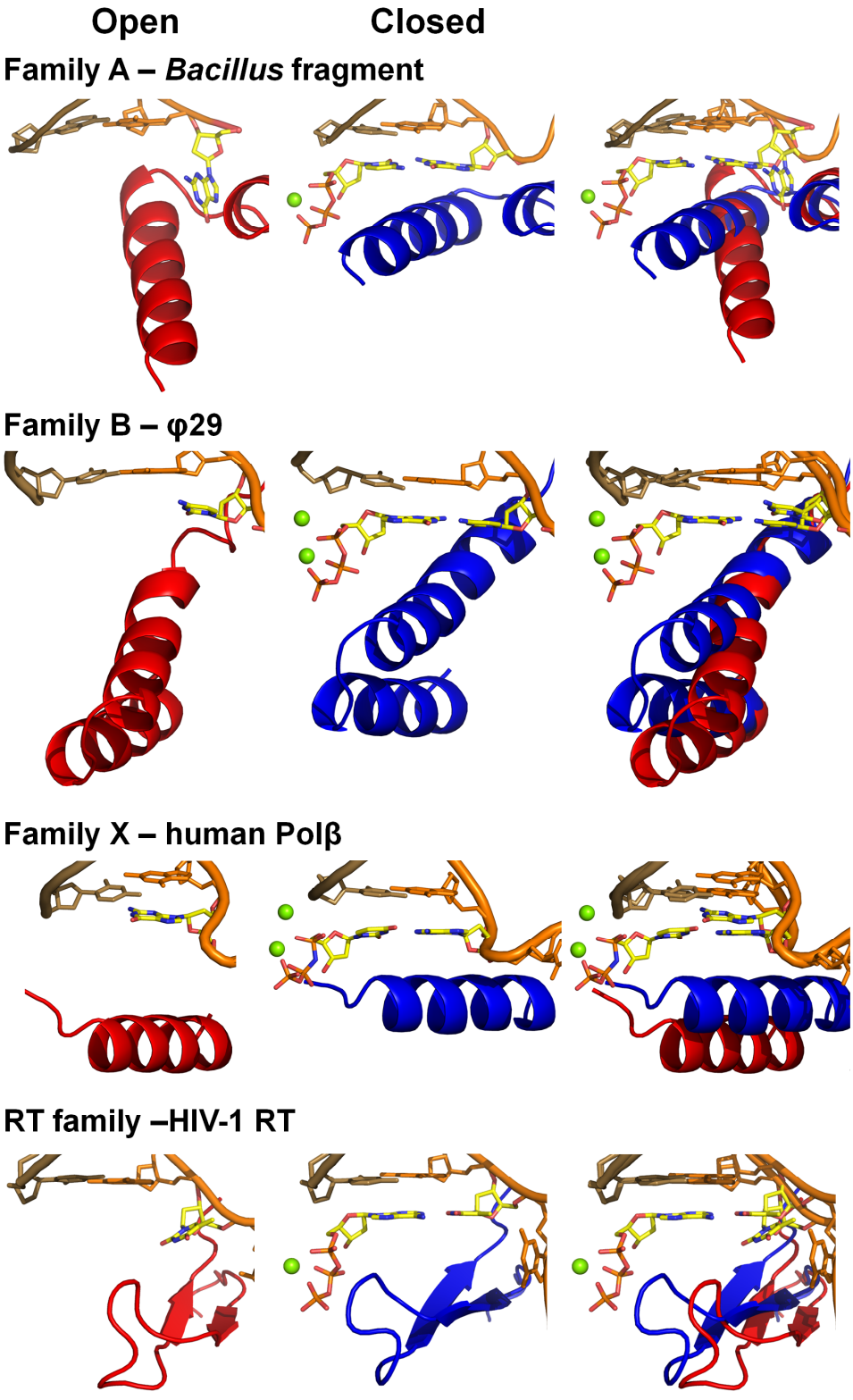
Family A DNA polymerases

Family A DNA polymerases are among the most well characterized structurally. Our model system for studying DNA replication, BF polymerase, also belongs to this family. Crystal structures for five members of the A family have been determined (Table 1-1).

Table 1-1 Crystal structures of DNA polymerases from family A

	Complex					
Polymerase	**Apo**	**Binary**	**Ternary**	**Mismatch**	**Lesion**	**Editing**
E. coli Klenow fragment	[24, 45]					[46-49]
T7 DNA polymerase			[50]		[51-54]	
Taq Pol I	[55, 56]	[37, 57]	[37, 58, 59]			
Bacillus fragment	[28]	[40, 60, 61]	[40, 62]	[61]	[62-65]	
Human Pol γ	[66]					

Figure 1-3 Open and closed conformation of DNA polymerase.
Four families of DNA polymerases have open-binary and closed-ternary crystal structures determined: family A *Bacillus* fragment open (1L3U) [40] and closed (4DQI) [27]; family B φ29 DNA polymerase open (2PZS) and closed (2PYJ) [67]; family X human Pol β open (3ISB) [68] and closed (2FMQ) [32]; RT family HIV-1 RT open (3KJV) and closed (3KK2) [35]. The left, middle, and right columns correspond to the open (red), closed (blue), and superimposed conformations. DNA primer (sand), template (orange), base pair prior to chemistry (yellow), and metal ions (green sphere) are also shown.


Open
Closed
Family A – *Bacillus* fragment
Family B – φ29
Family X – human Polβ
RT family –HIV-1 RT

Accurate DNA replication

High-fidelity DNA polymerases from the A, B, C, and RT families (proofreading-defective derivatives) exhibit 10^4 to 10^6-fold selectivity for incorporating their correct substrates against incorrect ones [69]. Five different sites [37, 40, 70, 71], pre-insertion, insertion, catalytic, post-insertion, and DNA duplex-binding region, function at various stages along the reaction pathway to ensure accurate DNA replication. A cartoon representation highlighting the fidelity filter sites at open and closed conformations of *Bacillu*s DNA polymerase I large fragment is shown in Figure 1-4 [40].

Pre-insertion site is a pocket formed by the loop between O and O1 helices and is where the next incoming template nucleotide resides. Upon the binding of a correct dNTP, the O helix undergoes a large conformational change from the open to the closed state. dNTP is paired with the complementary template base at the insertion site and its α-phosphate is aligned with regard to the 3' hydroxyl group of the primer 3' terminus at the catalytic site where two universally conserved aspartates and two divalent metal ions reside. Since substrate recognition at the insertion site accounts for the majority of DNA polymerase specificity [69], how incorrect nucleotides are discriminated against here is key to the understanding of replication fidelity.

At the insertion step, geometric and shape complementarity between the nascent base pair and the polymerase active site surface is one of the main determinants of specificity [69, 72]. Incorrect nucleotide substrates with aberrant base-pair shape and geometry cannot fit into the tight binding pocket which leads to nucleotide discrimination. Water exclusion at the tightly fitted active site prior to chemistry has been suggested to further amplify the free energy difference between the correct and incorrect nucleotides contributing to selectivity [73]. Conformational intermediates along the reaction pathway have been proposed to serve as “kinetic checkpoints” against incorrect nucleotide incorporation [73]. The crystallographically observed conformational

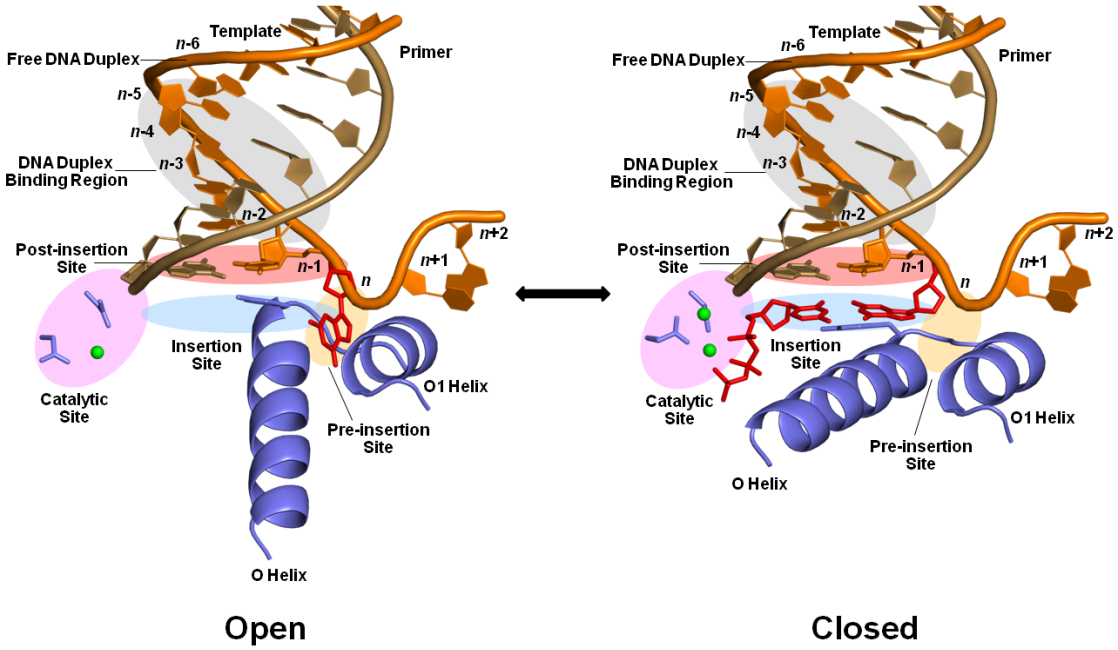


Figure 1-4 Fidelity filter sites of a high-fidelity DNA polymerase.
DNA polymerase replication fidelity filters. Shaded areas correspond to fidelity filters: pre-insertion site (*n*+1, orange), insertion site (*n*, blue), catalytic site (magenta), post-insertion site (*n*-1, pink), and DNA duplex-binding region (*n*-2 to *n*-5, gray). DNA primer (copper) and template strands (orange) are also shown. The O helix (blue) transitions from an open (left panel) to a closed (right panel) conformation. Cognate-shaped base pairs (red) are positioned for catalysis in the closed state. This figure combines information derived from three structures: open (1L3U) [40], closed (2HVI [62] and 3EZ5 [74]).

changes for DNA polymerases from different families are consistent with the possible existence of intermediate checkpoints [73] in between the open and closed states (Figure 1-3).

After the incoming nucleotide is incorporated onto the primer 3' terminus, DNA translocates, pyrophosphate is released, and the polymerase resets to the open state. The nascent base pair is placed at the post-insertion site where the minor groove edge of the base pair is read-out by protein residues providing another opportunity for error-checking [5, 69]. The readout is continued in the DNA duplex binding region up to four base-pairs away from the post-insertion site. The free DNA duplex region no longer retains interaction with the polymerase.

Nucleotide base and sugar selectivity

The substrate for DNA polymerase (dNTP) is composed of a base ring, a sugar moiety, and a triphosphate tail. In cells, nucleotides with five different bases, A, G, C, T, and U, and two types of sugar moieties, 2'-deoxyribose and ribose, exist at the same time. Two questions central to DNA replication fidelity and genome stability are:

1. How does DNA polymerase recognize and discriminate against nucleotide base-pair mismatches?

2. How does DNA polymerase recognize and select the correct deoxyribonucleotide sugar moiety to incorporate while discriminating against ribonucleotide?

Nucleotide base selectivity

In addition to the T•A and C•G cognate pairing schemes originally proposed by Watson and Crick more than 50 years ago [75], twelve base-pair mismatches, pyrimidine•purine (T•G, G•T, C•A, and A•C), purine•purine (G•G, G•A, A•G, and A•A), and pyrimidine•pyrimidine (T•T, T•C, C•C, and C•T), can arise at the DNA polymerase active site. Extensive solution

studies show that mismatch incorporation rates vary greatly depending on the specific mismatch, the intrinsic property of the polymerase, and the DNA sequence flanking the incorporation site [5, 76-78].

To understand the mechanisms of how DNA polymerase discriminates against incorrect base composition, structures of mismatches bound at the active site of DNA polymerase are required. Previously, all twelve mismatches were captured at the active site of *Bacillus* DNA polymerase I large fragment (*Bacillus* fragment, BF) after they were incorporated into the DNA duplex [61]. Six of the mismatches (T•G, G•T, G•G, A•G, T•T, and C•T) were paired and the base pairing schemes are shown in Figure 1-5. Three (A•A, G•A, and C•C) were frayed without base pairing and three (A•C, C•A, and T•C) were disordered. G•T and C•T were further placed at various positions at the DNA duplex-binding region and the free DNA duplex region. In all cases, mismatches with non-cognate base-pairing schemes cause disruptions to the polymerase active site which presumably slows down further incorporation and provides an opportunity for the exonuclease domain to excise the mismatch [61].

In contrast to the thorough structural studies of mismatches after chemistry [61], few mismatches have been captured prior to incorporation. Recently, a structure of a ddTTP·dG mismatch bound to the insertion site of BF polymerase was determined [79]. In addition, structures of a dGTP•dT mismatch bound to the active site of lower fidelity human DNA polymerase β [80] and Dpo4 [81] were determined. A complete understanding of base-pair selectivity prior to chemistry requires more structures of mismatches bound at the insertion site of high-fidelity DNA polymerases to be captured.

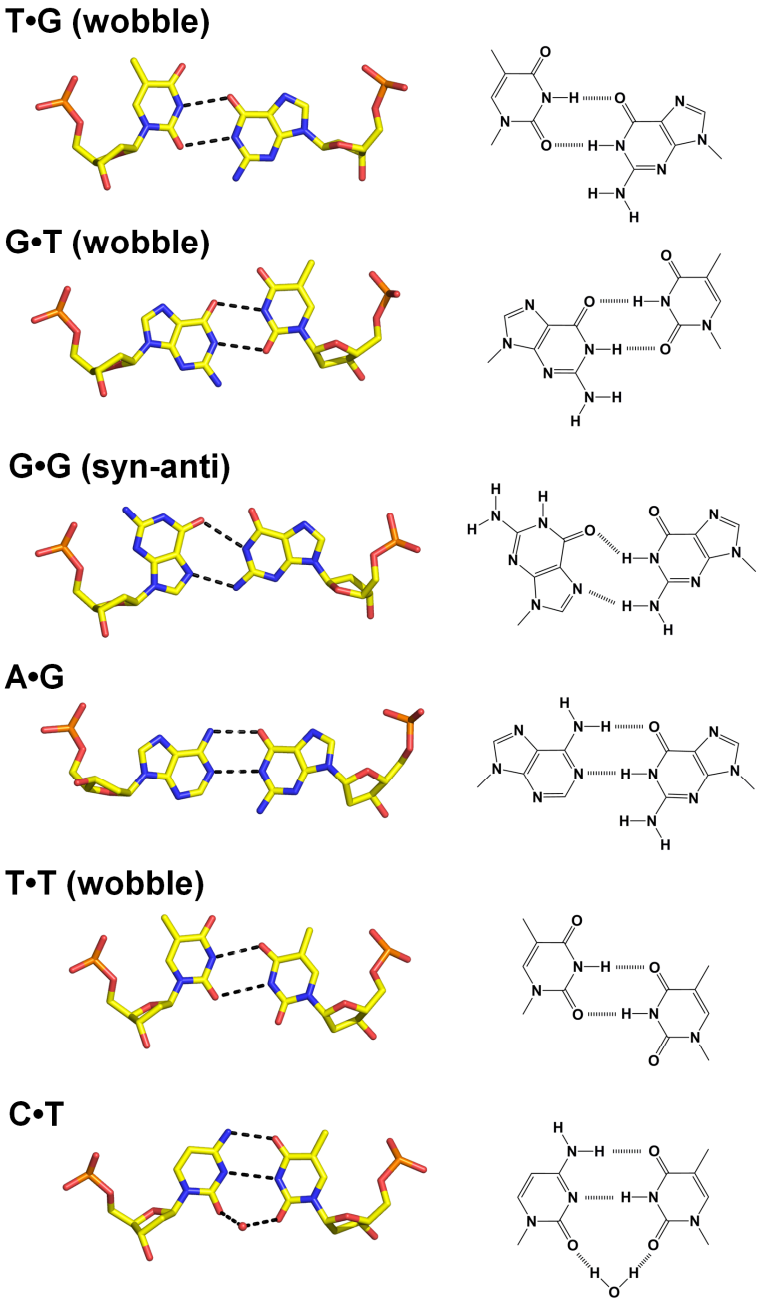


Figure 1-5 Mismatches captured after incorporation at the post-insertion site (*n*-1) of *Bacillus* fragment.
Left panel, cartoon representation of the base pair; right panel, Chemdraw figures showing details of the hydrogen bond interactions. The PDB codes for the structures are: 1NJX, 1NJW, 1NK4, 1NK0, 1NJY, and 1NJZ [61].

Nucleotide sugar recognition

In addition to base-pair mismatch discrimination, DNA polymerase also selects its natural substrate deoxyribonucleotide (dNTP) for incorporation over ribonucleotide triphosphate (rNTP) and 2′,3′-dideoxyribonucleotide triphosphate (ddNTP) [9, 10, 82] (Figure 1-6).

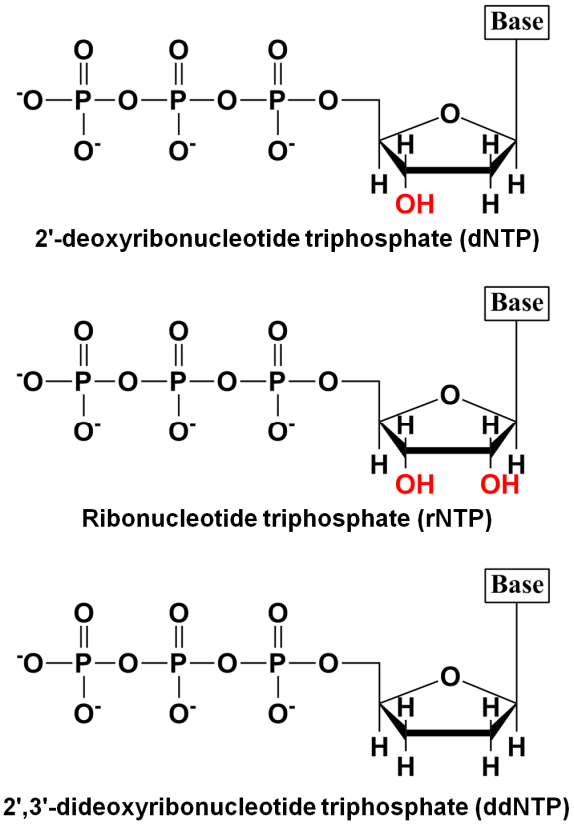


Figure 1-6 2'-deoxy-, ribose-, and 2',3'-dideoxy-nucleotides

Steric gate and ribonucleotide discrimination

In cells, DNA polymerases copy DNA in the presence of ribonucleotides which are 10-2,000-fold more abundant than their deoxy counterparts [10, 83-85]. Although DNA polymerase exhibits a high level of selectivity for deoxy- over ribose sugar, ribonucleotides are possibly the most common non-canonical substrates incorporated by the polymerase during initial DNA

replication [85] and subsequent gap filling DNA synthesis [10] due to their higher concentrations. The misincorporated ribonucletides can lead to replicative stress and genome instability by slowing down DNA replication and causing DNA strand to be more prone to cleavage [86].

A number of studies have been carried out to elucidate the mechanisms of ribonucleotide discrimination. Based on kinetic and structural studies of DNA polymerases and variants from the A, B, X, Y, and RT families, a residue in the vicinity of the sugar 2' position has been proposed to serve as a "steric gate" against ribonucleotide incorporation [9, 10] (Figure 1-7). The steric gate is a universally conserved glutamate residue in family A [9, 87-89], but for most polymerases from the B, X, Y, and RT families, it is either a tyrosine or a phenylalanine [10]. When the steric gate residue is mutated to amino acids with a smaller side chain, e.g. alanine, ribonucleotide selectivity is dramatically compromised [9, 10]. Structural analyses of six families of DNA polymerases show that in all families but X, the side chain of the steric gate is in the vicinity of the nucleotide sugar moiety 2' position consistent with the mutagenesis data [10]. In family X, the backbone carbonyl of the steric gate residue is placed closer to the sugar 2' position (Figure 1-7) [90]. This structural feature is likely to account for the observations that some members in the X family, when the steric gate residue is mutated, show relatively mild ribose sugar selectivity change while others exhibit no discrimination against ribonucleotide [10].

Although the steric gate residue plays a role in ribose sugar selectivity, removal of the steric block does not enable the polymerase to incorporate ribonucleotide as efficiently as deoxyribonucleotide, and it also greatly compromises the incorporation efficiency of dNTP [27]. Previous FRET and smFRET studies on the Klenow fragment show that the polymerase adopts intermediate conformations upon the binding of ribonucleotide [91, 92]. Therefore, to fully understand ribonucleotide selectivity, structures of high-fidelity DNA polymerases in complex with ribonucleotide are desired.

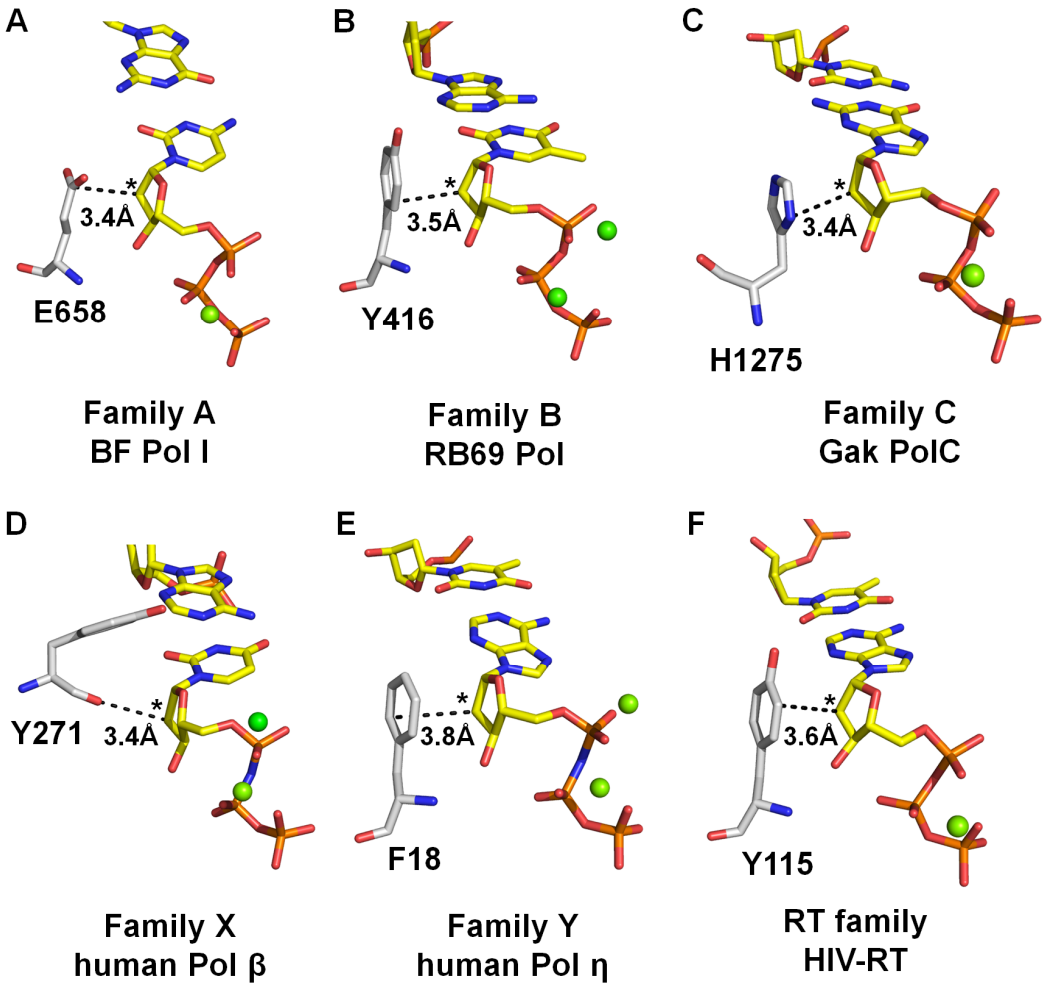


Figure 1-7 Relative position of the "Steric gate" residue and sugar C2' atom at the active site of representative members of six DNA polymerase families.
A, A-family, *Bacillus* Fragment (BF) (pdb code: 4DQI). B, B-family, Enterobacteria phage RB69 DNA polymerase (3NCI) [29]. C, C-family, *Geobacillus kaustophilus* DNA polymerase PolC (3F2B) [31]. D, X-family, human DNA polymerase beta (2FMP) [32]. E, Y-family, human DNA polymerase eta (3MR2) [34]. F, Reverse Transcriptase (RT) family, HIV-1 RT (3KK2) [35]. Selection of the structures for polymerase families was based on availability and resolution of the ternary complexes.

Dideoxynucleotide recognition

Family A DNA polymerases also recognize 3'-OH lacked in synthetic 2', 3'-dideoxynucleotide (ddNTP). As chain terminators, dideoxynucleotides are key components of DNA sequencing reaction and have clinical applications as antiviral drugs (e.g. 2′,3′-dideoxycytidine) [21]. Solution mutagenesis studies in family A DNA polymerases have shown that a conserved aromatic residue on the O helix [88] plays a key role: phenylalanine discriminates strongly against ddNTP while tyrosine substitute allows ddNTP to be incorporated with similar efficiency as dNTP [82, 93, 94]. Some members of the family A have a tyrosine at this position as the wild-type enzyme (e.g. T7 DNA polymerase, human mitochondria DNA polymerase γ) while some have a phenylalanine (e.g. Klenow fragment, Klentaq, *Bacillus* fragment) [88]. Although we and others have taken advantage of tyrosine at this position to promote incorporation of dideoxynucleotide to the DNA primer 3′ end as a chain terminator in order to trap complexes prior to chemistry [37, 50, 62], a structure of ddNTP bound ternary complex in the presence of phenylalanine is lacking, as is our understanding of dideoxynucleotide discrimination.

Mutagenic DNA replication

Although polymerases achieve very high specificity when replicating DNA, some mismatches are still incorporated at a low frequency, leading to spontaneous mutagenesis [5]. It has been proposed that mismatches may conform to the constrained DNA polymerase ternary active site by mimicking the shape of cognate base pairs through base tautomerization [75, 95], ionization [96, 97], and syn-anti conformational changes [69, 95, 98]. However, direct evidence for base-pair mismatch misincorporation through tautomerization or ionization is still lacking. Recently, a G•T mismatch has been observed at the human DNA polymerase β insertion site

which adopts a Watson-Crick base-pair shape through base ionization and is aligned for chemistry [80]. There is still no direct structural evidence for the long proposed rare tautomer hypothesis for spontaneous mutagenesis [75, 95].

Bacillus fragment as a model system

Bacillus DNA polymerase I large fragment (*Bacillus* fragment, BF) (67.7 kDa) belongs to family A. It contains a polymerase domain and a vestigial 3'-5' exonuclease domain that lacks proofreading activity [28]. BF is a close homologue of other members in the family, such as Klenow fragment (48.4% sequence identity) [99] and Klentaq (large fragment of *Taq* DNA Pol I) (52.3% sequence identity) [100].

Four characteristics make BF polymerase an ideal model system to study DNA replication fidelity mechanisms. First, large quantities of pure, stable, and catalytically active BF protein can be obtained readily making X-ray crystallographic studies possible. Second, BF can carry out processive DNA synthesis even in the crystal which can be taken advantage of to capture various polymerase-substrate complexes. Third, BF crystal can diffract to 1.4 Å resolution which allows detailed structural comparisons between complexes. Finally, all five fidelity filter sites critical for replication fidelity are present at the BF polymerase active site (Figure 1-4).

In the past 15 years, a library of high-resolution structures of BF polymerase in complex with cognate base pairs [40, 60], mismatches [61, 79, 101], and lesions [62-65] have been determined which have greatly advanced our understanding of replication fidelity (Appendix A).

Outline and objectives of this research

To study structural mechanisms of accurate and mutagenic DNA replication, structures of mismatches and incorrect sugar substrates captured at high-fidelity DNA polymerase active site are required. The goal of this research is to use the *Bacillus* DNA polymerase I large fragment as a model system to capture complex structures of BF, DNA, and various nucleotide substrates at fidelity filter sites especially at the pre-chemistry insertion site which is critical for substrate selectivity [69]. At the beginning of my research, there had been no structure of base-pair mismatch or incorrect sugar substrate bound at the insertion site of high-fidelity DNA polymerases.

I have been able to successfully co-crystallize BF polymerase with all 12 mismatches which resulted in a wide spectrum of substrate binding modes and polymerase O helix conformations. Among them, a C•A (primer•template) mismatch was of special interest because this mismatch mimics the shape of a cognate Watson-Crick base pair prior to chemistry. Although it has long been proposed by Watson and Crick that mismatches can mimic cognate base-pair shape by adopting their rare tautomeric forms [75] which leads to spontaneous mutagenesis [95], such a structure has never been captured before. Here, in the presence of Mn^{2+}, a C•A mismatch adopts the cognate base-pair shape by tautomerization and is properly aligned at the polymerase active site for incorporation. In addition, in the presence of Mg^{2+}, the C•A mismatch formed a wobble pair and was misaligned for chemistry. The mismatch was further placed at the post-insertion site, the DNA duplex binding region, and the free DNA duplex region revealing dynamic base pairing schemes. Together, the C•A mismatch captured at various fidelity filter sites illustrated how both accurate and mutagenic DNA replication can occur at the polymerase active site (Chapter 3).

In addition to base-pair mismatch discrimination, DNA polymerases exhibit a high degree of selectivity for deoxynucleotides over ribo- or dideoxy-nucleotides [9, 10, 87]. Ten high-resolution crystal structures of BF DNA polymerases and mutants, DNA duplex, and deoxy-, dideoxy- or ribo-nucleotide substrate were determined to elucidate the mechanism by which incorrect nucleoside sugar substrates are discriminated against. Central to this mechanism is the motion of the O helix from an open state to a fully closed state where the α-phosphate of the incoming nucleotide is aligned with the DNA primer terminus for chemical incorporation. This motion creates an ensemble of potential binding sites that bind incorrect incoming nucleotides through direct and water-mediated interactions by trapping the helix at various positions along its trajectory, misaligning their functional groups thereby compromising incorporation rates (Chapter 4).

Structures of the mismatches reveal more examples of how accurate replication is achieved. Mismatches (G•G, T•T, and T•G) trap BF polymerase at non-productive intermediate conformations and are misaligned for incorporation. Each one of these mismatch complexes is distinct in terms of the base pairing scheme, BF polymerase conformation, and the active site water structure. Intriguingly, the G•G mismatch adopts a novel “up-side-down” conformation and is stabilized at the active site by extensive direct and water-mediated interactions within the base pair and with the surround polymerase active site residues (Chapter 5).

These examples combined with the structural information from a library of BF polymerase complex structures generated previously allow us to reach a unified conclusion of structural mechanisms for nucleotide substrate selectivity by DNA polymerase (Chapter 6).

Chapter 2 Methods

Protein preparation

Cloning

In this dissertation, five BF mutants, D598A, E658A, F710Y, D598A/E658A, D598A/F710Y, were constructed using QuickChange Site-directed Mutagenesis (Stratagene, La Jolla, CA). Primers used for mutagenesis are shown in Table 2-1.

Table 2-1 Primer sequences for site-directed mutagenesis

Mutation	Primer sequence (5'-3')
D598A	GAAAGTCGTGCGACCC**GCT**ACAAAGAAGGTGCATACG
	CGTATGCACCTTCTTTGT**AGC**GGGTCGCACGACTTTC
E658A	GCCGCCGACTACTCACAAATT**GCA**TTGCGCGTCCTCGCC
	GGCGAGGACGCGCAA**TGC**AATTTGTGAGTAGTCGGCGGC
F710Y	CGTCAGGCGAAGGCGGTCAAC**TAT**GGGATCGTTTACGGG
	CCCGTAAACGATCCC**ATA**GTTGACCGCCTTCGCCTGACG

Gene expression

Transformation*:* In a 14 ml round bottom tube, add 1 μl DNA plasmid and 100 μl BL21* cells on ice for 30 min. Heat shock at 42 °C for 45 sec. Then put the tube on ice for 2 min. Add 1 ml Super Optimal Broth (SOC medium). Incubate at 37 °C for 1.5 hours. Plate 50-100 μl cell culture on a kanamycin plate and incubate 37 °C overnight.

Small-scale growth*:* Pick 1 colony into LB (100 ml) (1:50 amplification). Add 100 μl kanamycin stock (50 mg/ml) and incubate 37 °C overnight.

Large-scale growth*:* Add 40 ml overnight culture to 2 L growth with 2 ml kanamycin until OD reaches 0.6 (0.06 in nanodrop). Add 1 mM IPTG and induce at 37 °C for 3 hours.

Harvest cells: 3500 rpm, 15 min to harvest cells. Pull cell pellet out to a bottle (50 ml volume) and measure the weight. Store at -20 °C.

Protein purification

Buffer preparation:

Lysis Buffer: 50 mM Tris pH 8, 1 mM EDTA, 0.1% Nonidet P40 substitute, 0.1% Tween20, 10 mM BME.

Buffer A: 50 mM Tris pH 7.5, 1 mM EDTA, 0.1% Nonidet P40 substitute, 0.1% Tween20, 10 mM BME.

Buffer A low salt: 50 mM Tris pH 7.5, 150 mM NaCl, 1 mM EDTA, 0.1% Nonidet P40 substitute, 0.1% Tween20, 10 mM BME.

Buffer A high salt: 50 mM Tris pH 7.5, 500 mM NaCl, 1 mM EDTA, 0.1% Nonidet P40 substitute, 0.1% Tween20, 10 mM BME.

Buffer B: 50 mM Tris pH 7.5, 1 mM EDTA, 10 mM BME

Buffer B low salt: 50 mM Tris pH 7.5, 150 mM NaCl, 1 mM EDTA, 10 mM BME.

Buffer B high salt: 50 mM Tris pH 7.5, 800 mM NaCl, 1 mM EDTA, 10 mM BME.

***Cell Lysis*:** Thaw and resuspend cells in ~250 mL lysis buffer on ice. Add ~25 mg lysozyme and incubate >30 min. Lyse the cells in the cell cracker. Heat until the lysate reached 65 °C and incubate for 15 min. Cool on ice. Pellet cell debris and precipitated E. *coli* proteins by spinning at 28000 g/20 min/4 °C. Transfer the supernatant to a 25000 MWC dialysis membrane and dialyzed against 2 L buffer A overnight. Equilibrate 55 mL Q Sepharose column in Buffer A.

Purification:

Q Sepharose column: Load the protein onto Q Sepharose at 2 mL/min. Wash the column with ~8 column-volume (CV) Buffer A at 1.5 mL/min. Elute overnight in a 2CV gradient from 0-500 mM Buffer A at 1 mL/min. Wait for 45 mL for the gradient to get through the column

then began collecting 2 mL fractions. Run samples of selected fractions on 7.5% Nextgel. Pool fractions and dialyze against 2 L Buffer A +150 mM NaCl in a 25000 MWC membrane for ~2.5 hours.

Heparin column: Hook up 2×5 mL HighTrap Heparin columns in tandem and equilibrated with 150 mM buffer B. Load protein onto the Heparin column through line A at 2 mL/min and followed with ~20 mL dialysis buffer to wash all protein onto the column. Wash with Buffer B and 150 mM NaCl until the UV stabilized (~8 CV). Elute in a gradient from 150-800 mM NaCl over 100 mL (10 CV) at 1 mL/min and collect 1.5 mL fractions. Ran samples of select fractions on a 7.5% Nextgel. Pool fractions and load pooled fractions into a 25000 MWC dialysis membrane and dialyzed against 2 L Buffer B overnight.

Concentrate the dialyzed protein to reach desired concentration around 20.0 mg/ml (by nanodrop) for crystallization (0.828 extinction coefficient correction).

DNA substrate preparation

Complementary DNA primer and template are designed as shown in Table 2-2. DNA oligos are synthesized at Midland Certified Reagent Co. (Midland, TX) at GF grade (powder form) and are dissolved in DNA resuspension buffer (10 mM cacodylate•Na, 0.5 mM EDTA, 500 mM NaCl, and 10 mM $MgSO_4$) to a final concentration of 9 mM. Same volume of primer and template solution are mixed together and annealed (heat the mixture up to 85 °C for 5 min, then gradually cool 1 °C/90sec to 4 °C). The final concentration of the DNA duplex ready for crystallization is 4.5 mM.

Crystallization reagent preparation

Subtle changes in preparing crystallization reagents can make a big difference in BF crystallization and catalysis in crystal. A detailed protocol to prepare each reagent to obtain optimal BF crystals is described below.

Saturated $(NH_4)_2SO_4$: Purchased from J. T Baker (high purity). Add filtered water to excessive amount of $(NH_4)_2SO_4$ and shake vigorously. Filter (0.2 μm) before use.

MES (2-(N-Morpholino)ethanesulfonic acid) solution: 1 M pH 5.8 or 7.3. Mix 1 M MES acid (pH 3.15) and 1 M MES base (pH 9.98) solutions to obtain desired final pH. Filter (0.2 μm) and store with 0.02% Azide added.

- MES, free acid: Purchased from Calbiochem (475893), pH 3.15

- MES, sodium salt: Purchased from SIGMA (M38885-100G), pH 9.98

MgSO4: Purchased from Mallinckrodt; make 1M stock solution, add 0.02% Azide, and 0.2 μm filter.

MnSO4: Purchased from Fluka; make 1M stock solution, add 0.02% Azide, and 0.2 μm filter.

MPD (Hexylene glycol or 2-methyl-2,4-pentanediol): Purchased from Fluka. Use non-sticky pipet tips to draw up solution and wait 10 sec before removing. Clean off edge of the tip. Then eject solution slowly.

Nucleotide: Ultrapure ddNTPs and rNTPs were purchased from USB Co. (Cleveland, OH), and dNTPs from Promega Co. (Madison, WI), all at a concentration of 100 mM.

Table 2-2 DNA primer and template for crystallization

Co-crystallization to place a desired substrate at BF insertion site (*n*)		
Name	Primer (bottom) and template (top) sequences	Nucleotides and metal ions used in cocrystallization
ddXTP•dY (Running start)[a]	5′- CATYZGAGTCAGG -3′ 3′-CTCAGTCC-5′	ddXTP, Mg^{2+}/Mn^{2+}
d/dd/rXTP•dY (Cold start)[b]	5′- CATYGGAGTCAGG -3′ 3′-$_{dd}$CCTCAGTCC-5′	d/dd/rXTP, Mg^{2+}/Mn^{2+}
Catalysis in crystal to place a desired substrate into DNA duplex region		
Base pair position	Primer (bottom) and template (top) sequences	Nucleotides used in catalysis in the crystal
X•Y (*n*-1)	5′- GACGTYCGTGATCGCA-3′ 3′-GCACTAGCG-5′	d/dd/rXTP
X•Y (*n*-2)	5′- GACGTYCGTGATCGCA-3′ 3′-XGCACTAGCG-5′	dATP
X•Y (*n*-3)	5′- GACGTYCGTGATCGCA-3′ 3′-XGCACTAGCG-5′	dATP, dCTP
X•Y (*n*-4)	5′- GACGTYCGTGATCGCA-3′ 3′-XGCACTAGCG-5′	dATP, dCTP, dGTP
X•Y (*n*-6)	5′- GACGTYCGTGATCGCA-3′ 3′-XGCACTAGCG-5′	dATP, dCTP, dGTP, dTTP

[a]Running start: Z is designed to be complementary to base X. ddXTP is first chemically incorporated by BF to act as a dideoxy chain terminator to stop chemical reaction thereby trapping the next ddXTP prior to chemistry and paired with the next incoming template Y. D598A/F710Y mutant BF [62] should be used here.

[b]Cold start: pre-synthesized primer with a ddC acting as the chain terminator at the primer 3' terminus. This duplex is used to capture desired base pair placed at the wild-type polymerase insertion site.

BF crystallization protocol

Crystallization of BF, DNA duplex, and substrate ternary complex (closed form)

Mismatches and incorrect sugar substrates were captured prior to chemistry by co-crystallization utilizing a dideoxy chain terminator to prevent the reaction. BF protein used for this purpose carries the D598A mutation, which has been previously identified to destabilize a crystal contact thereby favoring the closed state in the crystal [40]. Further side chain mutations can be engineered on the D598A background. Detailed steps to obtain BF-DNA-nucleotide ternary complex (Closed form) crystals are described below and illustrated in Figure 2-1.

Protein solution preparation

- Spin BF D598A derivative protein for 2 min at 13.2 rpm at 4 °C right before use.

- DNA duplex is designed and prepared as described above.

- Add BF protein and DNA duplex at 1:3 molar ratio, with a final concentration of 150 μM and 450 μM, respectively.

- Add nucleotide substrate and divalent metal solution at 1:2 molar ratio, with a final concentration between 10-30 mM and 20-60 mM, respectively. In most cases, 10 mM nucleotide and 20 mM $MgSO_4$/$MnSO_4$ are used in crystallization. However, when partial occupancy of nucleotide is observed, a higher concentration of 30 mM nucleotide and 60 mM $MgSO_4$/$MnSO_4$ may be desirable. When adding $MnSO_4$ to a final concentration of 60 mM, the stock solution should be added gradually and be mixed well along the way to prevent possible protein precipitation.

- Incubate protein, DNA duplex, nucleotide substrate, and $MgSO_4$/$MnSO_4$ at room temperature for 30-60 min before setting up crystal trays.

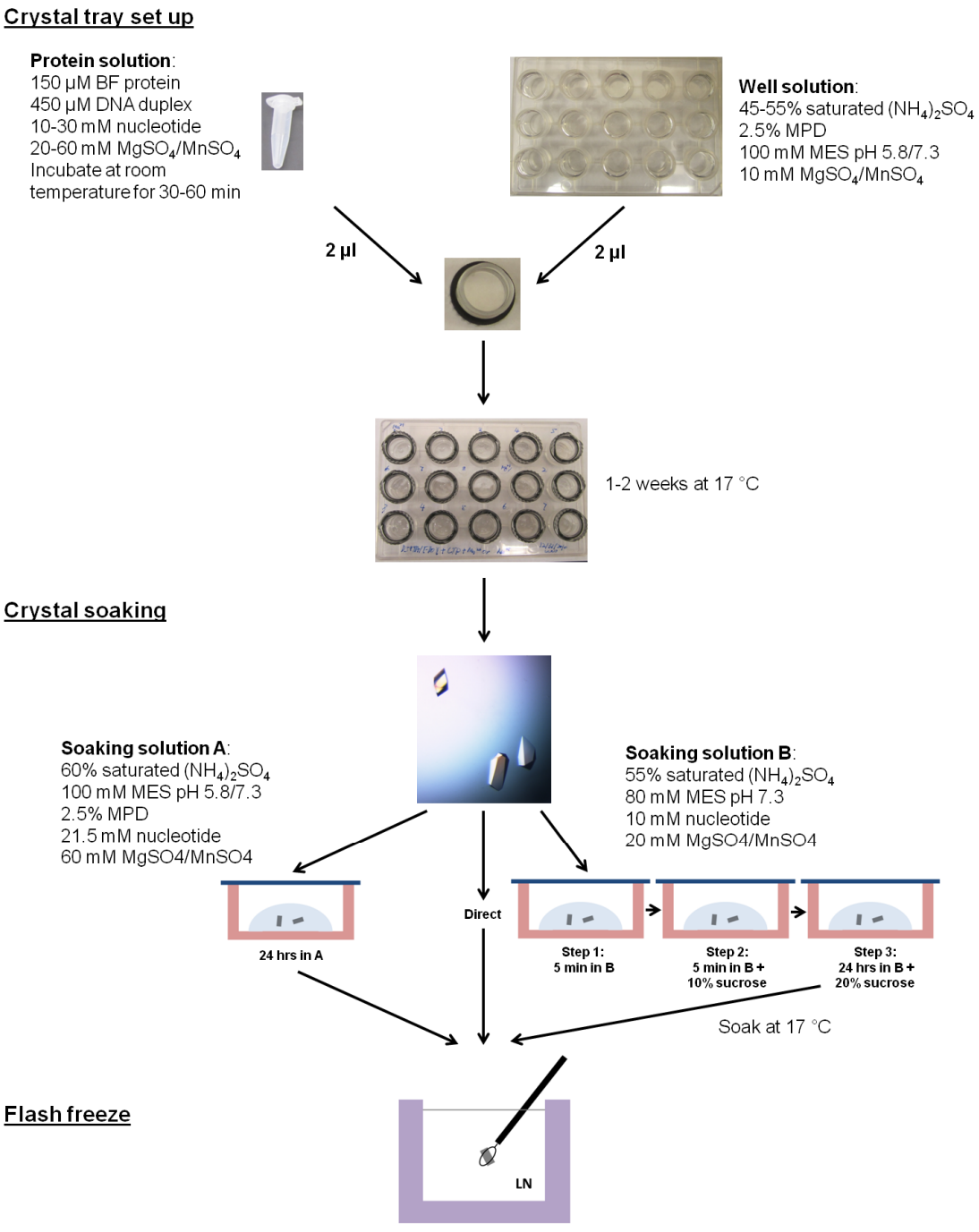


Figure 2-1 Protocol to obtain ternary complex crystals prior to chemistry.

Well solution preparation

- Well solution consists of 40-60% saturated $(NH_4)_2SO_4$, 2.5% MPD, 100 mM MES pH 5.8 or pH 7.3, and 10 mM $MgSO_4$ or $MnSO_4$.

- 500 µl well solution is normally made in a 1.5 ml eppendorf tube using stock solutions described above. Turbo mix the solution and spin at 13.2 rpm for 2 min.

Crystal tray setup

Hanging-drop protein crystallization trials are set up manually using EasyXtal 15-Well tools (X-seal) with greaseless screw-in crystallization support by Qiagen.

- Pipet 450 µl well solution into each of the wells.

- Pipet 2 µl of protein solution onto the support, then add 2 µl well solution onto the protein solution droplet without mixing.

- Immediately flip the support, place it above the well, and screw tight.

- After a tray is finished, place it on flat surface at 17 ℃.

Crystals normally appear after 5-7 days and grow to full size in two weeks. Crystals in droplets at a lower precipitant concentration may take longer to show up and to mature.

Crystal freezing

Crystals can be flash frozen in liquid nitrogen by one of the three methods described below:

- Directly out of the crystallization drop.

- After soaking in stabilization solution containing higher concentration of nucleotide substrate: 60% saturated $(NH_4)_2SO_4$, 100 mM MES pH 5.8 or pH 7.3, 2.5% MPD, 21.5 mM d/dd/rNTPs, and 60 mM $MgSO_4$/$MnSO_4$. This can be used when the nucleotide is not well ordered or is at partial occupancy.

- After soaking in a 3-step cryoprotectant solution: first, 55% saturated $(NH_4)_2SO_4$, 80 mM MES pH 7.3, 10 mM d/dd/rNTPs, 20 mM $MgSO_4/MnSO_4$ (soak for 5-10 min); second, the same solution as in step one with 10% sucrose added (soak for 5-10 min); third, the same solution as in step one with 20% sucrose added (soak for 24 hours at 17 °C). This method has been shown to shift the conformational equilibrium of the O helix.

Crystallization of BF and DNA duplex binary complex (open form)

Protein solution preparation

- Spin wild-type BF for 2 min at 13.2 rpm at 4 °C right before use.

- DNA duplex is designed and prepared as described above.

- Add BF protein and DNA duplex at 1:3 molar ratio, with a final concentration of 150 μM and 450 μM, respectively.

- Add 20 mM $MgSO_4/MnSO_4$.

- Incubate protein, DNA duplex, and $MgSO_4/MnSO_4$ on ice for 30-60 min before setting up crystal trays.

Well solution preparation and crystal tray setup

The same as described for ternary complex crystallization.

Crystal freezing

Crystals are cryoprotected in cryo-solution (60% saturated $(NH_4)_2SO_4$, 100 mM MES pH 5.8, 24% sucrose) for 5-10 min before they are flash frozen in liquid nitrogen.

Catalysis in the crystal

BF polymerase is capable of processive DNA synthesis even inside crystals [40, 60]. By designing the experimental condition for accurate (in the presence of Mg^{2+}) or mutagenic (in the

presence of Mn^{2+}) nucleotide incorporation [102, 103], this property can be utilized to capture incorrect nucleotide with mismatched base or wrong sugar moiety at various sites in the DNA duplex region.

The steps to carry out catalysis in the crystal are illustrated in Figure 2-2. Open form binary complex crystals at an approximate size of 100 μm × 100 μm × 250 μm are transferred into stabilization solution containing the right combination of nucleotides and metal as shown in Table 2-2. The soaking well is sealed using a glass slide with vacuum grease and placed at 17 °C for 24 hours to allow the reaction to occur. Crystals are frozen as described for the binary complex crystals in the previous session.

Data collection, data processing, and structure refinement

Diffraction data were collected at SIBYLS and SER-CAT beamlines and processed with XDS [104] or HKL2000 [105]. Free reflections were generated in XDSCONV [104] or CCP4 [106] by combining inherited free reflections from the starting model and 5% randomly selected reflections beyond the resolution of the starting model.

Structures were determined and refined using starting model Crystal Form II (closed conformation, 2HVI [62]) or Crystal Form I (open conformation, 1L3T, 1L5U, 1L5U, and 1L3V [40] for (*n*-1), (*n*-3), (*n*-4), and (*n*-6) positions respectively) in REFMAC5 [107] or PHENIX [108]. Model building was carried out in Coot [109]. Composite omit maps were generated in CNS [110]. All figures and superpositions were prepared in PyMOL (Schrödinger, LLC.).

Crystal Form I and II both belong to the space group P2(1)2(1)2(1). In Crystal Form I, there is one molecule in the asymmetric unit. In Crystal Form II, there are two molecules in the asymmetric unit.

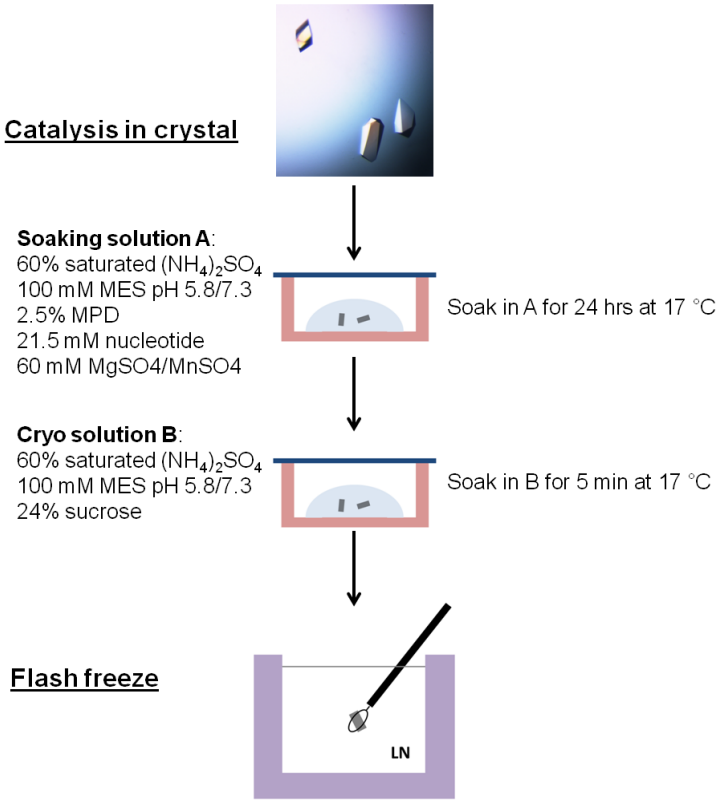


Figure 2-2 Protocol of catalysis in crystal by BF polymerase.

Chapter 3 Structural evidence for the rare tautomer hypothesis of spontaneous mutagenesis

Summary

Even though high-fidelity polymerases copy DNA with remarkable accuracy, some base-pair mismatches are incorporated at low frequency, leading to spontaneous mutagenesis. Using high-resolution X-ray crystallographic analysis of a DNA polymerase that catalyzes replication in crystals, we observe that a C•A mismatch can mimic the shape of cognate base pairs at the site of incorporation. This shape mimicry enables the mismatch to evade the error detection mechanisms of the polymerase, which would normally either prevent mismatch incorporation or promote its nucleolytic excision. Movement of a single proton on one of the mismatched bases alters the hydrogen-bonding pattern such that a base pair forms with an overall shape that is virtually indistinguishable from a canonical, Watson-Crick base pair in double-stranded DNA. These observations provide structural evidence for the rare tautomer hypothesis of spontaneous mutagenesis, a long-standing concept that has been difficult to demonstrate directly.

Introduction

High-fidelity polymerases replicate double-stranded DNA with remarkable accuracy [111]. Fidelity is achieved by a successive series of conformational changes and molecular recognition events encoded at different sites on the polymerase surface such that mismatches are either prevented from incorporating or are excised within a few nucleotides past their incorporation point [2, 20, 25, 73]. At the site of covalent incorporation, shape complementarity between the polymerase surface and the edges of correctly paired bases is the dominant

mechanism that determines specificity [72, 112]. Here, mismatched base pairs or lesions that do not conform to this stereochemical constraint misalign their incoming triphosphate moiety relative to the 3' OH of the growing primer terminus, leading to rejection of the incorrect or damaged nucleotides [20, 25, 73]. However, modified bases that maintain the stereochemistry of cognate base pair edges are readily incorporated [72, 112, 113]. Nevertheless, polymerases do incorporate mismatched nucleotide base pairs at low frequency, leading to spontaneous mutagenesis [111].

The mechanism by which spontaneous replication errors occur has long been the subject of intense speculation. In their original paper on the structure of DNA, Watson and Crick recognized that tautomerization alters the hydrogen-bonding patterns and therefore could enable mismatches to assume the structure of canonical base pairs [75]. This notion was elaborated in the rare tautomer hypothesis of spontaneous mutagenesis, which states that mutations arise through the formation of high-energy tautomers at low frequency [95, 113]. However, it has been challenging to obtain direct structural evidence for this mechanism. In the absence of polymerase, mismatches do not adopt a canonical base pair structure in DNA [2, 4]. Recently, a T•G mismatch has been observed to adopt a canonical base pair structure in a polymerase, due to an ionization event demonstrating that non-canonical hydrogen bonding pattern can arise in a polymerase [80]. Here we present the structure of a C•A mismatch in the active site of a high-fidelity DNA polymerase, the *Bacillus stearothermophilus* DNA polymerase I large fragment (*Bacillus* fragment, BF), an enzyme that has been used extensively to study the structural enzymology of nucleotide incorporation [28, 40, 60, 61].

The C•A mismatch has the advantage that only tautomers give rise to cognate base pair mimicry, whereas ionization leads to “wobble” base paring (Figure 3-1). We show that under

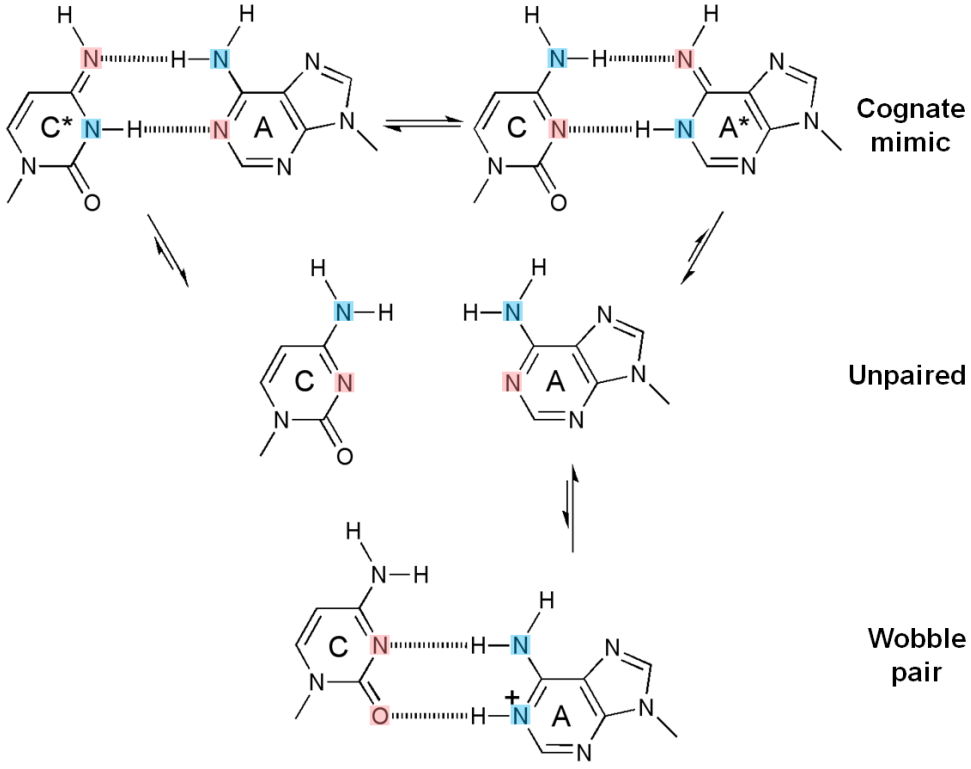


Figure 3-1 Inferred protonation states of C•A base pairs observed in the structures.
Inferred protonation states of C•A base pairs observed in the structures. In their canonical tautomeric state C and A do not pair (middle), because the two extracyclic amines clash. If either C or A tautomerizes (asterisk), a hydrogen-bonded base pair that mimicks a cognate shape can form (top). If A ionizes, a wobble base pair can form (bottom) [114, 115]. Hydrogen bond donors and acceptors are colored blue and pink respectively.

conditions which stabilize an enzyme conformation that places a nucleotide at the site of incorporation, the C•A mismatch adopts a tautomeric cognate base pair shape, whereas otherwise it forms an ionized wobble that cannot be incorporated. Wobble base pairs have been observed in isolated DNA by X-ray crystallography [114] or NMR [115]. We observe the C•A mismatch within the double helix past the position of incorporation, where it adopts a cognate base pair conformation or a wobble, depending on site location on the polymerase surface. These structures unambiguously demonstrate that tautomeric base pairs can form in the polymerase active site, providing strong support of the rare tautomer hypothesis through direct structural evidence.

Results

Five sites at which fidelity filters are encoded [37, 67, 70, 71] have been identified for BF [40] (Figure 3-2): the pre-insertion site where the incoming DNA template strand resides; the insertion site where the incoming nucleotide triphosphate pairs with the template; the catalytic site where the metals and catalytic groups are aligned; the post-insertion site where the 3′ end of the nascent duplex strand resides; and a four-base-pair DNA duplex-binding region. Recognition events at the insertion site are the most critical for replication fidelity [111, 112], and results from the binding energy arising from shape complementarity between the enzyme and cognate base pairing of the incoming nucleotide with the template strand [69], and precise alignment of the catalytic groups on the enzyme with the reactive groups on the substrates [23, 116]. This process involves concerted motions of the polymerase O helix [2, 25], base pairing of the incoming nucleotide with the template strand, and positioning of the triphosphate moiety next to

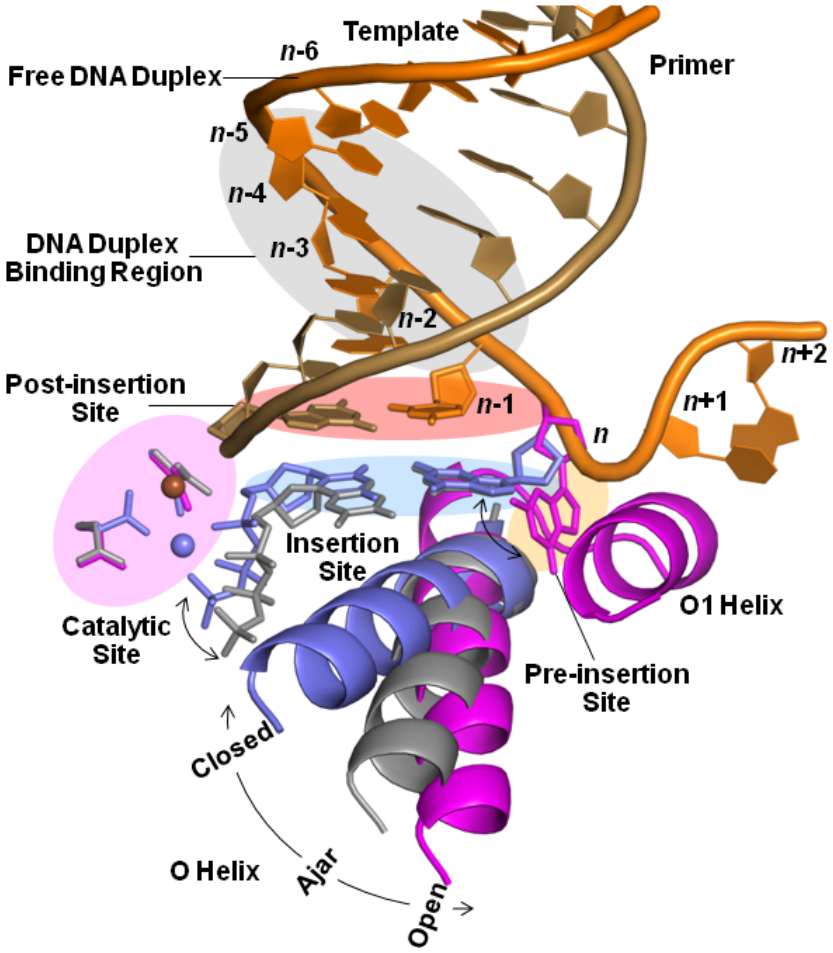


Figure 3-2 DNA polymerase replication fidelity filters.
Shaded areas correspond to fidelity filters: pre-insertion site (*n*, orange), insertion site (*n*, blue), catalytic site (magenta), post-insertion site (*n*-1, pink), and DNA duplex-binding region (*n*-2 to *n*-5, gray). DNA primer (copper) and template strands (orange) are also shown. The O helix transitions from an open (magenta) through an ajar (gray) to a closed (blue) conformation. Cognate-shaped base pairs (blue) are positioned for catalysis in the closed state. Non-canonical shapes (grey) tend to be selected against in the ajar conformation. The polymerase makes hydrogen bonds with the minor groove of base pairs positioned at sites *n*-1 to *n*-5 in the duplex-binding region following incorporation. This figure combines information derived from four structures: open (1L3U) [40], ajar (3HP6) [79], and closed (2HVI and 3EZ5) [62, 74].

Table 3-1 Crystallographic data collection and refinement statistics

	ddCTP•dA cognate (*n*)	**ddCTP•dA wobble (*n*)**	**dCTP•dA (*n*)**	**ddTTP•dA (*n*)**	**ddATP•dT (*n*)**	**ddGTP•dC (*n*)**	**C•A (*n*-1)**	**C•A (*n*-3)**	**C•A (*n*-4)**	**C•A (*n*-6)**
Data collection										
Resolution (Å)	100-1.59	100-1.58	100-1.73	100-1.52	100-1.61	100-1.62	50-1.53	100-1.65	50-1.65	50-1.60
Outer shell (Å)	1.68-1.59	1.68-1.58	1.83-1.73	1.61-1.52	1.71-1.61	1.71-1.62	1.62-1.53	1.76-1.65	1.75-1.65	1.70-1.60
R_{sym}	7.4(56.4)[a]	4.4(60.1)	5.8(58.3)	6.9(49.2)	6.6(49.3)	7.3(49.3)	3.3(48.0)	4.5(48.9)	4.6(47.0)	4.3(49.7)
$I / \sigma I$	17.6(3.2)	21.9(2.4)	18.0(2.6)	14.9(4.2)	14.2(3.6)	12.2(3.5)	26.2(4.1)	19.9(3.4)	21.2(3.4)	22.2(3.8)
Completeness (%)	99.0(99.9)	95.4(89.5)	98.3(99.6)	97.5(95.4)	97.2(91.4)	98.3(99.6)	97.9(99.6)	98.1(91.6)	97.4(99.5)	99.8(99.2)
Redundancy	8.3	5.4	5.8	7.2	5.6	5.5	4.8	4.9	5.0	5.7
Refinement										
Resolution (Å)	87.7-1.59	71.1-1.58	34.2-1.73	88.4-1.52	88.4-1.61	43.0-1.62	41.1-1.53	40.7-1.66	40.8-1.65	45.3-1.60
No. reflections	192561	188106	149632	222603	185284	180731	120984	96815	95402	111078
R_{work} / R_{free}[b]	18.5/20.7	21.1/23.9	19.5/22.4	18.5/21.1	19.3/22.4	18.7/21.5	17.0/18.5	17.6/20.2	16.7/19.0	17.8/19.2
No. non-hydrogen atoms										
Total	11542	11408	11476	12090	11962	11936	6036	5896	5988	6041
Solvent	1202	1468	1368	1823	1696	1687	873	697	775	670
B-factor	27.8	27.3	29.0	23.6	25.7	25.9	26.6	28.0	25.4	29.6
R.m.s. deviations										
Bond lengths (Å)	0.013	0.008	0.009	0.009	0.010	0.008	0.011	0.009	0.010	0.010
Bond angles (°)	1.489	1.169	1.193	1.245	1.266	1.221	1.358	1.223	1.243	1.290
Expected maximal error (Å)[c]	0.051	0.056	0.060	0.044	0.050	0.053	0.077	0.088	0.081	0.074
Expected minimal error (Å)[d]	0.009	0.010	0.013	0.008	0.011	0.012	0.011	0.019	0.021	0.014
Crystal Form[e]	II	II	II	II	II	II	I	I	I	I
PDB code	3PX6	3PX4	3PX0	3PV8	3THV	3TI0	3TAN	3TAP	3TAQ	3TAR

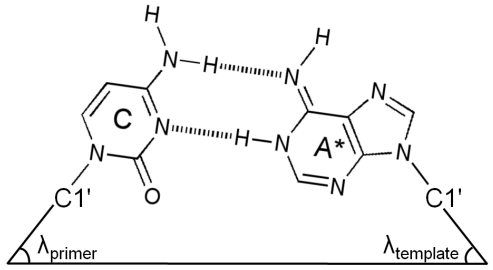
[a] Numbers in parentheses correspond to parameter values in the outer resolution shell.

[b] 5% free reflections were generated in XDSCONV [104] by combining inherited free reflections from the starting model (1L3T, 1L5U, 1L5U, and 1L3V for Crystal Form I C•A (*n*-1), C•A (*n*-3), C•A (*n*-4), and C•A (*n*-6) respectively and 2HVI for all Crystal Form II) and randomly selected reflections beyond resolution of the starting model.

[c,d] Expected maximal and minimal error were calculated in SFcheck [117].

Table 3-2 DNA base pair parameters at the insertion site and the duplex region

Base Pair	λ_{primer}(°)[a]	$\lambda_{template}$(°)[b]	$d_{C1'-C1'}$(Å)	Shear(Å)	Stretch(Å)	Stagger(Å)	Buckle(°)	Propeller(°)	Opening(°)
Insertion site									
Molecule A[c]									
C•A cognate	58.7	54.4	10.5	0.29	-0.03	0.04	5.92	-3.23	5.69
C•A wobble	70.7	51.9	10.1	1.49	-0.07	-0.01	3.45	-5.20	15.01
T•A	56.0	57.4	10.4	-0.16	-0.13	0.09	3.21	-6.16	2.97
A•T	57.7	56.7	10.4	0.08	-0.14	-0.10	2.29	-7.01	1.50
G•C	58.0	54.5	10.6	-0.06	-0.14	-0.20	-2.94	-14.66	1.60
C•G[d]	55.1	55.4	10.6	0.19	-0.15	0.15	4.66	-3.89	2.52
Watson-Crick[e]	56.7±1.4	56.0±1.3	10.5±0.11	0.01±0.15	-0.14±0.01	-0.02±0.16	1.81±3.31	-7.93±4.68	2.14±0.71
Molecule B[c]									
C•A cognate	59.6	52.4	10.4	0.62	-0.12	-0.15	6.69	- 3.31	5.80
C•A wobble	NA								
T•A	55.8	55.7	10.5	-0.02	-0.14	-0.02	7.56	-5.36	1.60
A•T	59.7	55.1	10.5	-0.03	-0.04	-0.04	2.19	-5.56	1.23
G•C	58.7	54.4	10.6	-0.18	-0.14	-0.13	-1.03	-11.96	1.56
C•G[d]	56.9	53.5	10.6	0.46	-0.09	-0.18	8.51	-4.62	3.03
Watson-Crick[e]	57.8±1.8	54.7±0.9	10.6±0.1	0.06±0.28	-0.10±0.05	-0.09±0.08	4.31±4.51	-6.88±3.41	1.86±0.80
Duplex region									
C•A(*n*-1)	64.2	44.4	10.3	1.95	-0.42	-0.27	18.44	-18.84	2.33
Watson-Crick(*n*-1)[f]	58.1±2.1	58.0±0.8	10.2±0.2	0.01±0.22	-0.11±0.09	-0.31±0.25	22.6±5.9	-6.89±2.83	4.19±1.76
C•A(*n*-3)	57.8	53.0	10.3	0.77	-0.32	-0.09	14.00	-16.83	1.00
Watson-Crick(*n*-3)[f]	57.2±0.8	56.6±0.9	10.4±0.1	0.04±0.16	-0.16±0.04	-0.01±0.17	9.87±5.01	-13.4±2.7	3.11±2.16
C•A(*n*-4)	65.8	43.7	10.6	2.13	-0.37	-0.58	12.54	6.43	1.10
Watson-Crick(*n*-4)[f]	56.4±1.1	53.3±1.4	10.7±0.1	0.13±0.23	-0.11±0.04	-0.20±0.21	5.82±2.85	7.46±2.56	-0.07±2.52
C•A(*n*-6)	64.5	45.2	10.4	2.06	-0.40	-0.17	1.08	-17.09	5.66
Watson-Crick(*n*-6)[f]	54.5±0.5	53.8±0.8	10.8±0.1	0.09±0.07	0.02±0.16	-0.15±0.11	2.06±4.71	-11.0±2.8	-0.52±0.94
C•A1[g]	64.0	45.7	10.4	-1.71	-0.45	0.22	11.06	-8.83	-2.06
C•A2[g]	72.3	49.3	10.2	2.17	-0.27	0.27	-10.28	-12.27	9.05



[a,b] λ_{primer} and $\lambda_{template}$ are defined as the angle between the glycosidic bond of primer or template nucleotide and the line drawn between the C1′ atoms of the base pair (see inserted panel on the right). $d_{C1'-C1'}$ is the distance between the C1′ atoms of the base pair. All other base pair parameters are defined [118]. All values were calculated in 3DNA [119].

[c] There are two molecules in the asymmetric unit in Crystal Form II. For C•A cognate and C•A wobble structures, molecule 1 (chains D, E, and F) is more ordered than molecule 2 (chains A, B, and C). Chain naming follows previously published structures [62, 79]. Each molecule contains the BF polymerase, DNA primer, and template. In the structure of the ddCTP•dA cognate, the O helix of molecule 2 is in the closed conformation and ddCTP adopts a near-cognate shape. In the structure of the ddCTP•dA wobble, the O helix of molecule 2 is in the open conformation, the loop between O helix and N helix is partially disordered, and there is no base pairing at the insertion site. In the structures of four Watson-Crick base pairs, the structure of molecule 2 is similar to molecule 1 with the exception of some amino acid side chain conformations.

[d] This structure was determined previously (2HVI) [62].

[e] Average values and standard deviations were calculated over all four cognate base pairs observed at the insertion site.

[f] Averaged values of cognate base pairs in previously observed complexes [40]. At the post-insertion site and *n*-3 position in the DNA duplex-binding region, average values of all four cognate base pairs (1L3S, 1L3T, 1L3U, and 1L5U) are shown. At the *n*-4 position, average values of C•G, A•T, and G•C pairs (1L3S, 1L3T, and 1L5U) are shown. At the *n*-6 position, average values of T•A, A•T, and C•G pairs (1L3S, 1L3T, and 1L3U) are shown.

[g] Values of the two C•A mismatches observed in a DNA dodecamer structure (1D99) [114].

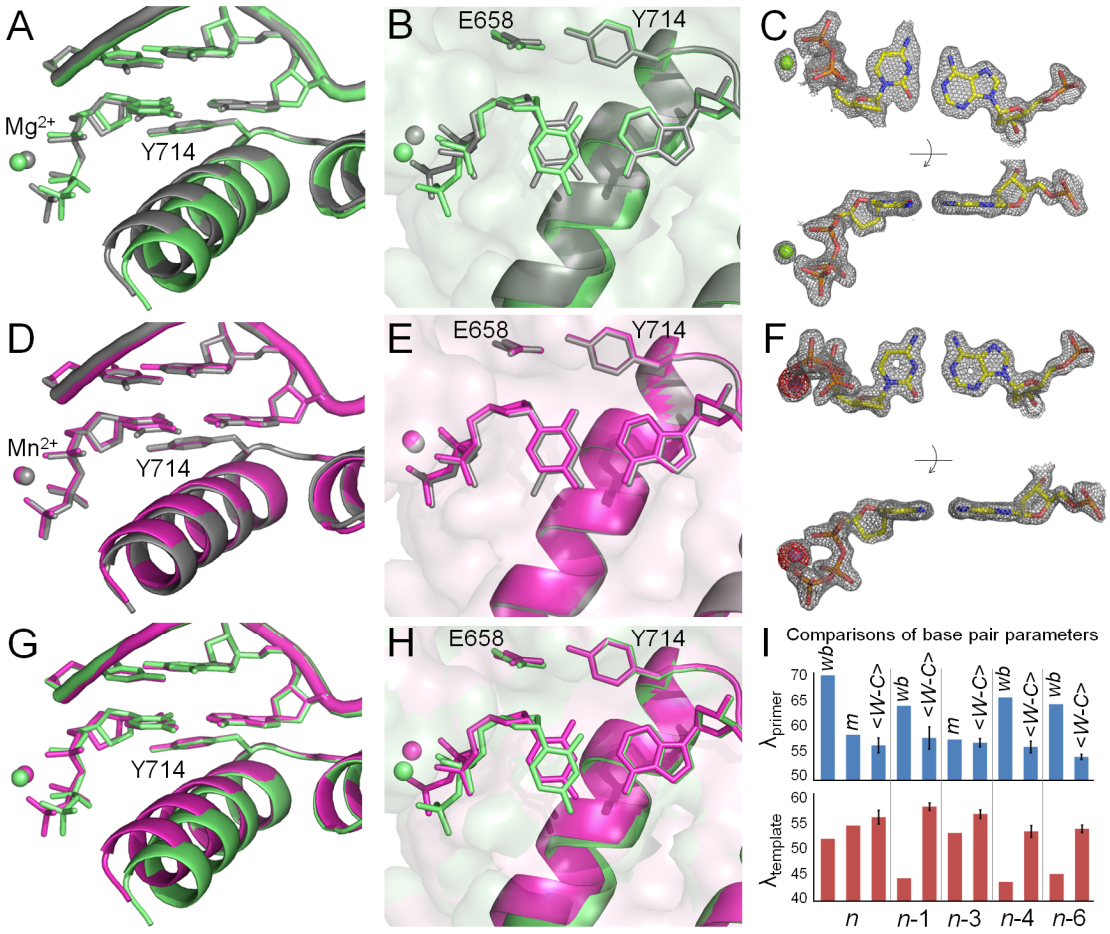
the binuclear metal center in the polymerase catalytic site and the 3′ hydroxyl of the nascent strand [23, 69].

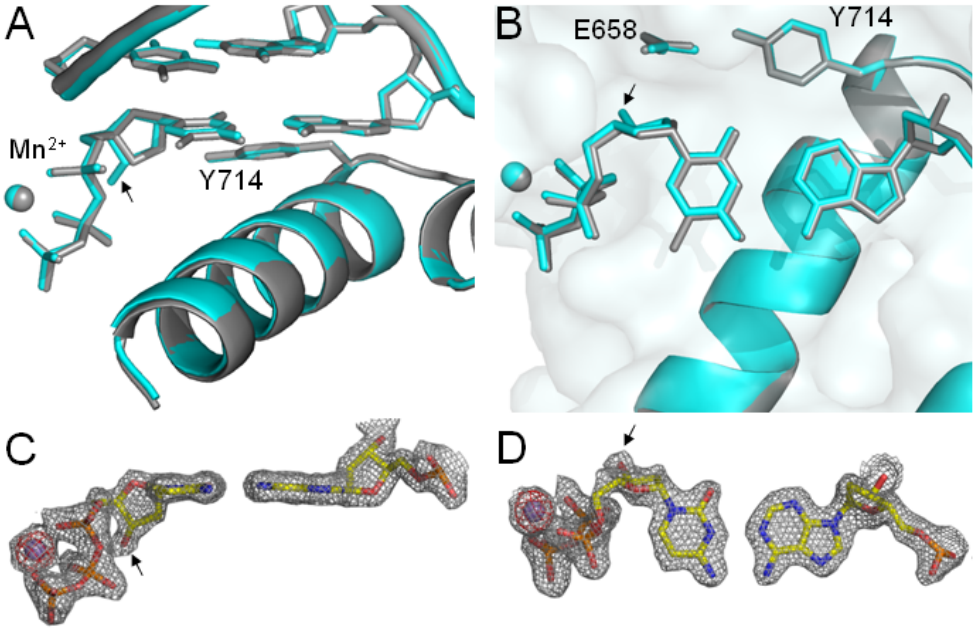
The C•A Mismatches Can Form a Cognate Base Pair Shape at the Insertion Site

Substitution of Mg^{2+} with the mutagenic Mn^{2+} ion in the active site significantly enhances the misincorporation of C•A mismatches [102, 103]. Here we compare high-resolution structures of C•A mismatches positioned at the insertion site in the presence of Mg^{2+} or Mn^{2+} (Figure 3-3 and 3-4; Table 3-1). In the 1.58Å resolution Mg^{2+} structure, we find that the C•A mismatch forms a non-cognate "wobble" (Figure 3-1) which does not match the shape of a T•A base pair placed in the same position (Figure 3-3 A-C). Furthermore, the polymerase O helix does not adopt the fully closed conformation associated with productive catalysis [37, 40, 50], but remains in an "ajar" conformation that tends to prevent non-cognate shapes from moving into the closed conformation necessary for the chemical incorporation step [79]. The structure of the triphosphate moiety also is distorted from that observed in a cognate base pair. Taken together, these differences between cognate and mismatch recognition events are expected to interfere with catalysis, preventing mismatch incorporation.

By contrast, in the presence of Mn^{2+} (observed at 1.59Å resolution), the C•A mismatch exhibits all the hallmarks of cognate base pair recognition and incorporation (Figure 3-3 D-F): shape matching, triphosphate alignment, and O helix closure. Furthermore, in this cognate conformation the cytosine O2 atom forms a hydrogen bond with a water molecule that is anchored by three polymerase side chains (Figure 3-5A). This hydrogen bond is absent in the wobble base pairs (Figure 3-5B), but its equivalent is found in all four cognate base pairs (Figure 3-5 C-F), indicating that it represents a critical feature of correctly formed base pair edge recognition at the insertion site.

Figure 3-3 Comparison of C•A mismatch and T•A cognate base pairs placed at the polymerase insertion site.
(A-B) The C•A wobble (green) and T•A base (gray) pair obtained in the presence of Mg^{2+}. For the C•A wobble pair, the O helix adopts the ajar conformation (A), the triphosphate is distorted, and the catalytic site is incompletely assembled (B). (D-E) The C•A cognate shape (magenta) obtained in the presence of Mn^{2+}. Comparison with a T•A base pair shows that the O helix is closed (D), the triphosphate is undistorted, and the active site fully assembled (E). (C, F) Two views of composite omit maps of the C•A base pair (contoured at 1.2σ, C•A wobble; contoured at 2σ, C•A cognate) (green, Mg^{2+}; purple, Mn^{2+}). The presence of Mn^{2+} is confirmed by anomalous difference map (red, contoured at 4σ). (G, H) Superposition of C•A wobble and C•A cognate at two different views showing the structural differences between the wobble and cognate conformations of this mismatch. (I) Variations of minor groove angles of C•A mismatch structures (wb, wobble; m, cognate mimic) or average cognate, Watson-Crick base pair structures (<W-C>) captured at five different positions. λ_{primer} and $\lambda_{template}$ are defined as the angle between the glycosidic bond of primer or template nucleotide and a line between the C1' atoms of the base pair. Complete tables of all nine base pair parameters are included (Table 3-2). Analysis shown here is based on molecule 1 of the two molecules in the asymmetric unit; molecule 2 is described in Table 3-2. The capture of a nucleotide at the insertion site involves the use of dideoxy analogs [62]. Additional structures were determined with a 2'-deoxycytidine triphosphate which confirms the results described here (Figure 3-4).


A
Mg2+
Y714
B
E658
Y714
C
D
Mn2+
Y714
E
E658
Y714
F
G
Y714
H
E658
Y714
I
Comparisons of base pair parameters
λprimer
λtemplate
wb
m
<W-C>
n
n-1
n-3
n-4
n-6



Base pair parameters of dCTP•dA at the polymerase insertion site

Base pair	λ_{primer}(°)	$\lambda_{template}$(°)	$d_{C1'-C1'}$(Å)	Shear(Å)	Stretch(Å)	Stagger(Å)	Buckle(°)	Propeller(°)	Opening(°)
Molecule 1									
dCTP•dA	53.8	56.6	10.6	0.03	-0.01	0.02	5.84	-1.78	4.29

Figure 3-4 Structure of a dCTP•dA base pair at the polymerase insertion site.
A complex of dCTP•dA was captured by exchanging ddCTP in ddCTP•dA crystals with dCTP. Comparisons with a cognate dideoxy base pair captured at the insertion site show that dCTP•dA adopts a similar shape. (A, B) In both deoxy (cyan) and dideoxy (grey) structures, the O helix is closed (A), the triphosphate is undistorted, and the active site fully assembled (B). (C, D) Two views of composite omit maps (gray) at 1.5σ contour of the dCTP•dA base pair showing the cognate base pair shape and 3' hydroxyl group (arrow). The presence of Mn^{2+} ion (purple), which is observed in the deoxynucleotide, is confirmed by anomalous difference map (red). The structural comparisons were based on the more ordered molecule A.

The Cognate Shape of the Mismatch Corresponds to a Tautomer

Only tautomerization of C or A results in a hydrogen-bonding pattern that enables formation of a base pair mimicking the cognate A•T shape [95] (Figure 3-1). However, a similar shape also could arise from deamination of the cytosine to form uracil, which is rare, but can occur (in the range of 10^{-10} sec^{-1} by *in vitro* measurement) [120]. We carried out mass spectrometric analysis of the crystallization drops and determined that cytosine, not uracil, is present (Figure 3-6). The mimicry of the T•A shape by a C•A mismatch is therefore the consequence of stabilizing a non-canonical tautomerized state, consistent with the original proposal of the rare tautomer hypothesis [75, 95].

Structures of the C•A Mismatch Placed at the DNA Duplex Region

Non-canonical protonation states of the C•A base pair also are observed at the other fidelity filters. At the post-insertion site and duplex DNA-binding region (Figure 3-7), in addition to steric shape complementarity, readout of hydrogen bonds in the minor groove contributes to edge recognition [69, 111]. Using successive rounds of nucleotide incorporation in the crystal [40, 60], we have placed the C•A mismatch at the post-insertion site, as well as the *n*-3 and *n*-4 sites of the DNA duplex-binding region. At all three locations the two nucleotides form base pairs, indicating formation of non-canonical protonation states. At the *n*-3 position the C•A mismatch forms a near-cognate interaction and accordingly is tautomerized (Figure 3-7D), whereas at the post-insertion (Figure 3-7 A-C) and *n*-4 (Figure 3-7E) sites a wobble is adopted corresponding to ionization events. At the post-insertion site, the wobble results in a 0.5Å displacement of the 3′ hydroxyl. Otherwise, distortions are moderate compared to other mismatches captured at this position [61]. The effects of the wobble at the *n*-4 position are also mild and do not induce a

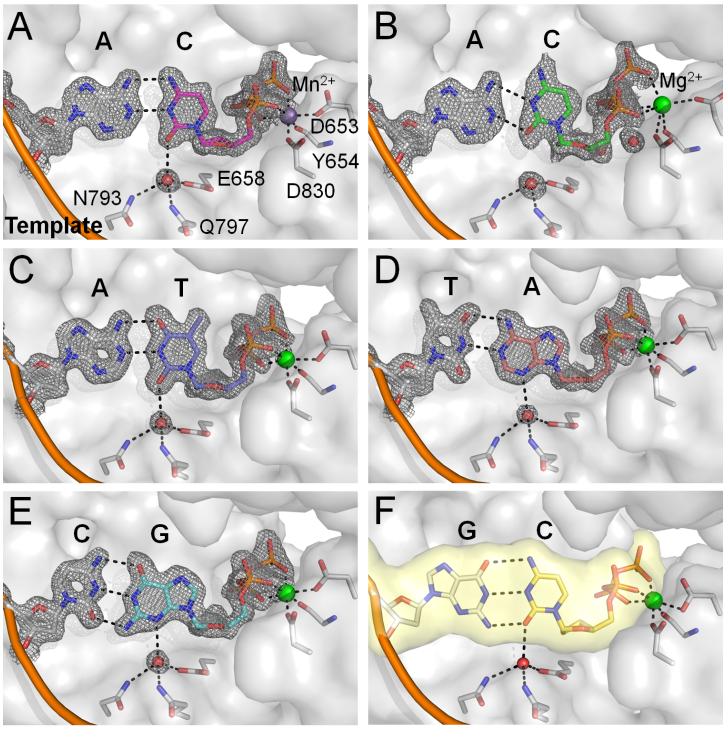
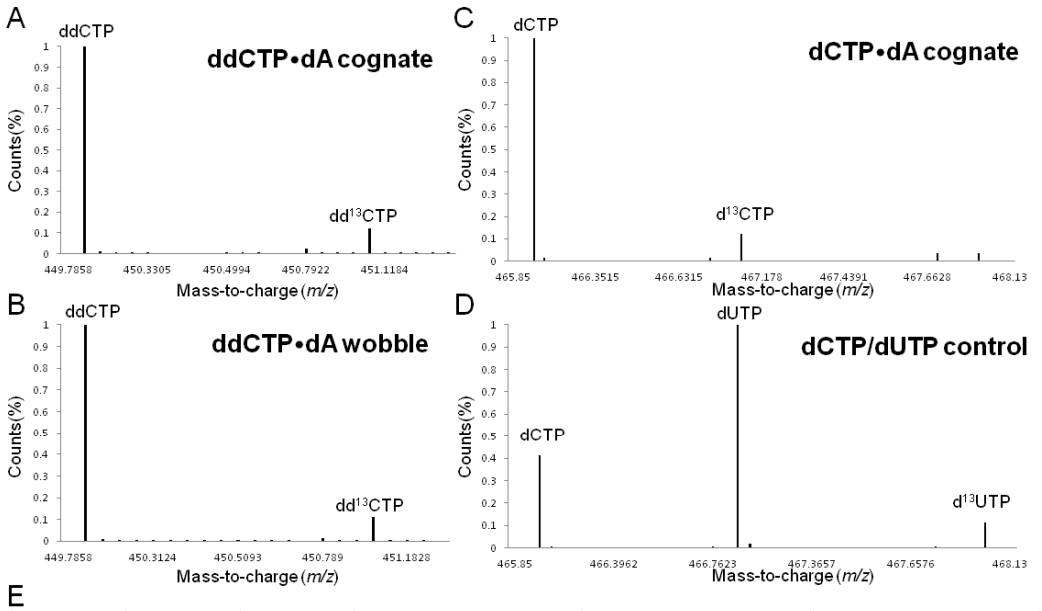


Figure 3-5 A water mediated hydrogen bond encodes edge recognition of cognate base pair shapes.

(A) C•A cognate shape mimic (Mn^{2+}, magenta); (B) C•A wobble (Mg^{2+}, green); (C) T•A; (D) A•T; (E) G•C; (F) C•G (from previously published structure; PDB code, 2HVI) [62]. Composite omit maps (gray) at 1.5σ (A, C-E) and 1.2σ (B) are shown around the base pairs and the anchored water molecule. Dashed lines (black) indicate hydrogen bonds.



E

Complex	Nucleotide	Molecular formula	Accurate mass (M-H)⁻ (all ^{12}C)		Accurate mass (M-H)⁻ (one ^{13}C)		^{13}C/^{12}C ratio (%)	
			Theoretical	Observed	Theoretical	Observed	Theoretical	Observed
ddCTP•dA cognate	ddCTP	$C_9H_{16}N_3O_{12}P_3$	449.9874	449.9883	450.9908	450.9907	11.1	12.1
	ddUTP	$C_9H_{15}N_2O_{13}P_3$	450.9714	-	451.9748	-		
ddCTP•dA wobble	ddCTP	$C_9H_{16}N_3O_{12}P_3$	449.9874	449.9880	450.9908	450.9907	11.1	11.2
	ddUTP	$C_9H_{15}N_2O_{13}P_3$	450.9714	-	451.9748	-		
dCTP•dA cognate	dCTP	$C_9H_{16}N_3O_{13}P_3$	465.9823	465.9833	466.9857	466.9846	11.1	12.1
	dUTP	$C_9H_{15}N_2O_{14}P_3$	466.9663	-	467.9697	-		
dCTP/dUTP control	dCTP	$C_9H_{16}N_3O_{13}P_3$	465.9823	465.9831	466.9857	-		
	dUTP	$C_9H_{15}N_2O_{14}P_3$	466.9663	466.9684	467.9697	467.9722	11.1	11.3

Figure 3-6 Investigation of cytosine deamination by mass spectrometry.
Mass spectrographs of the crystallization drops of ddCTP•dA cognate (A), ddCTP•dA wobble (B) crystals, soaking solution of dCTP•dA cognate crystals (C), and a solution with dUTP added to the dCTP•dA cognate soaking solution (D) are shown. Theoretical and observed masses of ddCTP or dCTP and correspondent deaminated ddUTP or dUTP are listed (E). The observed peak with the smaller *m*/*z* ratio in (A), (B), and (C) is consistent with ddCTP, ddCTP, and dCTP respectively. The peaks with the larger *m*/*z* in samples (A-C) are consistent with the correspondent ^{13}C isotope of the nucleotides respectively. The ratio of ^{13}C isotope to ^{12}C is within the error of experimental measurements (E). In the positive control containing both dCTP and dUTP (D), the peak with smaller *m*/*z* is consistent with dCTP, the intermediate peak with dUTP (d^{13}CTP is probably masked by dUTP), and the larger peak with d^{13}UTP. No detectable ddUTP or dUTP peaks were observed at correspondent *m*/*z* in the mass spectrographs of solutions where the C•A crystals were harvested from. Taken together, these results show that no detectable amount of ddUTP or dUTP is formed in the C•A crystals.

"memory effect" by transmitting distortions back to the active site, as has been observed in several mismatches [61]. At the *n*-6 position the C•A mismatch forms a wobble base pair (Figure 3-7F) similar to the structure obtained in free DNA dodecamer [114] and is no longer interacting with the DNA polymerase. These observations indicate that the altered hydrogen bonds between C and A at the filter sites arise from local interactions between the mismatch and the polymerase.

Discussion

The tautomeric form of the C•A mismatch mimics the shape of a cognate base pair in the insertion site prior to incorporation. Although this tautomeric form is of higher energy than the canonical protonation state [121], the local environment of the DNA polymerase can contribute to its stabilization in two ways: through binding interactions with features present in the tautomeric cognate stereochemistry but absent in the ground or ionized states, and by electrostatically altering the intrinsic equilibrium between the tautomers [122, 123]. The first effect is evident in the structure of the C•A mismatch (Figure 3-5). In its cognate shape, the cytosine O2 atom makes a hydrogen bond with a tightly bound water, whereas the wobble cannot present this acceptor in the appropriate geometry. This interaction is present in all four cognate base pairs placed at this site, and is equivalent to the minor groove readout mechanism that recognizes cognate base pairs at the post-insertion site and beyond [111]. This critical water is tightly bound by three residues that are highly conserved in the A-family polymerases (to which BF belongs), replaced by other hydrogen-bonding groups in the C-, X-, and Y- family polymerases, but absent in the B- family polymerases (Figure 3-8). This conservation pattern suggests that mismatch incorporation by polymerase-mediated perturbation of tautomeric equilibria also is present in polymerases other

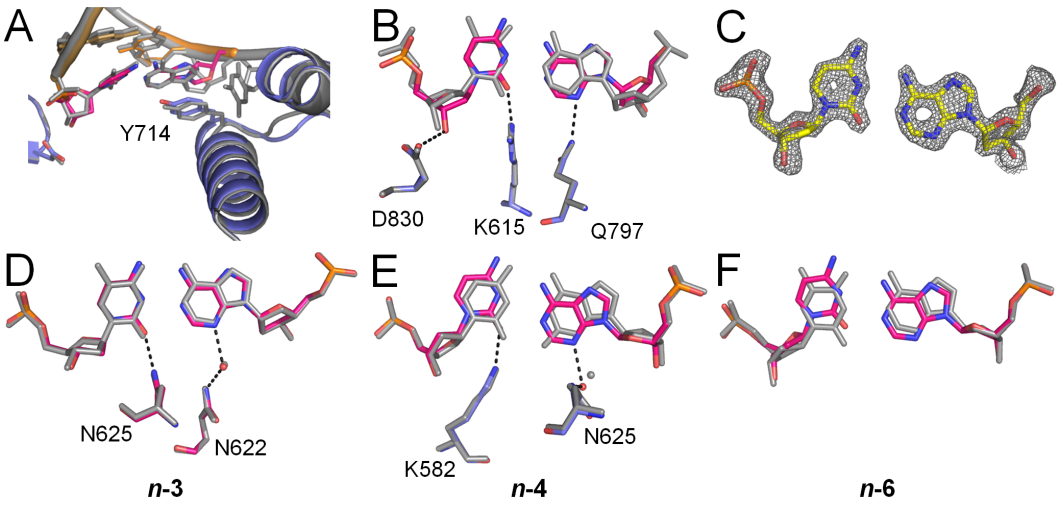


Figure 3-7 Comparison of C•A mismatch and T•A cognate base pair structures in the duplex region.

(A-C) The C•A wobble base pair captured at the post-insertion site (1.53Å resolution) showing overall structure (A), minor-groove interactions (B), and composite omit map (contoured at 1.8σ) around the mismatch (C). The next template base is disordered. (D, E) C•A adopts a near-cognate shape at the n-3 (D) position (1.65Å resolution) and a wobble shape at the n-4 (E) position (1.65Å resolution). At both positions, minor groove interactions are maintained. (F) The C•A wobble observed at the n-6 position (1.60Å resolution) where there are no contacts between the duplex DNA and the polymerase.

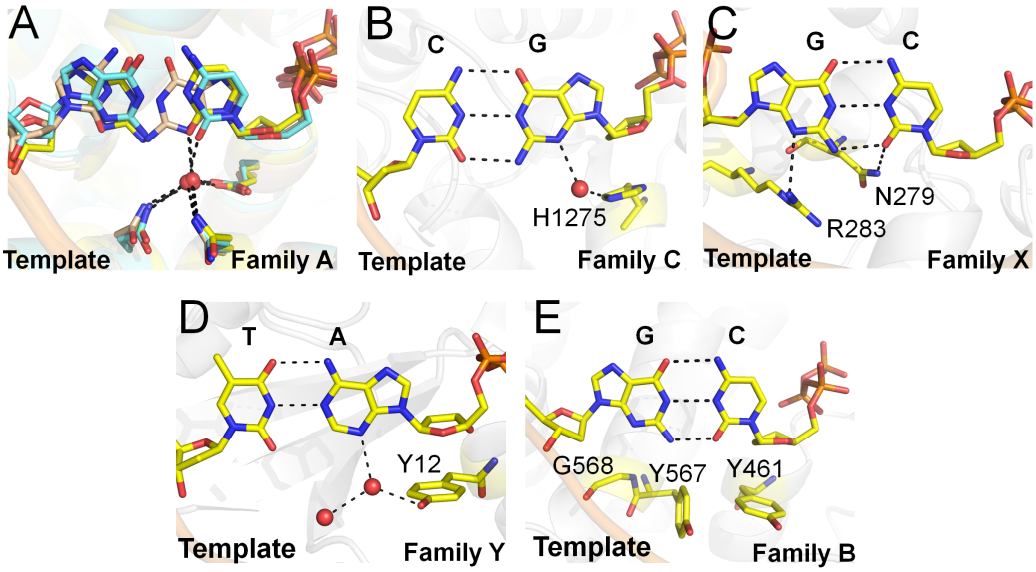
than BF, and varies with family, perhaps influencing their degree of fidelity in accordance with biological function.

The mutagenic effect of substituting Mn^{2+} for Mg^{2+} is probably also the consequence of enhancing binding to cognate base pair shapes. Mn^{2+} stabilizes the formation of the closed state relative to the ajar state (Figure 3-3) thereby enhancing the binding energy of the insertion site.

The intrinsic equilibrium of the tautomeric states is a function of the pK_a values of the two groups that change protonation states. These values are affected by the relative stabilities of dipoles within the base heterocycle, and therefore are strongly dependent on the local electrostatic environment [123]. Accordingly, the DNA polymerase could affect the spontaneous mutagenesis rate by shaping the local electrostatic field.

The results obtained here demonstrate that at least in one case spontaneous mutagenesis (prior to subsequent DNA repair processes) arises as a consequence of base tautomerization that enables a mismatch to assume the shape of a cognate base pair, consistent with the original rare tautomer hypothesis [75, 95]. Such tautomers can be stabilized by binding interactions that favor cognate stereochemical shapes, conformational equilibria that affect such binding energies, the chemical character of the residues encoding the readout, local electrostatics that alter the intrinsic tautomerization equilibria, and the chemical character of the bases themselves. These effects provide a general framework within which the observed differences in spontaneous mutagenesis frequencies of DNA polymerases, their mutants, and individual base pairs can be rationalized.

Figure 3-8 Insertion sites of representative members of five DNA polymerase families.
(A) Superposition of three members of the A family DNA polymerases: BF (yellow; PDB code, 2HVI) [62], *Thermus aquaticus* DNA polymerase I large fragment (cyan; 3KTQ) [37], and T7 bacteriophage DNA polymerase (pink;1T7P) [50]. The interactions between the water molecule and the base, and the three anchoring protein residues are conserved in all three complexes. (B) A member of the C family DNA polymerase, *Geobacillus kaustophilus* DNA polymerase PolC (3F2B) [31]. A water molecule makes similar interaction with the incoming nucleotide base which is coordinated by a single histidine instead of the anchoring side chains. (C) A member of the X family DNA polymerase, human DNA polymerase beta (2FMP) [32]. The incoming nucleotide is hydrogen-bonded directly to an asparagine side chain instead of a water. Similar interactions are also present in another member of the family, human DNA polymerase lambda (1XSN) [124]. (D) A member of the Y family DNA polymerase, *Sulfolobus solfataricus* DNA polymerase IV (Dpo4) (2AGQ) [81]. The water molecule contacting the nucleotide base is coordinated by a tyrosine instead of three anchoring residues. Similar interactions are also present in another member of the family, human DNA polymerase iota (1ZET) [125] (reviewed in [44]). (E) A member of the B family DNA polymerase, *Enterobacteria* phage RB69 DNA polymerase (3NCI) [29]. The base pair edges are read out by Van der Waals interactions only, perhaps augmented by weak electrostatic interactions mediated by the glycine and the two ring protons of tyrosine [29]. Similar interactions are also present in another member of the family, *Bacillus* phage phi29 DNA polymerase (2PYJ) [67]. The selection of the structure for a representative polymerase family was based on resolution of the ternary complex.


A
Template
Family A
B
C
G
H1275
Template
Family C
C
G
C
N279
R283
Template
Family X
D
T
A
Y12
Template
Family Y
E
G
C
G568
Y567
Y461
Template
Family B

Methods

Preparation of Protein

Wild-type and D598A/ F710Y mutant proteins were purified as described [28]. The D598A/F710Y double mutant [62] was used to capture ternary complexes. D598A destabilizes a crystal contact thereby favoring the closed state in the crystal [40]. F710Y facilitates incorporation of a 2',3'-dideoxynucleotide chain terminator [94] which prevents further incorporation and therefore traps ternary complexes before chemistry (some other members of the A family DNA polymerases have a wild-type Tyr at the equivalent position for example, T7 DNA polymerase) [88]. Wild-type protein was used to capture C•A at the post-insertion site, *n*-3, *n*-4, and *n*-6 positions.

BF Primer-template Complexes with Nucleotides Placed at the Insertion Site (Crystal Form II)

Unincorporated 2',3'-dideoxynucleoside triphosphates were trapped at the insertion site of complexes ddCTP•dA, ddTTP•dA, ddGTP•dC, and ddATP•dT by incubation of protein, primer-template duplexes (protein:DNA in a 1:3 molar ratio), dideoxynucleotides (10 mM), and Mg^{2+} or Mn^{2+} sulphate (20 mM) for 1hr at room temperature. Template sequences were designed such that a single nucleotide is incorporated to form dideoxy primer terminus thereby trapping the next nucleotide at the insertion site (Table 3-3). These reactions were used to set up crystallization as described previously [40] to obtain Crystal Form II crystals. dCTP was trapped at the insertion site by exchanging ddCTP, first soaking the crystals of ddCTP•dA (grown in the presence of both Mg^{2+} and Mn^{2+}) in a stabilization solution in the absence of nucleotides (60% saturated $(NH_4)_2SO_4$, 2.5% 2-methyl-2,4-pentanediol, 100 mM Mes pH 5.8, 30 mM $MnSO_4$ and 30 mM

$MgSO_4$) at 17 ˚C for 2d to remove the ddCTP, followed by soaking in the stabilization solution with 21.5 mM dCTP at 17 ˚C for 36 hrs to add dCTP (Figure 3-4).

BF Complexes with Mismatches Incorporated into the Primer-template Duplex (Crystal Form I)

The C•A mismatch positioned at the post-insertion site were obtained either by catalysis of primer-template in the crystal or by crystallization of a primer-template complex with a mismatch at the 3' primer terminus (Table 3-3). BF-DNA binary complexes were formed by incubating wild-type protein with the respective primer-template duplex (protein:DNA in a 1:3 molar ratio) in 20 mM $MgSO_4$ on ice for 1hr, followed by setting up the crystallization as described to obtain Crystal Form I. To incorporate a C•A mismatch by catalysis, crystals were soaked in a stabilization solution (51.5% saturated $(NH_4)_2SO_4$, 2.5% 2-methyl-2,4-pentanediol, 100 mM Mes pH 5.8) supplemented with 30 mM dCTP and 60 mM $MnSO_4$ at 17 ˚C for 24 hrs. There were no discernible structural differences between C•A mismatch positioned at the post-insertion site using either method (We present data obtained from the catalysis experiment, because it was collected to higher resolution). C•A mismatches were positioned at various sites in the DNA duplex region by adding nucleotides in various combinations and incubating in stabilization solution at 17 ˚C for 24 hrs to stimulate catalysis in the crystal of the C•A mismatch placed at the post-insertion site (obtained by either method): *n*-3, 15 mM dATP and dCTP; *n*-4, 10 mM dATP, dCTP, and dGTP; *n*-6, 7.5 mM dATP, dCTP, dGTP, and dTTP (Table 3-3).

Table 3-3 Substrates for preparing BF primer-template complexes

Nucleotides placed at the insertion site		
Name	Primer (bottom) and template (top) sequences	Nucleotides and metal ions used in co-crystallization
ddCTP•dA wobble	5′- CATAGGAGTCAGG -3′ 3′-CTCAGTCC-5′	ddCTP, Mg^{2+}
ddCTP•dA cognate	5′- CATAGGAGTCAGG -3′ 3′-CTCAGTCC-5′	ddCTP, Mn^{2+}
ddTTP•dA	5′- CATAAGAGTCAGG -3′ 3′-CTCAGTCC-5′	ddTTP, Mg^{2+}
ddATP•dT	5′- CATTTGAGTCAGG -3′ 3′-CTCAGTCC-5′	ddATP, Mg^{2+}
ddGTP•dC	5′- CATCCGAGTCAGG -3′ 3′-CTCAGTCC-5′	ddGTP, Mg^{2+}
Mismatches incorporated into the DNA duplex		
Position	Primer (bottom) and template (top) sequences	Nucleotides used in catalysis in the crystal
C•A (*n*-1)	5′- GACGTACGTGATCGCA-3′ 3′-GCACTAGCG-5′	dCTP
C•A(*n*-1) pre-synthesized	5′- GACGTACGTGATCGCA-3′ 3′-CGCACTAGCG-5′	None
C•A (*n*-3)	5′- GACGTACGTGATCGCA-3′ 3′-CGCACTAGCG-5′	dATP, dCTP
C•A (*n*-4)	5′- GACGTACGTGATCGCA-3′ 3′-CGCACTAGCG-5′	dATP, dCTP, dGTP
C•A (*n*-6)	5′- GACGTACGTGATCGCA-3′ 3′-CGCACTAGCG-5′	dATP, dCTP, dGTP, dTTP

Oligonucleotides were synthesized at GF grade from Midland Certified Reagent Co. (Midland, TX) and annealed to form duplexes as described (3). Ultrapure ddNTPs were purchased from USB Co. (Cleveland, OH), and dNTPs from Promega Co. (Madison, WI).

Diffraction Data Collection and Structure Determination

Crystals were flash frozen in liquid nitrogen either directly out of the crystallization drop (Crystal Form II) or after soaking in a cryoprotectant solution (60% saturated $(NH_4)_2SO_4$, 100 mM Mes pH 5.8, 24% sucrose) (Crystal Form I). Data were collected at SIBYLS and SER-CAT beamlines and processed with XDS [104]. Use of SIBYLS beamline at the Advanced Light Source, Lawrence Berkeley National Laboratory, was supported in part by the DOE program Integrated Diffraction Analysis Technologies (IDAT) and the DOE program Molecular Assemblies Genes and Genomics Integrated Efficiently (MAGGIE) under Contract Number DE-AC02-05CH11231 with the U.S. Department of Energy. Use of the Advanced Photon Source was supported by the U. S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. W-31-109-Eng-3.

Structures were determined and refined using starting model Crystal Form II (closed conformation, 2HVI) or Crystal Form I (open conformation, 1L3T, 1L5U, 1L5U, and 1L3V for C•A (*n*-1), C•A (*n*-3), C•A (*n*-4), and C•A (*n*-6) respectively) in REFMAC5 [107] and PHENIX [126]. Model building was carried out in Coot [109]. Data and refinement statistics are listed in Table 3-1. Composite omit maps were generated in CNS [110]. All figures and superpositions were prepared in PyMOL (Schrödinger, LLC.).

Mass Spectrometry Experiments

Crystallization drops of ddCTP•dA cognate (Mn^{2+}) and ddCTP•dA wobble (Mg^{2+}) crystals, or soaking solution of dCTP•dA cognate (Mn^{2+}, Mg^{2+}) crystals, or a positive control of dUTP added to the dCTP·dA soaking solution were analyzed using Mass Spectrometry to determine the amination state of the nucleotide (Figure 3-6). In all samples, only the expected

nucleotide was detected. Therefore the observed cognate structure of the C•A mismatch is not due to cytosine deamination.

Chapter 4 Structural factors that determine selectivity of a high-fidelity DNA polymerase for deoxy-, dideoxy-, and ribo-nucleotides

Summary

In addition to discriminating against base-pair mismatches, DNA polymerases exhibit a high degree of selectivity for deoxyribonucleotides over ribo- or dideoxy-nucleotides. It has been proposed that a single active site residue (steric gate) blocks productive binding of nucleotides containing 2' hydroxyls. Although this steric gate plays a role in sugar moiety discrimination, its interactions do not account fully for the observed behavior of mutants. Here we present ten high-resolution crystal structures and enzyme kinetic analyses of Bacillus DNA polymerase I large fragment (BF) variants complexed with deoxy-, ribo-, dideoxy-nucleotides, and a DNA substrate. Taken together, these data present a more nuanced and general mechanism for nucleotide discrimination in which ensembles of intermediate conformations in the active site trap non-cognate substrates. It is known that the active site O-helix transitions from an open state in the absence of nucleotide substrates to a ternary complex closed state in which the reactive groups are aligned for catalysis. Substrate misalignment in the closed state plays a fundamental part in preventing non-cognate nucleotide misincorpation. The structures presented here show that additional O-helix conformations intermediate between the open and closed state extremes create an ensemble of binding sites that trap and misalign non-cognate nucleotides. Water-mediated interactions, absent in the fully closed state, play an important role in formation of these binding sites, and can be remodeled to accommodate different non-cognate substrates. This mechanism may extend also to base-pair discrimination.

Introduction

DNA polymerases impose exquisite specificity for correctly paired nucleotide basepair formation through filtering mechanisms that serially select against non-cognate pairings [2, 4, 20, 25, 72, 73], thereby achieving the degree of fidelity required to maintain genomic integrity [111]. Additionally, DNA polymerases select nucleotides on the basis of their sugar identity, discriminating between the deoxyribonucleotide (dNTP) and their corresponding, 10-2,000-fold more abundant ribonucleotide (rNTP) pools on the basis of the 2' sugar hydroxyl present only in the latter [10, 83-85]. In addition to imposing the logic for the flow of genetic information [127], separating the synthesis of RNA and DNA also is necessary to prevent misincorporation of otherwise correctly paired ribonucleotides, which leads to genome instability [86]. Ribonucleotide misincorporation possibly is the most common replication error in initial DNA synthesis [85] or subsequent gap filling [10], slowing down replication [85, 87] and increasing susceptibility to strand cleavage. DNA polymerases also recognize the 3'-OH which is absent in synthetic dideoxyribonucleotides (ddNTPs) that are used for DNA sequencing and some antiviral drugs [21]. Understanding the mechanism by which DNA polymerases discriminate between different nucleotide pools based on the sugar moiety hydroxylation state is therefore both of fundamental and technological interest. Insights may aid in the development of therapeutics and synthetic RNA polymerases that enable RNA synthesis to be initiated using oligonucleotides rather than promoters [128, 129].

Although some specialized DNA polymerases are non-selective (e.g. human Pol μ, human terminal DNA transferases), most DNA polymerases exhibit appreciable preference for dNTP over rNTP incorporation, with widely varying selectivity coefficients, that range from ~10^2-~10^6 [10]. Mutagenesis studies on both high- [87, 89, 90, 130-135] and low-fidelity [136, 137] DNA polymerases have shown that mutation of a single residue in the active site

significantly lowers the sugar-based selectivity [10]. Molecular modeling suggested that this residue could interfere with binding of the rNTP 2'-OH, and that reduction of its steric bulk accounts for the lowering of specificity in these mutants [9, 87]. Structural studies of a high- [90] and low-fidelity [138] DNA polymerase are consistent with this "steric gate" hypothesis.

Although the putative steric gate residue plays an important role in sugar specificity, the steric gate hypothesis does not account fully for the observed behavior of the mutants. First, in all cases, the decrease in sugar-based nucleotide selectivity is accompanied by a loss of dNTP incorporation rates (Figure 4-1), indicating that the effects of the steric gate mutation on dNTP and rNTP incorporation are not independent. Second, even in the mutants, dNTP remains favored over rNTP incorporation [10]. Taken together these observations suggest that the steric gate residue does not act as a straightforward block of 2'-OH binding, as postulated by the steric gate hypothesis, but interacts with other components in the active site that affect nucleotide incorporation. Based on 1.58-1.95Å structures of wild-type and mutant *Bacillus* DNA polymerase I large fragment (*Bacillus* fragment (BF)) [28, 40, 60-65, 79, 101] ternary complexes with dNTP, rNTP, and ddNTP, we show that in addition to the steric gate, motions of a critical subdomain ("fingers domain") formed by the O helix play a major role in determining specificity by setting up a series of intermediate structures that can trap non-cognate nucleotides prior to formation of the catalytically active closed state. Furthermore, we show that the E568A steric gate mutation results in re-arrangements in the active site that can account for the consistent loss of dNTP incorporation.

DNA can be replicated in BF crystals, which has enabled many stages of the polymerase catalytic cycle to be captured crystallographically with a variety of substrates [40, 60-62, 65, 79, 101]. Conformational changes are key to the mechanism by which BF imposes a high degree of specificity. Briefly, dNTPs are incorporated at the insertion site in the "closed" conformational

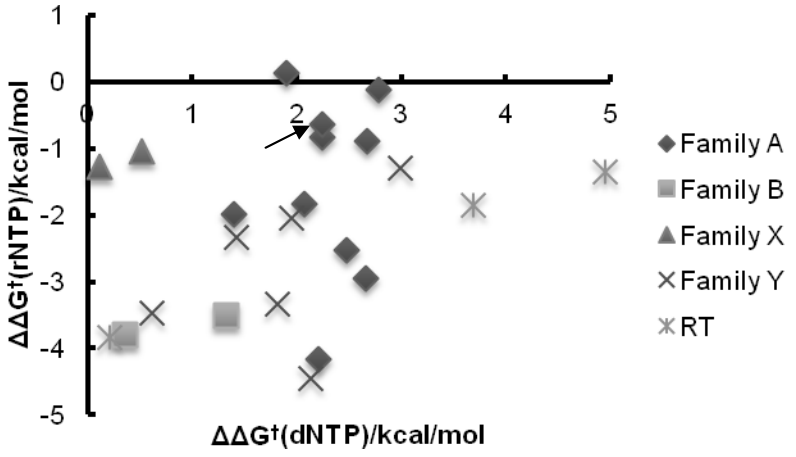


Figure 4-1 Plot of $\Delta\Delta G^{\dagger}$(rNTP) vs $\Delta\Delta G^{\dagger}$(dNTP) for DNA polymerases from five families with wild-type and "steric gate" mutant kinetic data available.
Data point for *Bacillus* DNA polymerase I large fragment is marked with an arrow. Original data and references are shown in Table 4-1.

$$\Delta\Delta G^{\dagger}(\mathrm{NTP}) = -\mathrm{RT}\ln\left(\left({}^{Mutant}_{NTP}k_{cat}/{}^{Mutant}_{NTP}K_m\right)/\left({}^{WT}_{NTP}k_{cat}/{}^{WT}_{NTP}K_m\right)\right)$$

Table 4-1 Summary of kinetic data of DNA polymerases from five families

	Protein	X	Selectivity		Loss (dXTP)[a]	Gain (rXTP)[b]	$\Delta\Delta G^{\dagger}$ (dXTP)[c]/ kcal/mol	$\Delta\Delta G^{\dagger}$ (rXTP)[c]/ kcal/mol
			WT	Steric gate mutant				
Family A								
	E. coli Klenow fragment [87]	C	3400	4.3	34	23	2.1	-1.8
		dTTP/rUTP	1700000	33	43	1200	2.2	-4.2
	Taq Pol I [130]	G	29000	5.3	59	91	2.7	-2.9
		A	24000	130	8.5	21	1.4	-2.0
		C	22000	10	45	48	2.5	-2.5
	Human Pol γ [89]	A	9300	68	38	3.8	2.2	-0.82
		G	1100	66	22	0.8	1.9	0.14
		C	6600	62	92	1.2	2.8	-0.11
		dTTP/rUTP	77000	222	76	4.3	2.7	-0.90
	Bacillus fragment (this study)	C	24000	170	46	3.0	2.2	-0.64
Family B								
	RB69 [131]	C	64000	10	9.4	370	1.3	-3.5
	φ 29 [132]	dTTP/rUTP	4400000	2300	1.9	>1000	0.35	-3.8
Family X								
	Human Pol λ [133]	dTTP/rUTP	3000	360	1.2	7.8	0.11	-1.3
	Human Pol β [90]	C	8200	670	2.3	5.5	0.51	-1.1
Family Y								
	Dpo4 [136]	G	18333	4	32	1367	2.1	-4.4
		C	5500	12	10	44	1.4	-2.3
		A	13448	3	19	224	1.8	-3.3
		dTTP/rUTP	20500	30	2.7	275	0.61	-3.5
	Dbh [137]	G	3400	3.7	28	33	1.95	-2.1
		C	7700	5	164	9.1	3.0	-1.3
Reverse Transcriptase								
	HIV-1 RT [134]	A	140000	5	3064	9.1	4.9	-1.4
		dTTP/rUTP	120000	15	397	20	3.7	-1.8
	MMLV RT [135]	dTTP/rUTP	16000	23	1.4	500	0.21	-3.8

[a] $\text{Loss (dXTP)} = ({}^{WT}_{dXTP}k_{cat}/{}^{WT}_{dXTP}K_m)/({}^{Mutant}_{dXTP}k_{cat}/{}^{Mutant}_{dXTP}K_m)$

[b] $\text{Gain (rXTP)} = ({}^{Mutant}_{rXTP}k_{cat}/{}^{Mutant}_{rXTP}K_m)/({}^{WT}_{rXTP}k_{cat}/{}^{WT}_{rXTP}K_m)$

[c] $\Delta\Delta G^{\dagger}(\text{NTP}) = -RT\ln({}^{Mutant}_{NTP}k_{cat}/{}^{Mutant}_{NTP}K_m)/({}^{WT}_{NTP}k_{cat}/{}^{WT}_{NTP}K_m)$

state of the DNA polymerase. In this form, the incoming base is paired with the template strand, its α-phosphate is aligned with the 3'-OH terminus of the primer strand in the presence of Mg^{2+} which enters the active site together with the dNTP. In the absence of dNTPs, the polymerase adopts an "open" conformation. These two conformational states are determined primarily by the motion of the O helix in the presence of dNTPs, which swings from the "open" state [37, 40], through intermediates that in some instances can be captured crystallographically [79, 101] to a "closed" form [37, 40, 50]. This motion combines bending of a hinge region at the O helix C-terminus (residue G711) with deformations within the helix [79]. Base pair mismatches are selected against by a variety of filters that distinguish correctly from incorrectly paired nucleotides by their hydrogen-bonding pattern or steric shape [61, 79, 101]. Prior to incorporation, the α-phosphate of mismatched nucleotides or lesions misaligns with the primer terminus 3'-OH, thereby reducing or preventing the chemical incorporation step [62, 79, 101]. Interference with the motion of the O helix is an important component of these misalignment processes: incorrect dNTPs positioned at the insertion site form only partially closed O helix conformations [79, 101]. Following their incorporation, mismatches or lesions introduce distortions of the template strand and alter the alignment of the 3'-OH to interfere with subsequent incorporation events [61, 63-65]. Here we report that in the sugar moiety discrimination process, O helix motions also are interfered with, resulting in non-productive conformations that differ from those observed for mismatches or lesions.

Results

Nucleotide incorporation selectivity − Nucleotide affinities and incorporation rates of dCTP, rCTP, and ddCTP were determined for wild-type BF and its E658A and F710Y mutants by pre-steady state kinetics (Table 4-2). These values are consistent with those determined for

other DNA polymerases in the A family [82, 87, 89, 94]. Wild-type BF favors dCTP incorporation over rCTP or ddCTP by 4 and 3 orders of magnitude respectively. The E658A mutant decreases the selectivity against rCTP by two orders of magnitude, primarily as a consequence of a 100-fold loss in the dCTP catalytic rate, consistent with the behavior of this mutation in other members of the DNA polymerase A family (Figure 4-1). The F710Y mutation renders incorporation of ddCTP essentially as efficient as dCTP, and slightly boosts the catalytic rate of the latter. This mutation also improves binding of rCTP, but compromises its incorporation rate by a similar degree.

Nucleotide positioning, O helix motion, and water structure − High-resolution structures of dCTP, rCTP, ddCTP complexes were obtained for the wild-type active site BF and the F710Y and E658A mutants using a primer lacking a 3'-OH of its terminus (Table 4-3). In all crystallized ternary complexes, a surface D598A mutation was included to destabilize a crystal contact that favors the open state in the crystal lattice [40]. The resulting crystal form contains two molecules in the asymmetric unit. In all complexes, molecule **1** is more ordered than **2**; accordingly, our comparative analysis is based on molecule **1**. In all complexes the nucleotide is positioned at the insertion site and forms hydrogen bonds with dG on the template strand. Mg^{2+} and Mn^{2+} were used in the rCTP complexes to determine the location of bound metal by anomalous scattering of the latter.

In the wild-type active site, the rCTP (Figure 4-2) and ddCTP (Figure 4-3) nucleotide triphosphates are displaced from the dCTP position. Consequently, the rCTP or ddCTP α-phosphates are not aligned properly with the primer strand terminus, consistent with the diminished k_{pol} rate of these nucleotides relative to dCTP. The structure of the rCTP complex with

Table 4-2 Pre-steady-state kinetic constants for use of dNTPs, rNTPs, and ddNTPs by BF polymerase and mutant derivatives

	K_d, µM			k_{pol}, s-1			Selectivity*	
Protein	dCTP	rCTP	ddCTP	dCTP	rCTP	ddCTP	dCTP/rCTP*	dCTP/ddCTP
wt	33.2	1,800	134	52.1	0.12	0.049	24,000	4,400
E658A	16.2	753		0.56	0.15		170	
F710Y	36.6	674	3.5	115	0.046	26.5	46,000	2.4

* Defined as the ratio of k_{pol}/K_d for dCTP vs rCTP incorporation opposite template dG.

Table 4-3 Crystallographic data collection and refinement statistics

	1 dCTP•dG (*n*)	**2 ddCTP•dG (*n*)**	**3 rCTP•dG Mg^{2+} (*n*)**	**4 rCTP•dG Mn^{2+} (*n*)**	**5 F710Y-rCTP•dG Mg^{2+} (*n*)**	**6 F710Y-rCTP•dG Mn^{2+} (*n*)**	**7 E658A-rCTP•dG Mg^{2+} (*n*)**	**8 E658A-rCTP•dG Mn^{2+} (*n*)**	**9 E658A-DNA (dCTP)**	**10 WT-rC•dG (*n*-1)**
Data collection										
No. complexes in asymmetric unit	2	2	2	2	2	2	2	2	1	1
Resolution (Å)	100-1.69	100-1.74	100-1.68	100-1.68	100-1.67	100-1.66	100-1.59	100-1.95	100-1.58	100-1.66
Outer shell (Å)	1.79-1.69	1.84-1.74	1.78-1.68	1.78-1.68	1.77-1.67	1.76-1.66	1.69-1.59	1.98-1.95	1.67-1.58	1.76-1.66
R_{sym} (%)	6.6(49.1) [a]	7.7(49.0)	6.2(46.6)	6.7(48.0)	5.8(49.3)	5.9(49.1)	5.7(47.8)	9.2(68.7)	5.9(46.7)	4.1 (48.9)
I / σI	14.6(3.3)	15.1(4.4)	21.1(4.2)	21.4(4.0)	21.5(3.7)	19.3(3.6)	26.3(4.1)	20.7(3.9)	24.7(4.3)	22.1 (3.3)
Completeness (%)	90.2(94.9)	89.8(100)	89.6(97.2)	90.9(94.4)	91.2(99.8)	90.6(93.4)	98.7(92.3)	94.5(100)	98.8(98.9)	99.8 (99.9)
Redundancy	5.2(5.0)	7.2(7.4)	9.9(5.9)	10.8(6.2)	9.1(5.6)	9.1(5.4)	9.4 (5.9)	6.2(6.2)	10.0(5.5)	4.8(4.8)
Refinement										
Resolution (Å)	28.8-1.69	29.4-1.74	54.4-1.68	64.3-1.68	79.7-1.67	71.1-1.66	79.4-1.60	37.4-1.95	67.3-1.58	45.4-1.66
No. reflections	148421	138100	152366	153529	156215	156480	192615	105634	112903	103589
R_{work} / R_{free}	16.2/18.9	17.2/20.1	18.0/20.3	17.7/20.1	18.1/20.6	18.2-20.3	16.5/18.5	17.2/20.5	16.4/18.5	16.6/18.8
No. non-hydrogen atoms										
Protein	9570	9426	9233	9233	9112	9112	9500	9459	4697	4673
DNA	850	850	772	772	790	790	832	832	365	448
Ligands & ions	58	56	−−	−−	29	29	30	60	−−	27
Solvent	1642	1327	1156	1140	1286	1187	1272	862	861	857
B-factor										
Protein	24.3	31.3	33.8	34.5	32.6	33.4	27.1	36.5	22.5	26.3
DNA	22.3	27.5	31.3	32.3	29.0	30.2	29.0	33.9	33.1	57.2
Ligands & ions	17.5	23.9	−−	−−	23.2	29.1	21.5	30.5	−−	20.3
Solvent	34.4	37.0	38.8	40.0	38.6	39.1	36.8	39.2	37.6	38.2
R.m.s. deviations										
Bond lengths (Å)	0.011	0.012	0.011	0.011	0.011	0.011	0.011	0.011	0.010	0.010
Bond angles (°)	1.334	1.364	1.334	1.336	1.366	1.372	1.352	1.338	1.344	1.292
Ramachandran plot										
Allowed (%)	100	100	100	100	100	100	100	100	100	100
Favored (%)	98.1	98.2	97.7	98.0	98.0	98.0	98.1	98.1	97.9	98.4
PDB code	4DQI	4DQP	4DS5	4DS4	4DSE	4DSF	4DQQ	4DQR	4E0D	4DQS

[a] Numbers in parentheses correspond to values in the outer resolution shell.

wild-type BF is complicated by the presence of a mixture of two rCTP conformations, both with misaligned α-phosphates; this situation is simplified in the rCTP complex of the F710Y variant (which retains a strong preference for dCTP over rCTP; Table 4-2) in which only one, misaligned conformation is observed (Figure 4-2). In all misaligned rCTP or ddCTP complexes the O helix fails to adopt the fully closed conformation (Figure 4-2, A and E; Figure 4-3, A and E). Furthermore, these complexes contain a water layer that is absent in the fully closed wild-type dCTP and F710Y ddCTP complexes (Figure 4-4). A water layer also is present in the fully open (PDB code: e.g., 1L3S [40]) and ajar conformations (PDB codes: 3HP6 [79] and 3PX4 [101]); its displacement is therefore part of the mechanism by which improper nucleotides are selected against.

ddCTP − The failure of ddCTP to displace the water layer present in the open state, with the accompanying lack of O helix closure and misalignment of the α-phosphate, is accompanied by the loss of two hydrogen bonds, formed between dCTP 3'-OH and its own β-phosphate oxygen and the backbone amide of E658 (Figure 4-3, C and D). These lost hydrogen bonds are compensated for in the F710Y mutant which is located on the O helix surface facing the bound nucleotide. In the Y710-ddCTP complex, the mutant phenolic hydroxyl substitutes for the 3'-OH to interact with the β-phosphate of bound ddCTP and backbone amide of E658 (Figure 4-3G). As a consequence, the ddCTP affinity improves sufficiently (~40-fold; Table 4-2) to displace the water layer and close the O helix (Figure 4-3, E and F; Figure 4-4C). The F710Y-ddCTP complex is essentially indistinguishable from the wild-type dCTP complex, accounting for their near-identical incorporation rates. We note that Tyr 710 is a natural variant in a number of family A DNA polymerases, for example T7 DNA polymerase [88].

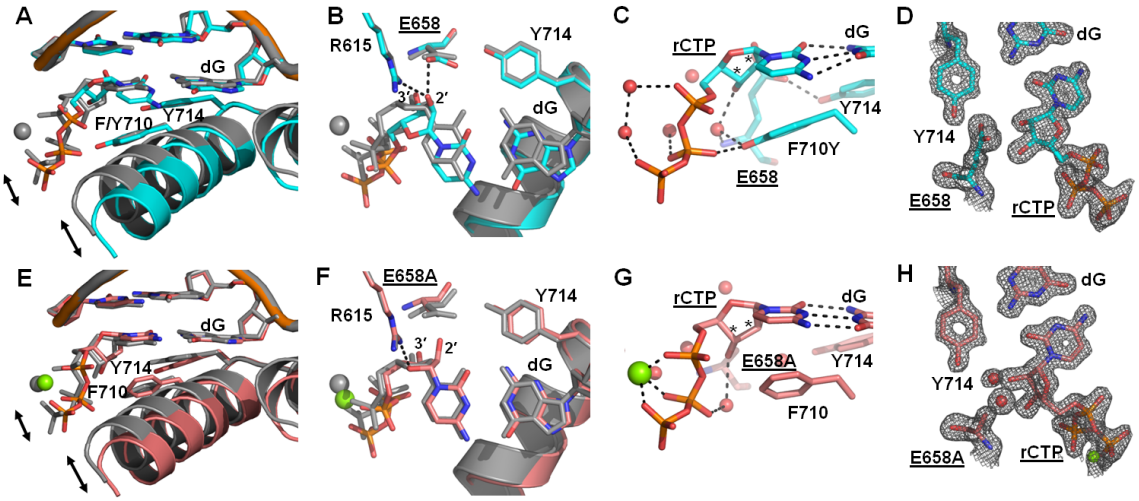


Figure 4-2 Comparison of rCTP and dCTP complexes.
A and B, two views comparing the F710Y BF mutant complex of rCTP (blue) with the wild-type complex of dCTP (gray). In the rCTP complex, the O helix does not fully close and the nucleotide triphosphate is misaligned (arrows emphasize structural differences). The interaction of Glu658 with rCTP 2′-OH contributes to the substrate misalignment. Arg615 forms a hydrogen bond with rCTP 2′-OH. The rCTP forms hydrogen bonds with the dG template which also moves; the base pair parameters are distorted from their canonical values (Table 4-4) (B). C, distortions of the rCTP triphosphate are stabilized by two water molecules such that Mg^{2+} fails to bind. E and F, two views comparing the rCTP complex of E658A (pink) with the wild-type dCTP complex (gray). E, arrows indicate rCTP misalignment and the failure of the O helix to close. F, the E658A mutation enables the rCTP base and sugar to occupy similar positions as observed for dCTP. Arg615 forms a hydrogen bond with the O4 atom of rCTP sugar as is observed in other complexes. G, the triphosphate conformation remains altered, but binds the metal ion at partial occupancy. The template base is not distorted and forms canonical hydrogen bonding with rCTP. D and H, composite-omit-maps (gray mesh) contoured at 1.5σ level. Water atoms (red) and Mg^{2+} ion (green or gray) are shown as spheres.

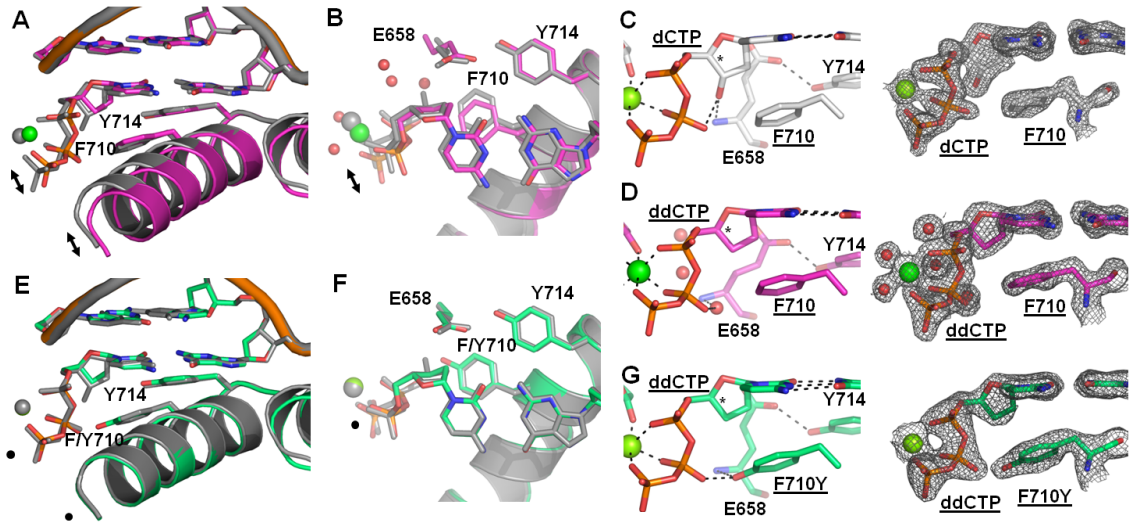


Figure 4-3 Comparison of ddCTP and dCTP complexes.
A-B, two views comparing wild-type BF complexes of ddCTP (magenta) and dCTP (gray). The ddCTP position is displaced and the O helix fails to close (arrows emphasize structural differences). E and F, comparison of the F710Y variant ddCTP complex (green) with wild-type BF dCTP complex (gray). Other than the absence of the 3'-OH, the two complexes are indistinguishable from one another. C, D, and G, close-up views of the active sites of the three ternary complexes illustrating critical interactions (left column) and their electron density maps (right column). C, the dCTP 3'-OH forms hydrogen bonds with its own β-phosphate oxygen and the Glu658 backbone amide. These two interactions are lost in ddCTP complex (D). Analogous interactions are formed by the Tyr710 phenolic hydroxyl in the F710Y mutant (G). Water atoms (red) and Mg^{2+} ion (green) are shown as spheres. Composite-omit-maps (gray mesh) are contoured at 1.5σ level.

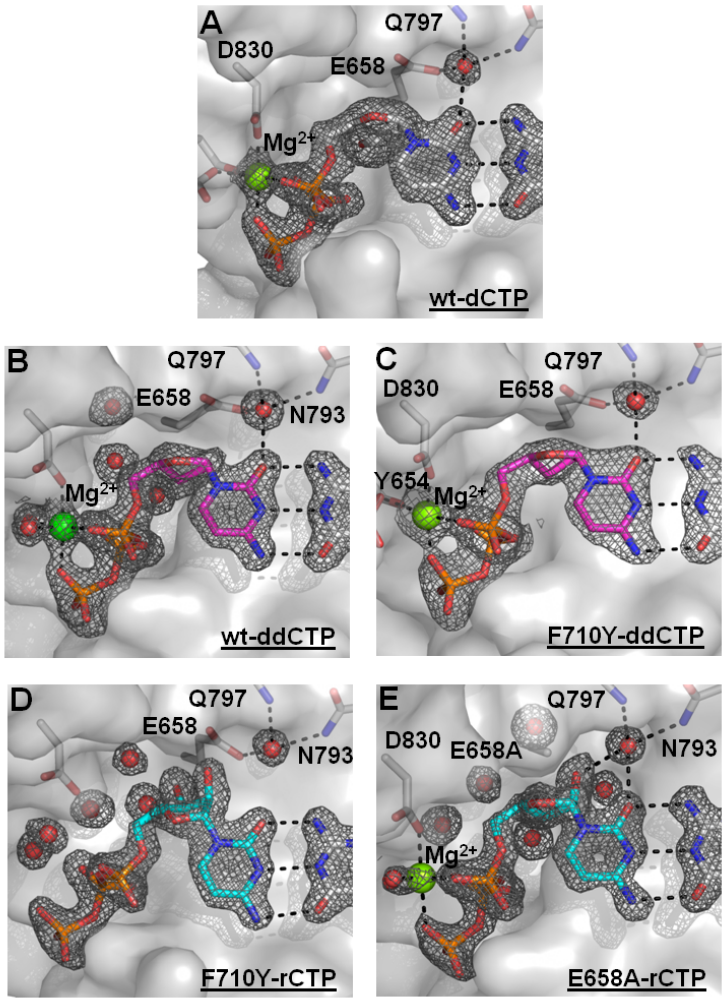


Figure 4-4 Water structure in the active site.
Nucleotide substrates, water molecules, and metal ions are shown for the wild-type active site bound to dCTP (A) and ddCTP (B); the F710Y variant bound to ddCTP (C) and rCTP (D); and the E658A variant bound to rCTP (E). The water layer is displaced in complexes where the nucleotide is properly aligned (A, C), but otherwise is present (B, D, E). In addition, a water molecule anchored by three conserved protein side chains [101] makes a hydrogen bond with the minor groove O2 atom of the cytosine base in all complexes except F710Y rCTP where rCTP is displaced. In E658A mutant structure, the anchoring interaction made by E658 is replaced by rCTP 2'-OH (E). Composite-omit-maps (gray mesh) of all complexes are shown around the highlighted features (all contoured at 1.5σ level).

rCTP − In the wild-type enzyme in the presence of Mg^{2+} or Mn^{2+}, rCTP adopts an ensemble of conformations, and the O helix exhibits an ensemble of states, none of which correspond to the fully closed form. This ensemble is simplified in the F710Y mutant in which both the bound rCTP and the O helix adopt a single, misaligned, non-closed conformation with either metal (Figure 4-2A). In this complex, four effects are observed that misalign the α-phosphate and interfere with nucleotide incorporation. First, rather than forming a steric block, the Glu658 carboxylate interacts with the rCTP 2'-OH (2.7 Å apart) (Figure 4-2B). Consequently, the position of the entire rCTP is displaced from that of dCTP to such an extent that even the template dG base moves to retain pairing between the two bases, albeit with base-pair parameters that deviate substantially from those of canonical Watson-Crick base pairs (Table 4-4). Furthermore, the 2'-OH forms a hydrogen bond with Arg615, which normally forms a hydrogen bond with the sugar O4 atom (Figure 4-2B). This residue also participates in recognizing cognate base pairs by reading out their minor groove hydrogen-bonding pattern [60]. Second, the rCTP displacement opens an aperture between the bound nucleotide and the wall of the active site opposite the O helix, which is filled with a water layer that adopts a different structure than was observed in the ddCTP complex (Figure 4-4D). Third, the triphosphate binds in an altered conformation which is stabilized by this water layer (Figure 4-5, A and B). The absence of an anomalous scattering signal revealed that in this conformation, the Mg^{2+} (substituted with Mn^{2+} to obtain an anomalous signal) that normally assembles with the nucleotide is absent and replaced by two water molecules (Figure 4-2C). The direct hydrogen bond between the 3'-OH and its own β-phosphate oxygen observed in the dCTP complex is replaced by the Tyr 710 phenolic hydroxyl (Figure 4-2C). This cyclic, non-covalent interaction within the nucleotide triphosphate has also been observed in ternary complexes of B-, C-, X-, and Y-family DNA polymerases and reverse transcriptase family (Figure 4-6). Finally, the direct interaction between 3'-OH and E658 main-

chain amide is replaced by a water-mediated hydrogen bond (Figure 4-2C). The stabilizing interaction of the bound rCTP β-phosphate with Tyr710 on the O helix and the Glu658 backbone amide via a water molecule presumably together improve affinity sufficiently (Table 4-2) to select a single conformation out of the ensemble present in the wild-type complex.

The loss of the interaction between the glutamate carboxylate and 2'-OH of the rCTP complex in the E658A single mutant does not suffice to form the closed state (Figure 4-2E). The diminished selectivity between rCTP and dCTP (Table 4-2) is primarily a consequence of a 100-fold loss of activity for dCTP.

dCTP − To elucidate the structural mechanism for the loss of dCTP incorporation in the E658A mutant, we cocrystallized this mutant with dCTP and Mg^{2+}. The resulting complex (Figure 4-7) contains only one molecule in the asymmetric unit, compared with the two observed in all other ternary complexes. This complex does not bind a dCTP at its active site (dCTP is observed to bind at a remote surface site, confirming its presence in the crystal). The polymerase conformation is similar to the open state [40, 60], but the C-terminal segment of the O helix (residues 707-714) has bent towards the primer DNA 3' terminus (Figure 4-7A). A conserved water molecule which usually is present in both open structures [40, 60] (Figure 4-7C) and ternary complexes [62, 79, 101] is displaced by the backbone carbonyl of V713 (Figure 4-7B). It also is present in all structures except this one, reported here. This water molecule normally interacts with the minor groove of the unincorporated dNTP-template base pair. Its hydrogen-bonding interactions are likely to be sufficiently strong that they perturb the intrinsic tautomeric equilibria of the free nucleotides [101]. Loss of this water molecule can therefore be expected to have significant impact on dNTP incorporation rates.

Table 4-4 **Base pair parameters of ribonucleotide captured at the insertion site**

Base pair	λ_{primer}(◦)[a]	$\lambda_{template}$(◦)[a]	$d_{C1'-C1'}$(Å)	Shear(Å)	Stretch(Å)	Stagger(Å)	Buckle(◦)	Propeller(◦)	Opening(◦)
F710Y-rCTP•dG (Mg^{2+})	54.6	51.4	10.6	0.42	-0.21	0.28	4.72	-18.03	1.47
E658A-rCTP•dG (Mg^{2+})	59.4	54.2	10.5	0.34	-0.13	-0.30	7.47	-2.30	1.27
Watson-Crick[b]	57.8±1.8	54.7±0.9	10.6±0.1	0.06±0.28	-0.10±0.05	-0.09±0.08	4.31±4.51	-6.88±3.41	1.86±0.80

[a] λ_{primer} and $\lambda_{template}$ are defined as the angle between the glycosidic bond of primer or template nucleotide and the line draw between the C1′ atoms of the base pair (see inserted panel on the right). $d_{C1'-C1'}$ is the distance between the C1′ atoms of the base pair. All other base pair parameters are defined [118]. All values were calculated in 3DNA [119].
[b]Average values and standard deviations over all four cognate base pairs observed at the insertion site were adapted from [101].

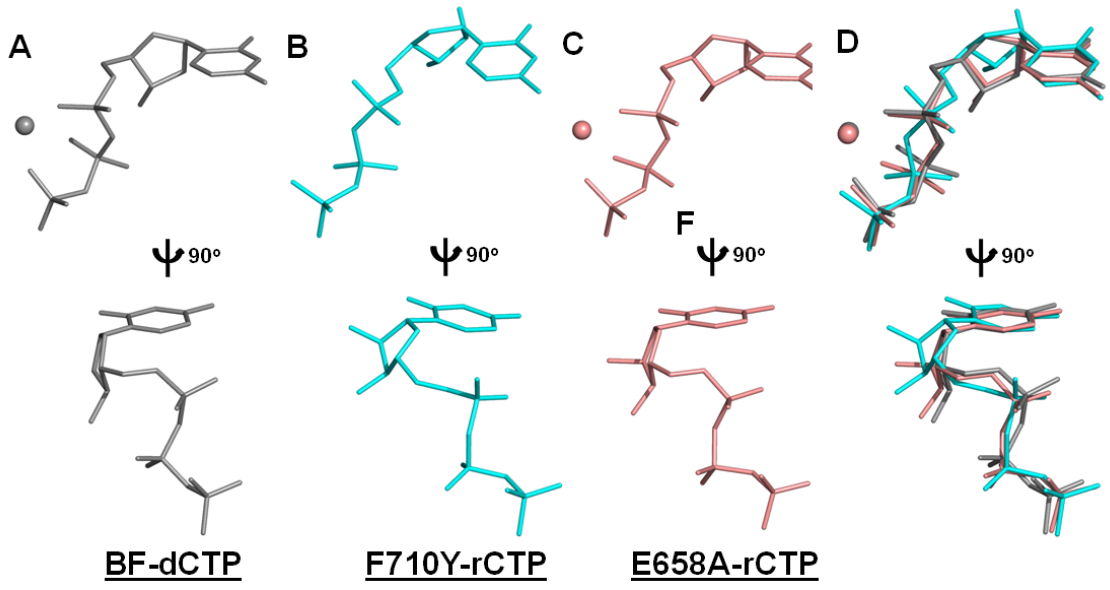


Figure 4-5 Nucleotide triphosphate conformations.
A-C, individual; D, superimposed. Gray: Wild-type active site-dCTP; blue: F710Y-rCTP; pink: E658A-rCTP. Two views are shown.

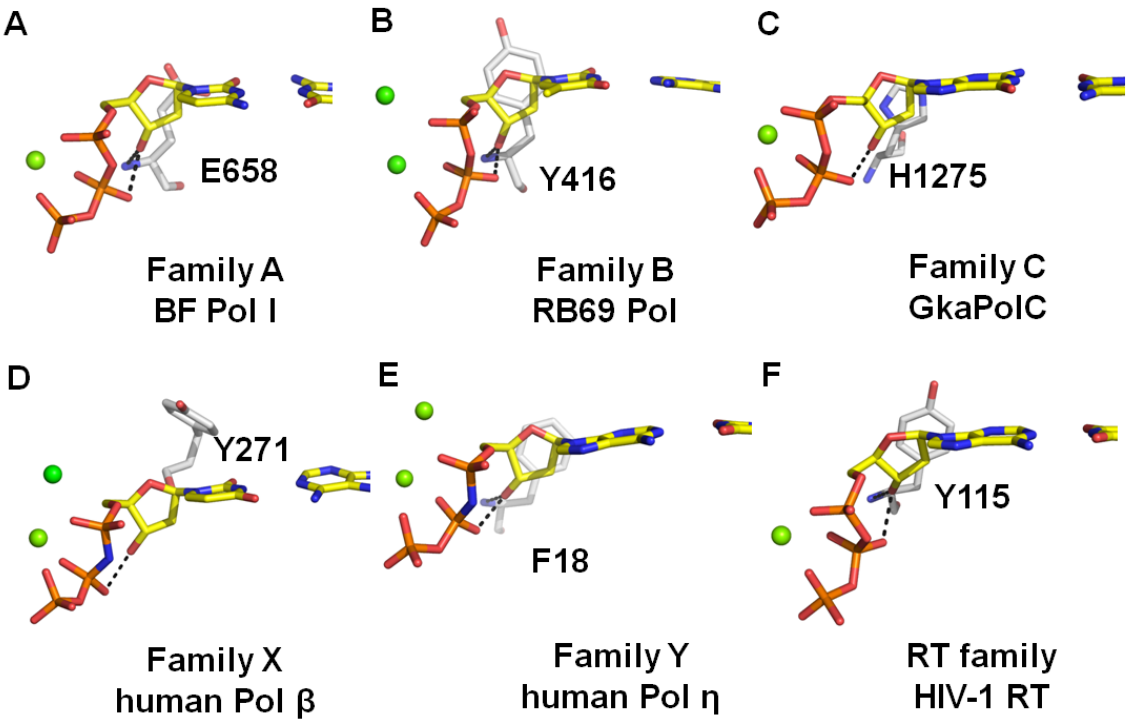


Figure 4-6 Conservation of the interaction between the dNTP 3'-OH and the β-phosphate at the insertion site of representative members of five DNA polymerase families and reverse transcriptase family.

A, A-family, Bacillus Fragment (BF) (this study). B, B-family, Enterobacteria phage RB69 DNA polymerase (pdb code: 3NCI) [29]. C, C-family, Geobacillus kaustophilus DNA polymerase PolC (3F2B) [31]. D, X-family, human DNA polymerase beta (2FMP) [32]. E, Y-family, human DNA polymerase eta (3MR2) [34]. F, Reverse Transcriptase (RT) family, HIV-1 RT (3KK2) [35]. Selection of the structures for polymerase families was based on availability and resolution of the ternary complexes.

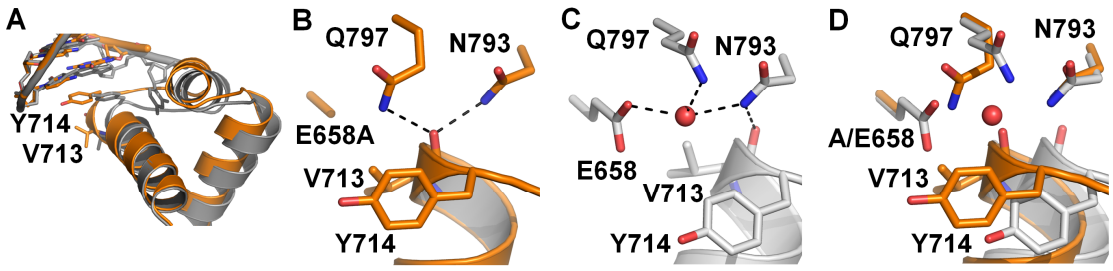


Figure 4-7 Comparison of E658A and wild-type BF-DNA binary complexes.
A, superposition of E658A (orange) and wild-type (gray) (pdb code: 1L5U [40]) complexes show that the mutation causes the C terminus of the O helix where Y714 resides to move toward the primer strand. B-C, Backbone carbonyl of V713 displaces the conserved water molecule coordinated by E658, Q797, and N793 in the wild-type structure. D, Q797 changes its side chain rotamer conformation and interacts with the V713 carbonyl together with N793.

The loss of this important water molecule in the E658A mutant is accompanied by several rearrangements in the active site, which are compensated for by the 2'-OH in the rCTP complex. Normally, Q797, N793 and E658 form a hydrogen-bonded network that coordinate this water. This network is broken in E658A such that Q797 adopts a different side chain rotamer conformation and the V713 carbonyl is displaced (Figure 4-7D). These motions together destroy the water coordination sphere. However in the rCTP complex, the 2'-OH is able to compensate for the loss of the carboxylate in the E658A, because in the properly formed closed state this hydroxyl occupies the position vacated by the carboxylate, restoring the hydrogen-binding network (Figure 4-4E). Despite the restoration of this critical water in the rCTP complex, the incorporation rate of this nucleotide remain low, because it still adopts distorted sugar and triphosphate conformations (Figure 4-2F; Figure 4-5, C and D). The direct interaction between the 3'-OH and β-phosphate is replaced by water; there is no interaction between the 3'-OH and the E658A backbone amide; the triphosphate binds Mg^{2+} (confirmed by Mn^{2+} anomalous signal), but not at a full occupancy (Figure 4-2G). The O helix remains open to a similar degree as observed for the F710Y-rCTP complex.

Base pairing of the incorporated rCTP − Covalently incorporated, cognate base pairs positioned at the post-insertion site form hydrogen bonds between their minor groove N3 (purines) or O2 atoms (pyrimidines) and Arg 615 and Gln 797 [60]. We incorporated rCTP at the primer strand 3'-terminus by catalysis in solution in the presence of Mn^{2+} [139] prior to crystallization. This binary complex between wild-type BF and DNA was crystallized in the absence of additional nucleotide. The resulting crystals contain one molecule in the asymmetric unit in the open form (Table 4-3). The rC is positioned in the post-insertion site, and retains the minor groove readout by R615 and Q797 observed for deoxynucleotide [60] (Figure 4-8), emphasizing the importance of selecting against rNTPs prior to their incorporation.

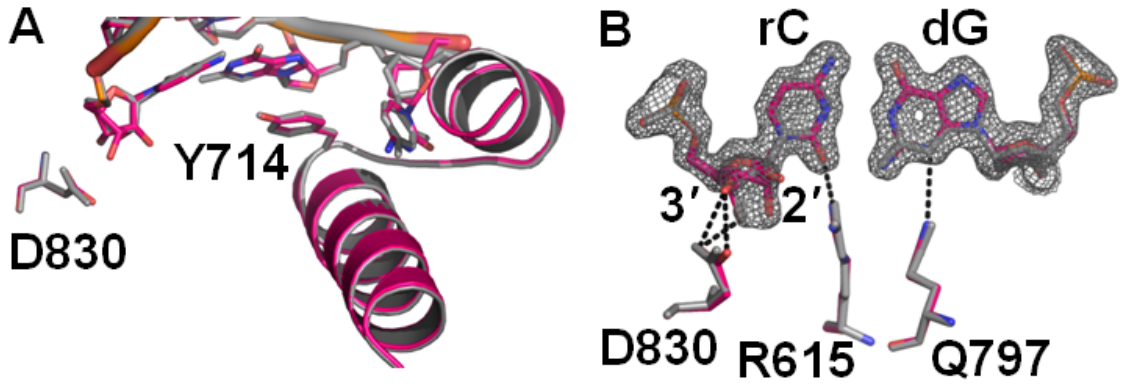


Figure 4-8 Comparion of rC•dG and dC•dG base pairs at the post-insertion site.
rC•dG (pink) and dC•dG (gray) comparison reveals similarities in their overall structure (A) and minor-groove hydrogen bonding (B). Composite-omit-map (gray mesh) is shown around the rC•dG base pair (contoured at 1.5σ level).

Discussion

The steric gate hypothesis postulates that discrimination between dNTPs and rNTPs is primarily the consequence of a steric blockage imposed by a single DNA polymerase residue on the rNTP 2'-OH [9]. Although reducing the bulk of this residue by mutation improves the incorporation rate of rNTPs relative to dNTPs, it invariably decreases dNTP incorporation (Figure 4-1). Furthermore, dNTP remains favored over rNTP incorporation in these mutants. These observations indicate that the "steric gate" residue functions not as a simple block, but also affects the mechanism of dNTP incorporation. The structural observations reported in this study suggest that the origins for this complex behavior lie within the dynamics of O helix motions [40, 79, 101]

We were able to capture rNTPs and ddNTPs at the "insertion site", paired with the template DNA, prior to reacting with the primer terminus and rNTP following its incorporation at the "post-insertion site". In the sugar moiety discrimination process, the O helix motions are interfered with, resulting in non-productive conformations that differ from those observed for mismatches or lesions [62, 79, 101]. Water-mediated interactions, which are absent in the closed state, contribute to the formation of the binding sites that trap the non-cognate nucleotides in the intermediate O helix conformations. These sites are placed in similar locations as observed for cognate nucleotide complexes, but differ in their interactions between the polymerase and bound nucleotide.

The O helix motions are therefore not simply a motion that distinguishes open and closed states, in which only the closed state binds NTPs. Instead the motions set up an ensemble of subtle "traps" that ensnare non-cognate NTPs. The binding of the water molecules that contribute to shaping the ensemble of non-cognate binding sites is affected by polymerase mutants, thereby altering the thermodynamic landscape of the non-cognate nucleotide traps. This mechanism

accounts for the large variation in specificities that has been noted [111, 140], and is consistent with a model based on the analysis of enzyme kinetic data in which specificity is a consequence of subtly different reaction pathways corresponding to the states particular to individual (mis)matched base pairs [73].

The ddCTP and rCTP complexes presented here illustrate these effects. The two water layers observed in the rCTP and ddCTP complexes (Figure 4-4) differ from each other, stabilizing aberrant interactions (or lack of interaction) specific to the respective nucleotide. In the case of the rCTP complex in the F710Y enzyme the interactions with water are sufficiently strong that they prevent binding of Mg^{2+} to the triphosphate. The remodeling of water structure in response to mutations is illustrated in the F710Y complex with ddCTP, in which the mutant Tyr710 phenolic hydroxyl replaces a bound water such that the closed state binding becomes dominant, eliminating the selectivity against ddNTP incorporation.

Water exclusion in the active site has long been proposed to enhance the effect of hydrogen-bonding interactions through lowering of the local dielectric thereby accentuating the energetic differences between correctly and incorrectly formed bases [141]. Here we show that the effects of water extends well beyond that of generalized electrostatic through dipole relaxation and involves specific, but readily remodeled interactions that contribute to creating the ensemble of binding sites necessary for kinetic checkpoints.

Although the recognition mechanisms by which sugar discrimination interferes with O helix motion and proper closed state assembly are distinct from base pair recognition, some polymerase residues play a role in both processes. This duality is responsible for the loss of dNTP incorporation in the E658A steric gate mutant, and also involves water-mediated effects. E658 forms part of a hydrogen-bonding network that holds in place a water molecule critical for reading out minor groove interactions between the NTP and the DNA template. In the E658A

mutation these interactions are destroyed, leading to the loss of the water and reduced dNTP incorporation. This loss is partially compensated for in the rNTP complex of this mutant, in which the 2'-OH takes the place of the missing carboxylate.

The subtle energetics generated by dynamically interconverting ensembles of conformational states is key to understanding protein folding [142], enzyme kinetics [73, 143, 144], and allostery [145, 146]. The mechanistic principles established by this recently emerged view of protein structure-function relationships are central to understanding DNA polymerase fidelity. Here we show structural features that lead to the creation of such thermodynamic landscapes via the interplay between protein dynamics, ligand binding, and solvent interactions.

Methods

Protein preparation − D598A, F710Y/D598A, E658A/D598A, F710Y, E658A mutants were constructed using QuickChange Site-directed Mutagenesis (Stratagene, La Jolla, CA). Wild-type and mutant proteins were expressed and purified as described [28]. Proteins containing the D598A mutation were used for crystallization to obtain closed form ternary complexes 1-9 (Table 4-3) of BF polymerase, DNA duplex, and incoming nucleotide substrate (paired with complementary template base) positioned at the insertion site prior to chemistry [40]. Wild-type protein was used to obtain the open form, binary complex 10 (Table 4-3) with ribonucleotide incorporated into the primer 3' terminus paired with template at the post-insertion site. Pre-steady state kinetics was carried out using proteins without the D598A mutation.

Crystallization of nucleotide substrates and DNA complexes − Oligonucleotides were synthesized by Midland Certified Reagent Co. (Midland, TX) at GF grade. Ternary complexes were crystallized using 5′-CCTGACTC$_{dd}$C-3′ as primer (dideoxynucleotide terminus traps the complex before chemistry) and 5′-CATGGGAGTCAGG-3′ as template. The primer and template

strands were annealed as described [40]. Ultrapure rCTP and ddCTP were purchased from USB Co. (Cleveland, OH) and dCTP was purchased from Promega Co. (Madison, WI). Protein, primer-template duplex (protein:DNA, a 1:3 molar ratio), nucleotides (dCTP or ddCTP, 10mM; rCTP, 30 mM; Mg^{2+} or Mn^{2+}, SO_4^{2-} salts; nucleotide:metal, 1:2 ratio) were incubated (1 hr, room temperature), and crystallized as described previously [40]. rCTP was incorporated (protein:DNA, 1:3 molar ratio; 30 mM rCTP; 60 mM $MnSO_4$;1 hr, room temperature) and crystallized in a binary complex using a 5′-GCGATCACGTA-3′ (primer) and 5′- GACGTACGTGATCGCA-3′ (template) DNA substrate as described [40].

Diffraction data collection and processing − All crystals were flash frozen in liquid nitrogen directly out of the crystallization drop. Data were collected at SIBYLS and SER-CAT beamlines. All crystals belong to the space group $P2_12_12_1$. The closed form crystals (complexes 1-8) contain two molecules in the asymmetric unit while the closed form complex 9 and the open form complex 10 contain one. Complex 8 was processed using HKL2000 [105] and all other were processed using XDS [104]. Complexes 1-6 exhibit ice ring formation, accordingly the reflections in shells 2.28-2.22 Å and 1.948-1.888 Å were excluded.

Structure determination – **Ternary complexes 1-8:** A closed form ternary complex (PDB code: 2HVI [62]) from which the active site residues on the O helix and preceding loop region (residues 681-727), incoming nucleotide substrate, metal ions, surface-bound sulfate (from the crystallization solution), and water molecules had been removed was used as the starting model for solving the structures of complexes 1-2, 4-5, and 7. Complexes 3, 6, and 8 are the Mg^{2+} or Mn^{2+} counterpart for complexes 4-5 and 7. Preliminary refinement showed that the two complexes with different metal ions are almost identical. Therefore, fully built and refined structures of complexes 4-5 and 7 with water molecules removed were used as the starting models for their metal-substituted counterparts.

Initial electron density maps for complexes 1-2, 4-5, and 7 were generated by Fourier synthesis after rigid body refinement using the modified starting model 2HVI in PHENIX [108]. Missing residues, nucleotide substrates, metal ions, and surface small molecules for all complexes were manually built in Coot [109] into the electron density map generated by iterative rounds of model building and refinement with PHENIX [108]. All residues in the complexes were evaluated by visual inspection to fit the map. Simulated annealing refinement was then applied to further improve the model. TLS groups were determined using PHENIX find_tls_groups [108] and TLS parameters were refined subsequently. Water molecules were added automatically in the refinement followed by visual inspection.

For the closed form ternary complexes 1-8, the two molecules in the asymmetric are not the same: molecule **1** (chains D, E, and F in the pdb) is more ordered than molecule **2** (chains A, B, and C) (chain naming follows previously published structures [62]). Each molecule in complexes 1-2, 7-8 adopts a single unique conformation (closed, ajar, or open). In complexes 3 and 4, molecule 1 does not adopt a unique conformation (molecule 2 is open and empty), and was modeled initially as a mixture of two distinct ajar O helix conformations. After refinement, additional densities around the O helix could be observed in difference maps, suggesting that there are additional conformations that have not been accounted for in the model. At least two distinct rCTP conformations were observed. In the final, deposited models for complexes 3-4, the conformational diverse portions of the O helix, its preceding loop region, and rCTP were omitted. For complexes 5-6, molecule 1 adopted a single ajar conformation with bound rCTP, whereas molecule 2 adopted an ensemble of conformations similar to complexes 3-4; in the deposited structures for complexes 5-6, these also were omitted from the model.

Closed form binary complex 9: The E658A mutant, DNA, dCTP, and Mg^{2+} complex was crystallized as described above for the closed form ternary complexes 1-8. However, the unit

cell dimensions of this complex resemble those observed for an open, binary crystal form [40, 60], and contains only one molecule in the asymmetric unit. The structure was determined by molecular replacement with Phaser [147], using as the search model the open, empty molecule of complex 7 which carries the same mutation and contains a DNA duplex with the same sequence . The output model was refined with simulated annealing in PHENIX [108] followed by iterative rounds of visual inspection, manual building, and refinement as described above. dCTP was not observed in the active site of this complex, but was observed to bind at a remote surface site [62, 79, 101], confirming its presence in the crystal.

Open form binary complex 10: an open form binary complex (PDB code: 1L5U) with a base pair positioned at the post-insertion site, was used as the starting model for solving the structure complex 10. The model for the rC nucleotide incorporated at the primer 3' terminus was built into the electron density map. The structure was refined as described above. There is one molecule in the asymmetric unit with a single, open conformation.

Structure quality checks − For all structure determinations, free reflections were generated in XDSCONV [104] or CCP4 [106] by combining inherited free reflections from the starting model and 5% randomly selected reflections beyond the resolution of the starting model for each complex. All structures were refined to good protein geometry with no Ramachandran outliers and ~98% residues in the Ramachandran favored region according to MolProbity [148]. Data, refinement statistics, Ramachandran plot summary, and Protein Data Bank accession codes are presented in Table 4-3. Composite omit maps were generated in CNS [110]. Figures and superpositions (all protein residues used) were prepared in PyMOL (Schrödinger, LLC.).

Pre-steady state kinetics of wild-type, E658A, and F710Y BF − Complementary oligonucleotides used in the solution studies were synthesized by Operon Biotechnologies, Inc. (Huntsville, AL). The template (5'-

TTACTTGACCAGATACACTGTCTTTGACACGTTGATGGATTAGAGCAATCACATCCA AGACTGGCTATGCACGAA-3') and fluorescently labeled primer (5'-[6-FAM] TCGTGCATAGCCAGTCTTGGATGTGATTGCTCTAATCCATCAACGTGTCAAAGACAG TGTATCTGGT-3') strands were annealed as described [40]. The DNA substrate and wild type, E658A, or F710Y proteins were diluted with reaction buffer (50 mM Tris-HCl, pH 8.0; 50 mM NaCl; 10 mM $MgCl_2$; 1 mM DTT) to 0.1 and 0.5 µM, respectively. BF-DNA complexes were mixed with equal volumes at various concentrations of dCTP, rCTP, or ddCTP. The reaction (room temperature) was quenched at different time points using four reaction volumes of quench solution (95% formamide (v/v); 25 mM EDTA). Reactions slower than 10 seconds (wt-rCTP, wt-ddCTP, E658-dCTP, E658-rCTP, and F710Y-rCTP) were executed manually. Reactions faster than 10 seconds (wt-dCTP, F710Y-dCTP, and F710Y-ddCTP) were executed using a KinTek RQF-3 Rapid Quench Flow instrument (KinTek Corporation, Austin, TX). Primer extension was quantified by capillary electrophoresis with fluorescence detection, using an ABI3100 Genetic Analyzer (Applied Biosystems, Foster City, CA). Pre-steady-state kinetic constants were determined and fit as described [79].

Chapter 5 Structural basis for ensembles of incorporation pathways of a high-fidelity DNA polymerase

Summary

In the previous two chapters, structural mechanisms for both mutagenic and accurate DNA replication have been elucidated. In this chapter, more examples of base-pair mismatches, G•G (primer•template), T•G, and T•T, bound to the BF polymerase active site at non-productive conformations and misaligned for chemistry were captured. A novel nucleotide binding mode was revealed for the G•G mismatch illustrating the dynamic nature of the nucleotide selection process. By comparing these mismatch structures together with previously determined complex structures of BF DNA polymerase, an ensemble of intermediate conformations in between the open and closed states emerged. These intermediate conformations together with active site water structure remodeling provide binding sites to trap and misalign incorrect substrates for incorporation. Only those with cognate base-pair shape and size can promote the fully closed state and be properly aligned for chemistry.

Introduction

DNA polymerase selects deoxyribonucleotide substrate with cognate base pair for incorporation while it discriminates highly against nucleotides with incorrect base [5, 69] and ribose sugar [9, 10]. The specificity of DNA polymerase is achieved at five fidelity filter sites [37, 67, 70, 71, 101]: the pre-insertion site; the insertion site; the catalytic site; the post-insertion site; and a four-base-pair DNA duplex-binding region. Shape complementarity between the polymerase active site and cognate base pairing of the nascent base pair [69], and precise

alignment of the functional groups at the insertion site are the most critical for replication fidelity [23, 111, 112, 116]. The pairing and aligning process involves a large scale motion of the polymerase fingers domain from an “open” to a “closed” state [2, 25] which has been observed in A, B, C, X, and RT families of DNA polymerases [31, 37-43].

Extensive solution studies to understand the specificity of DNA polymerases have been conducted leading to the identification of “kinetic checkpoints” that can serve as intermediate steps to screen the incoming nucleotide substrate and to reject incorrect ones [73]. Base-pair mismatches or incorrect sugar substrates are likely to be trapped at intermediate conformations between the open and closed states [73]. However, due to the lack of structures of incorrect nucleotide substrates bound at the polymerase active site prior to chemistry, structural evidence to support the ensemble of incorporation pathway to ensure replication fidelity was limited.

Recently, structural studies on a high-fidelity *Bacillus* DNA polymerase I large fragment (*Bacillus* fragment, BF) on deoxyribonucleotide selectivity over dideoxy- and ribo-nucleotide prior to chemistry revealed several intermediate conformations that trap incorrect substrates along the trajectory for fingers domain closure and the substrates are misaligned at the active site for incorporation [27]. Here we have captured three high-resolution structures of base-pair mismatches G•G (primer•template), T•G, and T•T bound at the BF DNA polymerase active site prior to chemistry. Comparisons of the structures of mismatches with those of cognate base pairs show that mismatches with non-cognate base pairing schemes are trapped at intermediate polymerase conformations and their functional groups are misaligned for chemistry. Among them, an extreme is the G•G mismatch which adopts a novel nucleotide binding mode that is drastically different from the previously observed ones illustrating the dynamic nature of nucleotide binding and trapping process.

These structures were further analyzed in comparison with previously determined high-resolution structures of BF polymerase (more than 60) with cognate base pairs [27, 40, 60, 62, 101], mismatches [61, 79, 101], lesions [62-65], and incorrect sugar substrates [27] bound at different fidelity filter sites. A unified picture of nucleotide selection by DNA polymerase has emerged: incorrect substrates with non-cognate base-pairing schemes or wrong sugar moiety trap the polymerase at an ensemble of non-productive intermediate conformations and are misaligned for chemistry; only those exhibiting complementary cognate base-pairing shape and size can promote fully closed polymerase conformation and be aligned for incorporation. Collectively, these observations provide the structural basis for the ensemble of the incorporation pathway of DNA polymerase to ensure replication fidelity.

Results

Ternary complex crystals – High-resolution crystal structures of ternary complexes of BF, DNA, and nucleotide were obtained capturing three base-pair mismatches, ddGTP•dG, ddTTP•dG, and ddTTP•dT, and one incorrect sugar substrate, rCTP•dG (Table 5-1). A dideoxynucleotide chain terminator was placed at the primer strand 3' terminus to prevent catalysis allowing us to trap the complexes prior to chemistry. A F710Y mutation was introduced to facilitate the incorporation of primer terminal dideoxy chain terminator [94]. A surface D598A mutation was included to destabilize a crystal contact that favors the open state in the crystal lattice [40]. The resulting crystal form contains two molecules in the asymmetric unit. In all ternary complexes, molecule **1** is more ordered than molecule **2.** In addition, Molecule **1** shows well ordered base pairing of the nucleotide and template base while molecule **2** is either open and empty or similar to molecule **1**. Therefore, our analyses were based on the molecule **1**.

Table 5-1 Crystallographic data collection and refinement statistics

	ddGTP•dG (*n*)	**rCTP•dG Mg^{2+} (*n*)**	**rCTP•dG Mn^{2+} (*n*)**	**ddTTP•dT (*n*)**	**ddTTP•dG (*n*)**
Data collection					
Wavelength	0.9712	1.0000	1.0000	0.9795	1.1158
Resolution (Å)	50-1.64	50-1.63	50-1.65	50-1.64	100-1.57
Outer shell (Å)	1.74-1.64	1.73-1.63	1.75-1.65	1.74-1.64	1.66-1.57
R_{sym} (%)	5.5(45.7)[a]	7.2(50.8)	10.9(50.9)	4.2(49.6)	4.8(46.2)
$I / \sigma I$	16.6(3.6)	20.3(3.4)	15.0(3.7)	22.9(3.9)	26.3(3.8)
Completeness	88.1(98.3)	100.0(99.8)	98.2(89.1)	91.2(99.9)	92.9(99.9)
Redundancy	5.1(5.0)	10.0(5.2)	10.0(5.6)	5.6(5.6)	8.6(5.3)
Refinement					
Resolution (Å)	46.9-1.64	35.0-1.63	34.2-1.65	43.1-1.64	79.6-1.57
No. reflections	158215	184316	175747	161680	191797
R_{work} / R_{free}	17.9/20.3	17.8/19.8	18.2/20.5	17.8/19.7	18.7/20.9
No. non-hydrogen atoms					
Total	11486	11646	11709	11657	11661
Solvent	1482	1340	1410	1449	1526
B-factor	27.1	30.8	30.8	30.6	30.1
R.m.s. deviations					
Bond lengths	0.006	0.006	0.006	0.006	0.006
Bond angles	1.076	1.110	1.112	1.089	1.103

[a] Numbers in the parenthesis correspond to the outer shell.

Novel "up-side-down" nucleotide binding mode − In both the G•G mismatch and rCTP structures, the incoming nucleotides are paired with the template dG at the polymerase insertion site. But different from mismatches [79, 101], lesions [62], or incorrect sugar substrates [27] captured previously at the same site, the base and sugar moieties of the ddGTP and rCTP have flipped "up-side-down" and the glycosidic bonds of incoming nucleotide and the template dG is in a trans orientation instead of the commonly observed cis orientation [149] (Figure 5-1). Both structures adopt intermediate O helix conformations (Figure 5-1A, D) and α-phosphates of ddGTP and rCTP are misaligned for chemical incorporation (Figure 5-1B, E). The incorrect nucleotides have to disassociate before they can be aligned properly for chemistry.

ddGTP•dG complex – Four structural features of the ddGTP•dG complex define the non-productive state of BF polymerase. First, ddGTP is misaligned at the active site by forming extensive interactions with the template base and the polymerase via direct and water-mediated hydrogen bonds (Figure 5-1C; Figure 5-2A). ddGTP adopts its syn conformer which is the more energetically favorable than the anti conformer preferred by other nucleotides [150]. In addition, it has flipped up-side-down and the Watson-Crick edge of the guanine base is facing the polymerase surface resulting in hydrogen-bond interactions with the E658 side chain and the water molecules at the active site. Base-pair parameters of ddGTP•dG are dramatically different from those for cognate base pairs captured under the same experimental condition (Table 5-2). Second, the triphosphate moiety of the ddGTP adopts an extended and distorted conformation and it fails to bind a metal ion (confirmed by the lack of anomalous signal from Mn^{2+} present in the crystallization solution) (Figure 5-3). Third, the O helix associated with ddGTP triphosphate adopts an intermediate conformation. Finally, ddGTP fails to displace a layer of water molecules (Figure 5-1C) that are absent in the catalytically potent fully closed form (Figure 5-1G) but is present in all non-productive complexes captured previously [27, 79, 101].

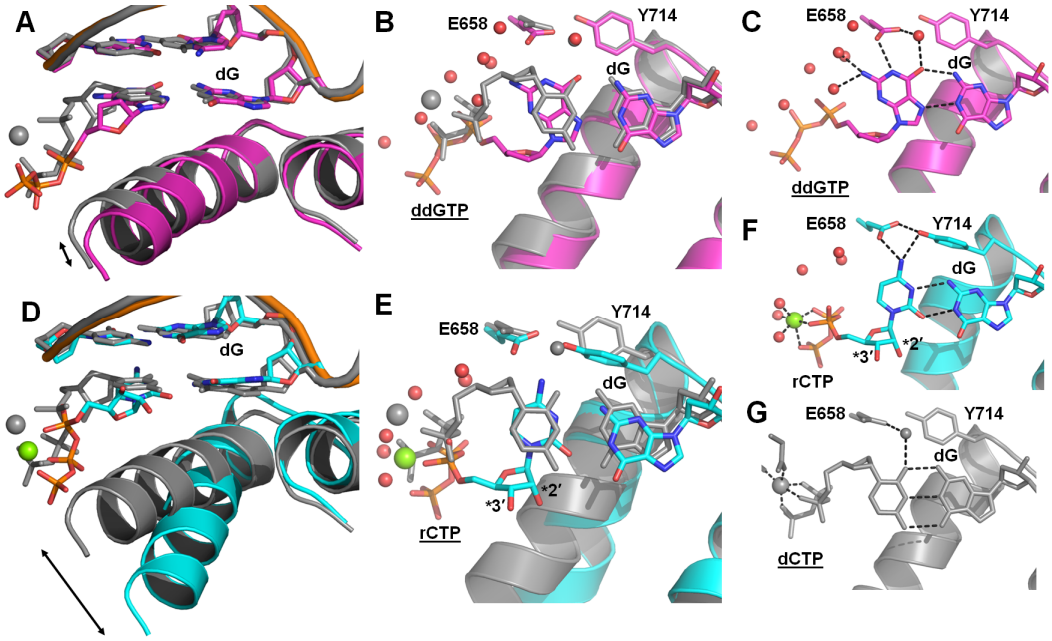


Figure 5-1 Comparison of ddGTP•dG, rCTP•dG, and dCTP•dG complexes.
A and B, two views comparing the BF complex of ddGTP (magenta) with that of dCTP (gray). ddGTP binds up-side-down and forms syn-anti base pair with the template dG. The O helix does not fully close and the nucleotide triphosphate is misaligned (arrows emphasize structural differences). The base pair parameters of this mismatch are dramatically distorted from their canonical values (Table 5-2). ddGTP fails to displace the water layer at the active site (shown as red spheres). C, ddGTP makes extensive hydrogen-bond interactions with the template dG and the polymerase through direct and water-mediated interactions. ddGTP triphosphate is distorted such that a metal ion fails to bind. D and E, two views comparing the rCTP complex of BF (cyan) with the dCTP complex (gray). Arrow indicates the failure of the O helix to close. rCTP binds up-side-down at the active site and forms a wobble pair with the template dG which is displaced toward the major groove. There is a water layer in between of the rCTP and the polymerase. The triphosphate conformation is different from that of dCTP, but it still binds a metal ion. The C-terminal turn of the O helix is distorted and Y714 side chain hydroxyl displaces the conserved water molecule observed in the wild-type dCTP complex. F, interactions are made between the rCTP base and E658 and Y714 side chains. G, dCTP forms a cognate base pair with the template dG and interacts with the conserved water molecule. The water layer is displaced and dCTP is aligned for chemistry. Mg^{2+} ions (green or gray) are shown as spheres.

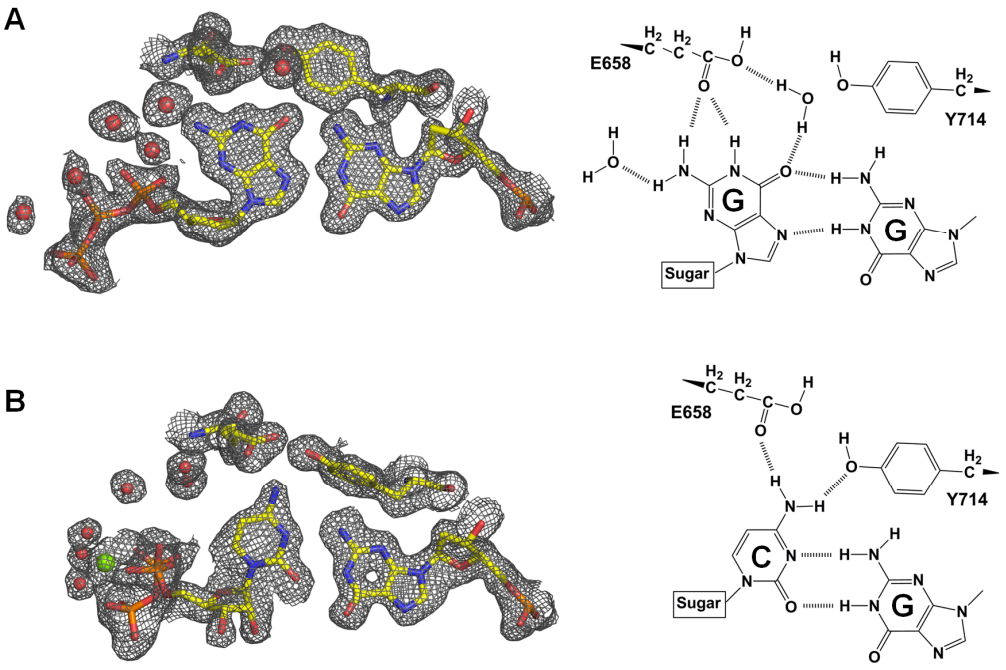


Figure 5-2 Electron density map and detailed interactions for G•G and rCTP•dG pairs. Left panel, composite-omit-maps (gray mesh) contoured at 1σ (ddGTP•dG) (A) and (rCTP•dG) (B). Right panel, interactions of ddGTP (A) or rCTP (B) at the active site.

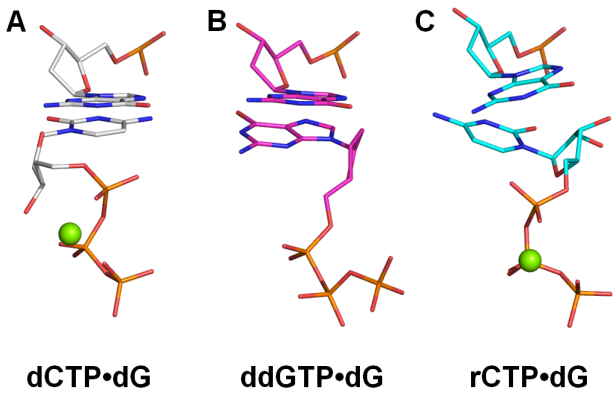


Figure 5-3 Nucleotide triphosphate conformations.
A, Wild-type active site-dCTP•dG; B, F710Y-ddGTP•dG; C, F710Y-rCTP•dG.

Table 5-2 Base pair parameters at the insertion site

Base pair	λ_{primer}(◦)[a]	$\lambda_{template}$(◦)[a]	$d_{C1'-C1'}$(Å)	Shear(Å)	Stretch(Å)	Stagger(Å)	Buckle(◦)	Propeller(◦)	Opening(◦)
ddGTP•dG	31.4	29.8	11.1	-5.48	-0.33	0.33	12.46	8.98	-112.60
rCTP•dG	63.0	38.8	10.5	0.09	-3.74	0.30	-28.07	13.92	-152.40
ddTTP•dG	61.7	39.1	10.8	2.43	-0.47	0.31	5.13	-8.39	-4.51
ddTTP•dT	72.9	41.4	8.8	2.29	-1.89	0.14	-6.24	-9.21	7.73
Watson-Crick[b]	57.8±1.8	54.7±0.9	10.6±0.1	0.06±0.28	-0.10±0.05	-0.09±0.08	4.31±4.51	-6.88±3.41	1.86±0.80

[a] λ_{primer} and $\lambda_{template}$ are defined as the angle between the glycosidic bond of primer or template nucleotide and the line draw between the C1′ atoms of the base pair (see inserted panel on the right). $d_{C1'-C1'}$ is the distance between the C1′ atoms of the base pair. All other base pair parameters are defined [118]. All values were calculated in 3DNA [119].

[b] Average values and standard deviations over all four cognate base pairs observed at the insertion site were adapted from [101].

rCTP•dG complex – Previously, an rCTP•dG structure was captured at the BF polymerase insertion site which forms three inter-base hydrogen bonds [27] (Figure 5-4A). Under slightly different experimental condition, however, we have captured an rCTP•dG complex where the incoming rCTP is bound up-side-down (Figure 5-1E). With this binding mode, rCTP and dG can no longer form a cognate base-pair shape due to steric clash between the hydrogen atoms of rCTP N4 and dG N2 (Figure 5-4B). Instead, rCTP•dG formed a wobble pair with two inter-base hydrogen bonds (Figure 5-1F) which is the most stable base-paring scheme when rCTP is bound up-side-down (Figure 5-4B). rCTP binding is further stabilized by interactions with E658 and Y714 side chains. The interaction with Y714 is made possible by the distortion of the C-terminus of the O helix where Y714 resides (Figure 5-1D). The phenolic hydroxyl group of Y714 replaces a conserved water molecule that reads-out the minor groove of the incoming nucleotide [27, 101] (Figure 5-1C, F, G). rCTP binds a metal ion (Mg^{2+}), the existence of which was confirmed by anomalous signal when Mg^{2+} is substituted by Mn^{2+}, but it is still misaligned for chemistry. The O helix adopts an intermediate conformation and the complex is trapped at a non-productive state.

More non-productive base-pair mismatch complexes – In addition, we have captured two mismatches, ddTTP•dG and ddTTP•dG, prior to chemical incorporation, which further illustrates how non-cognate base pairs trap the DNA polymerase at non-productive conformations (Figure 5-5). In both complexes, the mismatches form non-cognate base-pair shapes, BF polymerase adopts intermediate O helix conformations, and the functional groups are misaligned for chemistry.

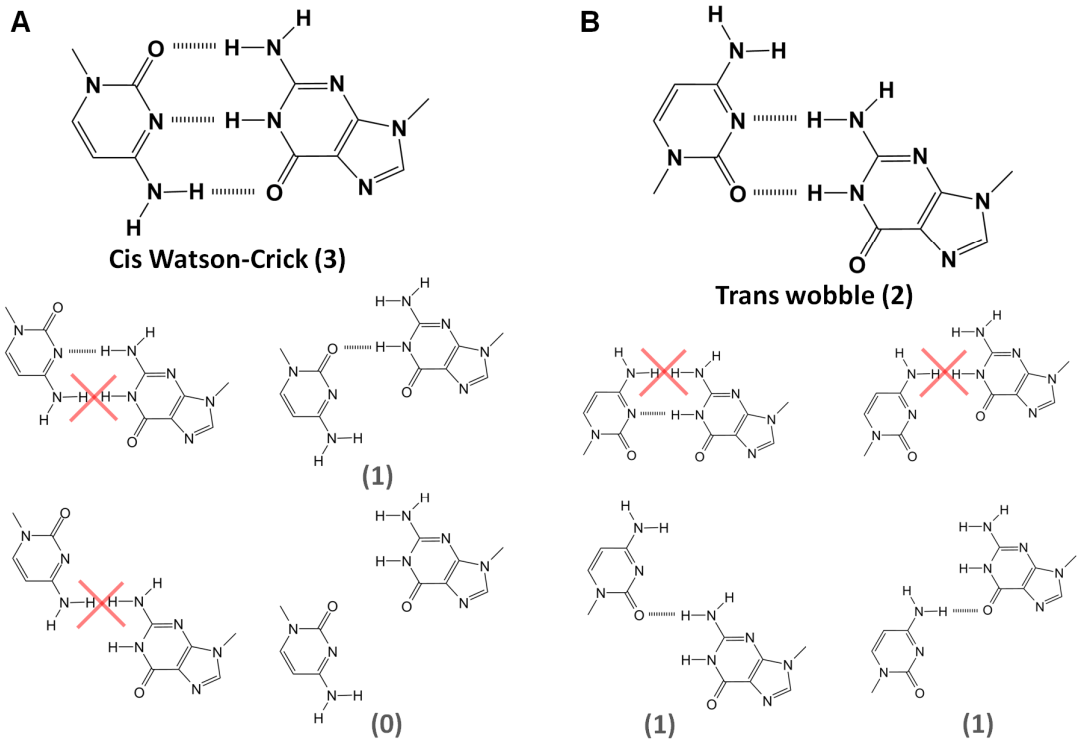


Figure 5-4 Base pairing schemes of cytosine and guanine.
A, C•G at cis glycosidic bond orientation. Canonical Watson-Crick C•G pair with three inter-base hydrogen bonds is shown (top panel). B, C•G at trans glycosidic bond orientation. Wobble C•G pair with two inter-base hydrogen bonds is shown (top panel). All other base pairing schemes are either less stable or impossible due to steric clash (marked with a red cross). Numbers in the parenthesis indicate how many inter-base hydrogen bonds are formed.

ddTTP•dG complex – T•G forms a wobble base pair with the incoming ddTTP moving toward the major groove while the template G is placed at the same spot as it is in a cognate C•G structure (Figure 5-5B). Displacement of ddTTP into the major groove creates a large aperture between the ddTTP base ring and the polymerase surface formed by E658 and Y714. In addition to the conserved minor-groove read-out water molecule, one more water molecule is bound between the anchored water and the ddTTP base. This wobble T•G mismatch forms stable interactions at the BF polymerase active site through two inter-base hydrogen bonds and two water-mediated interactions. ddTTP is misaligned and fails to displace the water molecules at the active site. The triphosphate is distorted and only the β-phosphate is directly anchored to a metal ion, the existence of which was confirmed by the presence of Mn^{2+} anomalous signal. BF polymerase adopts an intermediate O helix conformation.

ddTTP•dT complex – Similar to the T•G mismatch structure, ddTTP of the T•T mismatch is also displaced toward the base-pair major groove and there is a second water molecule bound between the anchored water and the ddTTP base ring (Figure 5-5E). T•T adopts a wobble base pair by forming two inter-base hydrogen bonds and two water-mediated interactions (Figure 5-5F). The incoming nucleotide is misaligned for chemistry. There is no metal bound to the distorted triphosphate and there is a layer of water molecules at the active site, the structure of which was remodeled compared with that for the ddTTP•dG complex (Figure 5-5 B, E). The BF polymerase adopts an intermediate conformation.

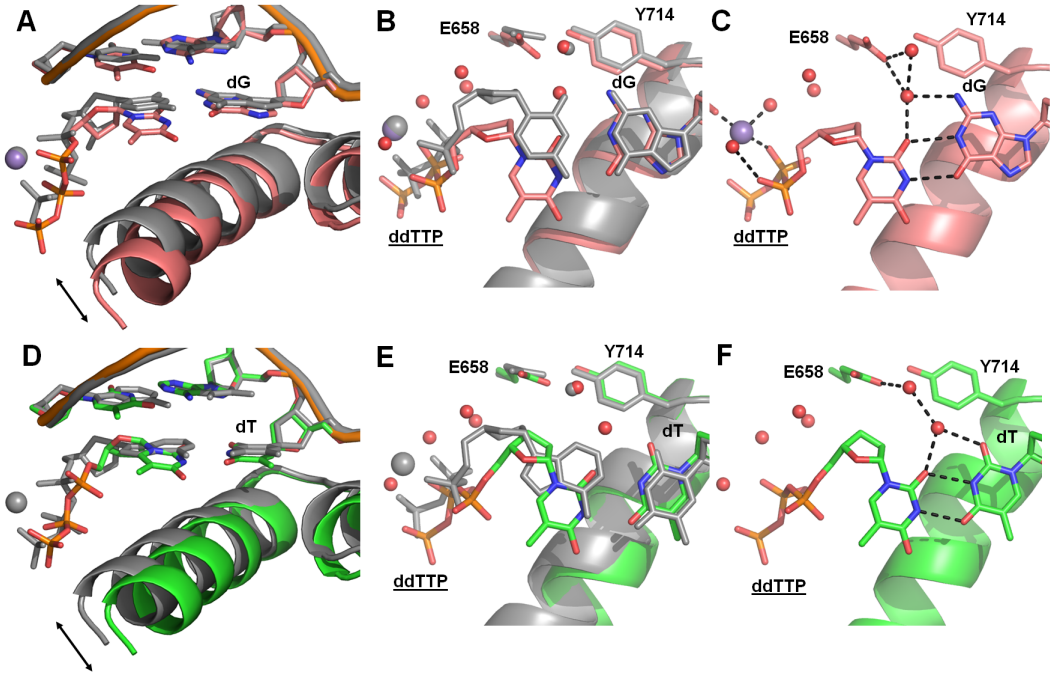


Figure 5-5 Comparison of ddTTP•dG and ddTTP•dT complexes with cognate base pair complexes.

A and B, two views comparing the ddTTP•dG complex (pink) with the dCTP•dG complex (gray). In the ddTTP complex, T•G forms a wobble base pair, the O helix does not fully close, and the nucleotide triphosphate is misaligned for chemistry (arrows emphasize structural differences). ddTTP fails to displace the water layer at the active site (shown as red spheres). C, ddTTP forms extensive interactions at the active site. D and E, two views comparing the ddTTP•dT complex (green) with the ddATP•dT complex (gray). Arrow indicates the failure of the O helix to close. ddTTP forms a wobble pair with the template dT and is displaced toward the base pair major groove. There is a water layer in between of the ddTTP and the polymerase. The triphosphate conformation is distorted and fails to bind a metal ion (lack of anomalous signal from the Mn^{2+} present in the crystallization solution). F, ddTTP forms extensive interactions at the active site. The base pair parameters of both mismatches are dramatically distorted from those of the cognate base pairs captured at the same position (Table 5-2). Mg^{2+} ions (green or gray) are shown as spheres.

Discussion

Novel nucleotide binding mode − Structures of up-side-down ddGTP and rCTP paired with template dG captured prior to chemistry reveal a new nucleotide binding mode. Instead of the cis glycosidic bond orientation observed previously for nucleotides bound at the BF polymerase insertion site, the incoming nucleotide can assume a trans orientation with regard to the template base and still form stable interactions at the active site which allows the structures to be captured crystallographically. These aberrant binding modes trap DNA polymerase at non-productive conformations and the nucleotide is misaligned for incorporation.

Non-productive states are inter-convertible – Structures of rCTP•dG [27] and ddTTP•dG [79] complexes have been captured previously under slightly different experimental conditions. BF polymerase shifts between two non-productive states in the presence of the same nucleotide substrate and DNA duplex: a more open intermediate when crystals were soaked in cryo-protecting solution with sucrose and a more closed intermediate without soaking. In addition to the two different non-productive enzyme conformations, the water structures are also remodeled under different experimental conditions (Figure 5-6). These observations indicate that the energetic differences between the non-productive states are likely to be small and are consistent with the hypothesis of conformational interconversions on the polymerase reaction pathway [73].

Ensemble of binding sites − By comparing the observations from this study with the structures determined previously of BF polymerase with mismatches [79, 101] and incorrect sugar substrates [27], a consistent picture emerges depicting that DNA polymerase creates an ensemble of non-productive binding sites consisted of intermediate O helix conformations between the open and closed states and active site water molecules that are failed to be displaced. Both the polymerase conformation and the water structure can be remodeled by DNA polymerase active site residues and the nature of the incoming nucleotide substrate (Figure 5-7).

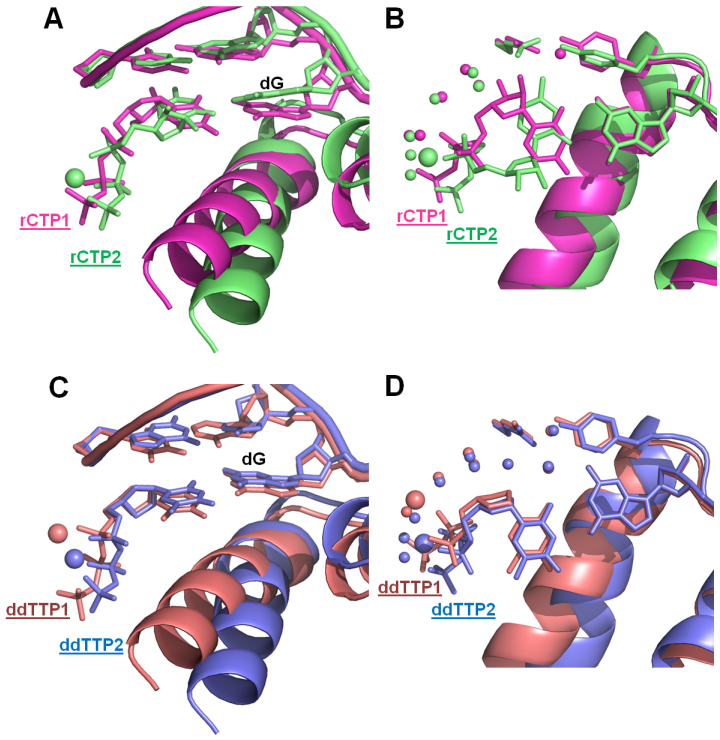


Figure 5-6 Comparisons of rCTP•dG and ddTTP•dG complexes captured under different experimental conditions.
A,B Two views comparing rCTP•dG captured in the presence (salmon) and absence (blue) of cryo-protecting solution. C,D Two views comparing ddTTP•dG captured in the presence (green) and absence (pink) of cryo-protecting solution.

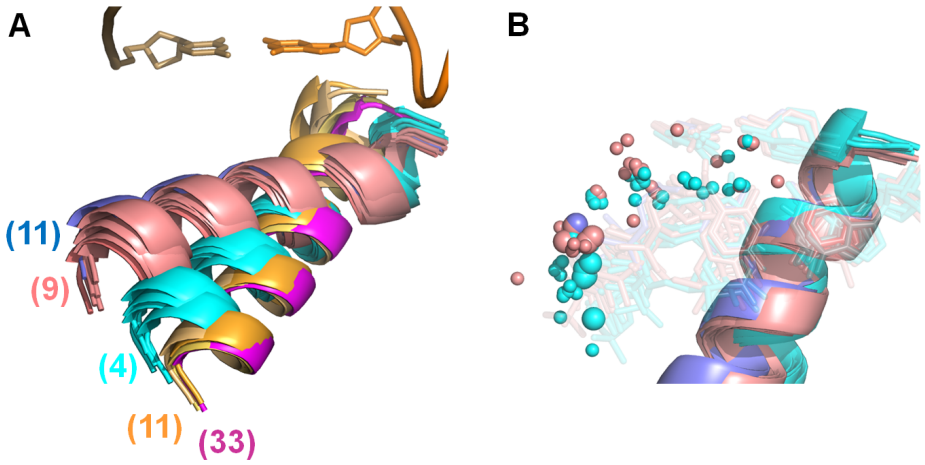


Figure 5-7 Superposition of structures of BF polymerase.
Superposition of structures of BF polymerase highlighting the conformational ensembles (A) and water structure remodeling (B).

Methods

Protein preparation − D598A/F710Y mutant BF was used to obtain closed form ternary complexes of BF polymerase, DNA duplex, and incoming nucleotide substrate at the insertion site prior to chemistry. D598A is a surface mutation that destabilizes a crystal contact thereby favoring the closed state in the crystal [62]. F710Y facilitates incorporation of a 2',3'-dideoxynucleotide chain terminator which prevents further incorporation and therefore traps ternary complexes before chemistry [27, 94]. This mutant protein was expressed and purified as described [27].

Crystallization of nucleotide substrates and DNA complexes − Oligonucleotides were synthesized by Midland Certified Reagent Co. (Midland, TX) at GF grade. BF-DNA(dG)-ddGTP, BF-DNA(dG)-ddTTP, and BF-DNA(dT)-ddTTP ternary complexes were crystallized using the same primer (5′-CCTGACTC-3′) and different templates (5′-CATGGGAGTCAGG-3′, 5′-CATGAGAGTCAGG-3′, and 5′-CATTAGAGTCAGG-3′, respectively). The template is designed such that ddNTP is first incorporated into the primer 3' terminus opposite of a complementary template base by BF polymerase to prevent further catalysis. The ddNTP is then trapped prior to chemistry to form desired mismatched base pair at the insertion site. BF-DNA-rCTP•dG ternary complex was crystallized using 5′-CCTGACTC$_{dd}$C-3′ as primer (pre-synthesized dideoxynucleotide terminus traps the rCTP complex prior to chemical incorporation) and the same template as for the BF-DNA(dG)-ddGTP complex. The primer and template strands were annealed as described [40]. Ultrapure ddNTP and rCTP were purchased from USB Co. (Cleveland, OH). Protein, primer-template duplex (protein:DNA, a 1:3 molar ratio), nucleotides (ddNTP or rCTP, 10 mM; Mg^{2+} or Mn^{2+}, SO_4^{2-} salts, 20 mM) were incubated (1hr, room temperature). The complexes were then crystallized as described previously [40].

Diffraction data collection − ddGTP•dG, ddTTP•dG, and ddTTP•dT crystals were flash frozen in liquid nitrogen directly out of the crystallization drop. rCTP•dG crystals were frozen after a three-step cryo-protecting procedure containing sucrose described previously [79]. The crystals grown in the presence of Mg^{2+} or Mn^{2+} were first transferred into stabilization solution supplemented with rCTP (55% saturated $(NH_4)_2SO_4$, 80 mM MES pH 7.3, 10 mM rCTP, and 20 mM $MgSO_4$ or $MnSO_4$, respectively) for 5 min. They were then transferred into stabilization solution with 10% sucrose added (soak for 5 min) before transferred into stabilization solution with 20% sucrose added (soak for 24 hrs at 17 ˚C). Data were collected at SIBYLS and SER-CAT beamlines.

Data processing and structure determination − All crystals belong to the space group $P2_12_12_1$ and contain two molecules in the asymmetric unit. Diffraction data were processed using XDS [104]. ddGTP•dG, ddTTP•dG, and ddTTP•dT crystals have ice rings and the reflections in shells (2.28-2.22 Å; 1.948-1.888 Å) were excluded. ddGTP•dG, ddTTP•dG, ddTTP•dT, and rCTP•dG-Mg^{2+} structures were determined using 2HVI [62] as the starting model as described [27]. Fully built and refined structure of rCTP•dG-Mg^{2+} complex with water molecules removed was used as the starting models for its Mn^{2+} counterpart. Free reflections were generated in XDSCONV [104] by combining inherited free reflections from the starting model and 5% randomly selected reflections beyond the resolution of the starting model. Data and refinement statistics are presented in Table 5-1. Composite omit maps were generated in CNS [110]. Figures and superpositions (Cα atoms of BF residues 646–655, 823–838, and 863–869 were used) were prepared in PyMOL (Schrödinger, LLC.).

Chapter 6 Summary and future direction

Summary

High-fidelity DNA polymerase replicates the genome with remarkably high accuracy. Yet, base-pair mismatches do occasionally arise leading to spontaneous mutagenesis. In this dissertation, examples of both accurate and mutagenic replication have been observed using X-ray crystallography at the active site of a model system for studying DNA replication, *Bacillus* DNA polymerase I large fragment. The series of high-resolution structures of base-pair mismatches and incorrect sugar substrates bound to fidelity filter sites of BF polymerase collectively illustrate a unified picture of nucleotide selectivity by DNA polymerase to ensure genome integrity.

Base pairs with cognate base-pair shape and size promote fully closed O helix conformation and are aligned properly at the active site for chemistry. These include all four cognate base pairs and a C•A mismatch that can mimic the shape and size of cognate base pairs.

In contrast, incorrect sugar substrates and mismatches are misaligned at a number of non-productive conformations which prevents the misincorpation of these non-cognate nucleotides. Central to the nucleotide selection mechanism is that incorrect substrates trap the polymerase at additional O helix conformations intermediate between the open and closed states creating an ensemble of binding sites that misalign non-cognate nucleotides. Water-mediated interactions, absent in the fully closed state, play an important role in formation of these binding sites, and can be remodeled to accommodate different non-cognate substrates.

Given the common presence of large scale DNA polymerase conformational changes and utilization of the same catalytic mechanism, these observations and their implications are likely to be extended to other high-fidelity DNA polymerases as well.

Future direction

Recently, a G•T mismatch bound at BF polymerase insertion site was captured which adopts a near-cognate shape, providing another example of how mutagenic replication occurs. Comparisons between the structure of G•T mismatch with that of a cognate A•T show that the two complexes are almost identical (Figure 6-1; Table 6-1). The G•T mismatch adopts a near-cognate shape and is properly aligned at the active site for chemistry. Base-pair parameters of the G•T mismatch are close to the average values of cognate base pairs captured at the same site (Table 6-2). BF polymerase is fully closed and the water layer at the active site is displaced. This complex was captured in the presence of Mn^{2+} the existence of which was confirmed by anomalous signal (Figure 6-1C).

Although a wide range of incorrect nucleotide substrates have been captured so far at the high-fidelity BF DNA polymerase active site, future efforts to capture more examples of mismatches and lesions under both accurate and mutagenic experimental conditions would be of interest. Additional structures may add more intermediate conformations to the current ensemble and reveal novel nucleotide binding modes that haven't been seen yet. We predict that new complexes are still likely to fall into the same general scheme of DNA polymerase substrate selectivity mechanisms. Some unstable and transient nucleotide binding modes and DNA polymerase conformations may not be captured using traditional X-ray crystallographic method. New experimental approach might be developed to fully characterize DNA polymerase, the sophisticated replication machinery.

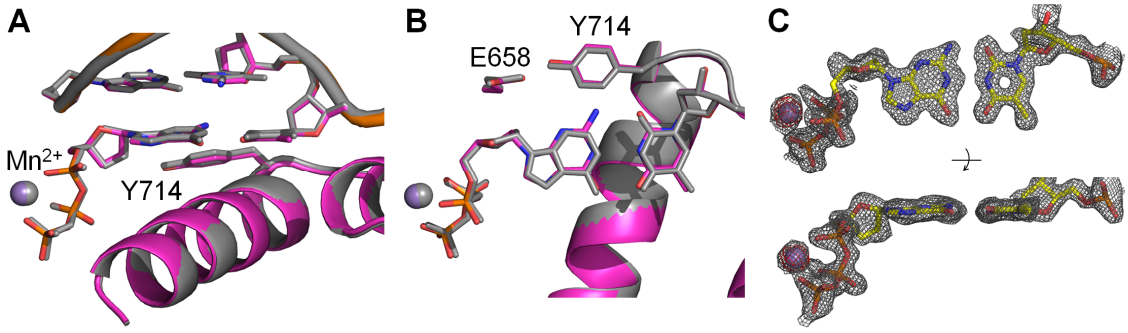


Figure 6-1 Comparisons of G•T mismatch and cognate A•T base pair at the insertion site
A, B, Two views showing the superimposed G•T mismatch (magenta) on cognate A•T base pair (gray). The two structures are almost identical. C, Composite-omit-map of the G•T base pair (gray mesh, at 1.2σ level). Anomalous map of the Mn^{2+} is also shown (red mesh, at 4.0σ)

Table 6-1 Crystallographic data collection and refinement statistics

	ddGTP•dT (*n*)
Data collection	
Wavelength	1.0000
Resolution (Å)	50-1.60
Outer shell (Å)	1.69-1.60
R_{sym} (%)	4.5(50.4)
$I / \sigma I$	21.3(3.6)
Completeness (%)	92.1(100)
Redundancy	5.7(5.6)
Refinement	
Resolution (Å)	30.1-1.60
No. reflections	179457
R_{work} / R_{free}	17.8/20.1
No. non-hydrogen atoms	
Total	11846
Solvent	1590
B-factor	30.0
R.m.s. deviations	
Bond lengths (Å)	0.006
Bond angles (°)	1.103

Table 6-2 Base pair parameters at the insertion site

Base pair	λ_{primer}(◦)[a]	$\lambda_{template}$(◦)[a]	$d_{C1'-C1'}$(Å)	Shear(Å)	Stretch(Å)	Stagger(Å)	Buckle(◦)	Propeller(◦)	Opening(◦)
ddGTP•dT	55.8	49.4	10.7	0.15	-0.29	-0.11	1.57	-8.37	-5.56
Watson-Crick[b]	57.8±1.8	54.7±0.9	10.6±0.1	0.06±0.28	-0.10±0.05	-0.09±0.08	4.31±4.51	-6.88±3.41	1.86±0.80

[a] λ_{primer} and $\lambda_{template}$ are defined as the angle between the glycosidic bond of primer or template nucleotide and the line draw between the C1′ atoms of the base pair (see inserted panel on the right). $d_{C1'-C1'}$ is the distance between the C1′ atoms of the base pair. All other base pair parameters are defined [118]. All values were calculated in 3DNA [119].

[b] Average values and standard deviations over all four cognate base pairs observed at the insertion site were adapted from [101].

Appendix A

Crystal structures of *Bacillus* DNA polymerase I large fragment deposited in the Protein Data Bank (PDB)

Structure description	PDB code	BF mutation	No. of molecules in asymmetric unit
Sugar discrimination [27]			
dCTP-dG-insertion	4DQI	D598A	2
ddCTP-dG-insertion	4DQP	D598A	2
rCTP-dG (Mg^{2+})-insertion	4DS5	D598A	2
rCTP-dG (Mn^{2+})-insertion	4DS4	D598A	2
F710Y-rCTP-dG (Mg^{2+})-insertion	4DSE	F710Y/D598A	2
F710Y-rCTP-dG (Mn^{2+})-insertion	4DSF	F710Y/D598A	2
E658A-rCTP-dG(Mg^{2+})-insertion	4DQQ	E658A/D598A	2
E658A-rCTP-dG(Mn^{2+})-insertion	4DQR	E658A/D598A	2
E658A-dCTP-dG(Mg^{2+})	4E0D	E658A/D598A	1
rC-dG(*n*-1)	4DQS	WT	1
C•A mismatch [101]			
ddCTP-dA cognate-insertion	3PX6	F710Y/D598A	2
ddCTP-dA wobble-insertion	3PX4	F710Y/D598A	2
dCTP-dA congate-insertion	3PX0	F710Y/D598A	2
ddTTP-dA-insertion	3PV8	F710Y/D598A	2
ddATP-dT-insertion	3THV	F710Y/D598A	2
ddGTP-dC-insertion	3TI0	F710Y/D598A	2
C-A(*n*-1)	3TAN	WT	1
C-A(*n*-3)	3TAP	WT	1
C-A(*n*-4)	3TAQ	WT	1
C-A(*n*-6)	3TAR	WT	1
Ajar intermediate T•G mismatch [79]			
ddTTP-dG insertion	3HP6	F710Y/D598A	2
dCTP-dG insertion-ajar	3HT3	V713P/D598A	2
dTTP-dG insertion	3HPO	Y714S/D329A	1
O^6-methyl-guanine [62]			
ddCTP-dG insertion	2HVI	F710Y/D598A	2
ddTTP-O^6MG insertion	2HHW	F710Y/D598A	2
ddCTP-O^6MG insertion	2HVH	F710Y/D598A	2
O^6MG pre-insertion	2HHX	WT	1

T-O^{6}MG(*n*-1)	2HW3	WT	1
T-O^{6}MG(*n*-2)	2HHV	WT	1
T-O^{6}MG(*n*-10)	2HHQ	WT	1
C-O^{6}MG(*n*-1)	2HHU	WT	1
C-O^{6}MG(*n*-2)	2HHT	WT	1
C-O^{6}MG(*n*-10)	2HHS	WT	1
Benzo[α]pyrene [63]			
C-Benzopyrene-G(*n*-1)	1XC9	WT	1
Aromatic amine adduct [64]			
G-AF-pre-insertion	1UA0	WT	1
C-G-AF(*n*-1)	1UA1	WT	1
Mismatches after incorporation [61]			
A-A(*n*-1)	1NK5	WT	1
C-T(*n*-1)	1NJZ	WT	1
C-T(*n*-2)	1NKE	WT	1
G-A(*n*-1)	1NK7	WT	1
T-T(*n*-1)	1NJY	WT	1
G-T(*n*-3)	1NK9	WT	1
G-T(*n*-4)	1NKB	WT	1
G-T(*n*-6)	1NKC	WT	1
A-G(*n*-1)	1NK0	WT	1
C-C(*n*-1)	1NK6	WT	1
G-G(*n*-1)	1NK4	WT	1
G-T(*n*-1)	1NJW	WT	1
G-T(*n*-2)	1NK8	WT	1
T-G(*n*-1)	1NJX	WT	1
8-oxoguanine [65]			
8oxoG-pre-insertion	1U45	WT	1
C-8oxoG(*n*-1)	1U47	WT	1
C-8oxoG(*n*-2)	1U48	WT	1
A-8oxoG(*n*-1)	1U49	WT	1
A-8oxoG(*n*-3)	1U4B	WT	1
Processive DNA synthesis [40]			
9bp	1L3S	WT	1
10bp	1L3T	WT	1

11bp	1L3U	WT	1
12bp	1L5U	WT	1
15bp	1L3V	WT	1
dCTP-dG	1LV5	D329A	2
Visualizing DNA replication [60]			
9bp	2BDP	WT	1
10bp (ddT-dA)	3BDP	WT	1
11bp	4BDP	WT	1
BF apo structure [28]			
Apo	1XWL/1BDP	WT	1

References

1. Kornberg, A. and T.A. Baker, *DNA replication*. 2 ed. 1992, New York, New York: W. H. Freeman and Company.

2. Johnson, K.A., *Conformational coupling in DNA polymerase fidelity.* Annu. Rev. Biochem., 1993. **62**: p. 685-713.

3. Joyce, C.M. and T.A. Steitz, *Function and structure relationships in DNA polymerases.* Annu. Rev. Biochem., 1994. **63**: p. 777-822.

4. Echols, H. and M.F. Goodman, *Fidelity mechanisms in DNA replication.* Annu. Rev. Biochem., 1991. **60**: p. 477-511.

5. Kunkel, T.A. and K. Bebenek, *DNA replication fidelity.* Annu. Rev. Biochem., 2000. **69**: p. 497-529.

6. Kunkel, T.A. and K. Bebenek, *Recent studies of the fidelity of DNA synthesis.* Biochim Biophys Acta, 1988. **951**(1): p. 1-15.

7. Bebenek, K. and T.A. Kunkel, *The Fidelity of Retroviral Reverse Transcriptases*, in *Reverse Transcriptase*, A.M. Skalka and S.P. Goff, Editors. 1993, Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY. p. 85-102.

8. Roberts, J.D. and T.A. Kunkel, *Fidelity of DNA Replication*, in *DNA Replication in Eukaryotic Cells: Concepts, Enzymes and Systems*, M.D. Pamphilis, Editor. 1996, Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY. p. 217-247.

9. Joyce, C.M., *Choosing the right sugar: How polymerases select a nucleotide substrate.* Proc.Natl.Acad.Sci.USA., 1997. **94**: p. 1619-1622.

10. Brown, J.A. and Z. Suo, *Unlocking the sugar "steric gate" of DNA polymerases.* Biochemistry, 2011. **50**(7): p. 1135-42.

11. Bebenek, K. and T.A. Kunkel, *Functions of DNA polymerases.* Adv Protein Chem, 2004. **69**: p. 137-65.

12. Delarue, M., et al., *An attempt to unify the structure of polymerases.* Protein Eng., 1990. **3**(6): p. 461-467.

13. Ito, J. and D.K. Braithwaite, *Compilation and alignment of DNA polymerase sequences.* Nucleic Acids Res, 1991. **19**(15): p. 4045-57.

14. Braithwaite, D.K. and J. Ito, *Compilation, alignment, and phylogenetic relationships of DNA polymerases.* Nucleic Acids Research, 1993. **21**: p. 787-802.

15. Steitz, T.A., et al., *A unified polymerase mechanism for nonhomologous DNA and RNA polymerases.* Science, 1994. **266**: p. 2022-2025.

16. Rothwell, P.J. and G. Waksman, *Structure and mechanism of DNA polymerases.* Advances in protein chemistry, 2005. **71**.

17. Garcia-Diaz, M., et al., *Structure-function studies of DNA polymerase lambda.* DNA Repair (Amst), 2005. **4**(12): p. 1358-67.

18. Prakash, S., R.E. Johnson, and L. Prakash, *Eukaryotic translesion synthesis DNA polymerases: specificity of structure and function.* Annu Rev Biochem, 2005. **74**: p. 317-53.

19. Ohmori, H., et al., *The Y-family of DNA polymerases.* Mol Cell, 2001. **8**(1): p. 7-8.

20. Rothwell, P.J. and G. Waksman, *Structure and mechanism of DNA polymerases.* Adv Protein Chem, 2005. **71**: p. 401-40.

21. Lewis, W., B.J. Day, and W.C. Copeland, *Mitochondrial toxicity of NRTI antiviral drugs: an integrated cellular perspective.* Nat Rev Drug Discov, 2003. **2**(10): p. 812-22.

22. Cann, I.K. and Y. Ishino, *Archaeal DNA replication: identifying the pieces to solve a puzzle.* Genetics, 1999. **152**(4): p. 1249-67.

23. Steitz, T.A., *DNA polymerases: structural diversity and common mechanisms.* J Biol Chem, 1999. **274**(25): p. 17395-8.

24. Ollis, D.L., et al., *Structure of large fragment of Escherichia coli DNA polymerase I complexed with dTMP.* Nature, 1985. **313**: p. 762-766.

25. Steitz, T.A. and Y.W. Yin, *Accuracy, lesion bypass, strand displacement and translocation by DNA polymerases.* Philos Trans R Soc Lond B Biol Sci, 2004. **359**(1441): p. 17-23.

26. Steitz, T.A. and J.A. Steitz, *A general two-metal-ion mechanism for catalytic RNA.* Proc Natl Acad Sci U S A, 1993. **90**(14): p. 6498-502.

27. Wang, W., et al., *Structural factors that determine selectivity of a high-fidelity DNA polymerase for deoxy-, dideoxy-, and ribo-nucleotides.* To be published, 2012.

28. Kiefer, J.R., et al., *Crystal structure of a thermostable Bacillus DNA polymerase I large fragment at 2.1 A resolution.* Structure, 1997. **5**(1): p. 95-108.

29. Wang, M., et al., *Insights into base selectivity from the 1.8 A resolution structure of an RB69 DNA polymerase ternary complex.* Biochemistry, 2011. **50**(4): p. 581-90.

30. Wang, J., et al., *Crystal structure of a pol a family replication DNA polymerase from bacteriophage RB69.* Cell, 1997. **89**(7): p. 1087-1099.

31. Evans, R.J., et al., *Structure of PolC reveals unique DNA binding and fidelity determinants.* Proc Natl Acad Sci U S A, 2008. **105**(52): p. 20695-700.

32. Batra, V.K., et al., *Magnesium-induced assembly of a complete DNA polymerase catalytic complex.* Structure, 2006. **14**(4): p. 757-66.

33. Beard, W.A. and S.H. Wilson, *Structure and mechanism of DNA polymerase Beta.* Chemical Reviews, 2006. **106**(2): p. 361-82.

34. Biertumpfel, C., et al., *Structure and mechanism of human DNA polymerase eta.* Nature, 2010. **465**(7301): p. 1044-8.

35. Lansdon, E.B., et al., *Visualizing the molecular interactions of a nucleotide analog, GS-9148, with HIV-1 reverse transcriptase-DNA complex.* J Mol Biol, 2010. **397**(4): p. 967-78.

36. Kohlstaedt, L.A., et al., *Crystal structure at 3.5 A resolution of HIV-1 reverse transcriptase complexed with an inhibitor.* Science, 1992. **256**(5065): p. 1783-90.

37. Li, Y., S. Korolev, and G. Waksman, *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., 1998. **17**(24): p. 7514-7525.

38. Franklin, M.C., J. Wang, and T.A. Steitz, *Structure of the replicating complex of a pol a family DNA polymerase.* Cell, 2001. **105**(5): p. 657-67.

39. Doublié, S., M.R. Sawaya, and T. Ellenberger, *An open and closed case for all polymerases.* Structure Fold Des, 1999. **7**(2): p. R31-5.

40. Johnson, S.J., J.S. Taylor, and L.S. Beese, *Processive DNA synthesis observed in a polymerase crystal suggests a mechanism for the prevention of frameshift mutations.* Proc. Natl. Acad. Sci. U. S. A., 2003. **100**(7): p. 3895-3900.

41. Huang, H., et al., *Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance.* Science, 1998. **282**(5394): p. 1669-75.

42. Sawaya, M.R., et al., *Crystal structures of human DNA polymerase beta complexed with gapped and nicked DNA: evidence for an induced fit mechanism.* Biochemistry, 1997. **36**: p. 11205-11215.

43. Wing, R.A., S. Bailey, and T.A. Steitz, *Insights into the replisome from the structure of a ternary complex of the DNA polymerase III alpha-subunit.* J Mol Biol, 2008. **382**(4): p. 859-69.

44. Pata, J.D., *Structural diversity of the Y-family DNA polymerases.* Biochim Biophys Acta, 2010. **1804**(5): p. 1124-35.

45. Beese, L.S., J.M. Friedman, and T.A. Steitz, *Crystal structures of the Klenow fragment of DNA polymerase I complexed with deoxynucleoside triphosphate and pyrophosphate.* Biochemistry, 1993. **32**(51): p. 14095-101.

46. Beese, L.S., V. Derbyshire, and T.A. Steitz, *Structure of DNA polymerase I Klenow fragment bound to duplex DNA.* Science, 1993. **260**(5106): p. 352-5.

47. Freemont, P.S., et al., *Cocrystal structure of an editing complex of Klenow fragment with DNA.* Proc.Natl.Acad.Sci.USA., 1988. **85**: p. 8924-8928.

48. Beese, L.S. and T.A. Steitz, *Structure of E. coli DNA polymerase I, large fragment, and its functional implications,* in *Nucleic Acids and Molecular Biology,* F. Eckstein and D.M.J. Lilley, Editors. 1989, Springer-Verlag: Berlin, Vol. 3. p. 28-43.

49. Beese, L.S. and T.A. Steitz, *Structural basis for the 3'-5' exonuclease activity of Escherichia coli DNA polymerase I: a two metal ion mechanism.* Embo J, 1991. **10**(1): p. 25-33.

50. Doublié, S., et al., *Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution.* Nature, 1998. **391**(6664): p. 251-258.

51. Brieba, L.G., et al., *A lysine residue in the fingers subdomain of T7 DNA polymerase modulates the miscoding potential of 8-oxo-7,8-dihydroguanosine.* Structure, 2005. **13**(11): p. 1653-9.

52. Brieba, L.G., et al., *Structural basis for the dual coding potential of 8-oxoguanosine by a high-fidelity DNA polymerase.* EMBO J, 2004. **23**(17): p. 3452-61.

53. Dutta, S., et al., *Crystal structures of 2-acetylaminofluorene and 2-aminofluorene in complex with T7 DNA polymerase reveal mechanisms of mutagenesis.* Proc Natl Acad Sci U S A, 2004. **101**(46): p. 16186-91.

54. Li, Y., et al., *Nucleotide insertion opposite a cis-syn thymine dimer by a replicative DNA polymerase from bacteriophage T7.* Nat Struct Mol Biol, 2004. **11**(8): p. 784-90.

55. Korolev, S., et al., *Crystal structure of the large fragment of Thermus aquaticus DNA polymerase I at 2.5-^{a} resolution: Structural basis for thermostability.* Proc.Natl.Acad.Sci.USA., 1995. **92**: p. 9264-9268.

56. Li, Y., et al., *Crystal structures of the Klenow fragment of Thermus aquaticus DNA polymerase I complexed with deoxyribonucleoside triphosphates.* Protein Sci., 1998. **7**: p. 1116-1123.

57. Eom, S.H., J. Wang, and T.A. Steitz, *Structure of Taq polymerase with DNA at the polymerase active site.* Nature, 1996. **382**: p. 278-281.

58. Li, Y. and G. Waksman, *Crystal structures of a ddATP-, ddTTP-, ddCTP, and ddGTP-trapped ternary complex of Klentaq1: insights into nucleotide incorporation and selectivity*. Protein Sci, 2001. **10**(6): p. 1225-33.

59. Li, Y., V. Mitaxov, and G. Waksman, *Structure-based design of Taq DNA polymerases with improved properties of dideoxynucleotide incorporation.* Proc Natl Acad Sci U S A, 1999. **96**(17): p. 9491-6.

60. Kiefer, J.R., et al., *Visualizing DNA replication in a catalytically active Bacillus DNA polymerase crystal.* Nature, 1998. **391**(6664): p. 304-7.

61. Johnson, S.J. and L.S. Beese, *Structures of mismatch replication errors observed in a DNA polymerase.* Cell, 2004. **116**(6): p. 803-16.

62. Warren, J.J., L.J. Forsberg, and L.S. Beese, *The structural basis for the mutagenicity of O(6)-methyl-guanine lesions.* Proc Natl Acad Sci U S A, 2006. **103**(52): p. 19701-6.

63. Hsu, G.W., et al., *Structure of a high fidelity DNA polymerase bound to a benzo[a]pyrene adduct that blocks replication.* J. Biol. Chem., 2005. **280**(5): p. 3764-70.

64. Hsu, G.W., et al., *Observing translesion synthesis of an aromatic amine DNA adduct by a high-fidelity DNA polymerase.* J. Biol. Chem., 2004. **279**(48): p. 50280-5.

65. Hsu, G.W., et al., *Error-prone replication of oxidatively damaged DNA by a high-fidelity DNA polymerase.* Nature, 2004. **431**(7005): p. 217-21.

66. Lee, Y.S., W.D. Kennedy, and Y.W. Yin, *Structural insight into processive human mitochondrial DNA synthesis and disease-related polymerase mutations.* Cell, 2009. **139**(2): p. 312-24.

67. Berman, A.J., et al., *Structures of phi29 DNA polymerase complexed with substrate: the mechanism of translocation in B-family polymerases.* EMBO J, 2007. **26**(14): p. 3494-505.

68. Beard, W.A., et al., *DNA polymerase beta substrate specificity: side chain modulation of the "A-rule".* J Biol Chem, 2009. **284**(46): p. 31680-9.

69. Kunkel, T.A., *DNA replication fidelity*. J Biol Chem, 2004. **279**(17): p. 16895-8.

70. Doublie, S. and T. Ellenberger, *The mechanism of action of T7 DNA polymerase.* Curr Opin Struct Biol, 1998. **8**(6): p. 704-12.

71. Brautigam, C.A. and T.A. Steitz, *Structural and functional insights provided by crystal structures of DNA polymerases and their substrate complexes.* Curr. Opin. Struct. Biol., 1998. **8**(1): p. 54-63.

72. Kool, E.T., *Active site tightness and substrate fit in DNA replication.* Annu. Rev. Biochem., 2002. **71**: p. 191-219.

73. Joyce, C.M. and S.J. Benkovic, *DNA polymerase fidelity: kinetics, structure, and checkpoints.* Biochemistry, 2004. **43**(45): p. 14317-24.

74. Golosov, A.A., et al., *The mechanism of the translocation step in DNA replication by DNA polymerase I: a computer simulation analysis.* Structure, 2010. **18**(1): p. 83-93.

75. Watson, J.D. and F.H.C. Crick, *Genetical implications of the structure of deoxyribonucleic acid.* Nature, 1953. **171**: p. 964-967.

76. Mendelman, L.V., et al., *Nearest neighbor influences on DNA polymerase insertion fidelity.* J.Biol.Chem., 1989. **264**: p. 14415-14423.

77. Eckert, K.A. and T.A. Kunkel, *Effect of reaction pH on the fidelity and processivity of exonuclease-deficient Klenow polymerase.* J.Biol.Chem., 1993. **268**: p. 13462-13471.

78. Kunkel, T.A. and P.S. Alexander, *The base substitution fidelity of eukaryotic DNA polymerases.* J.Biol.Chem., 1986. **261**: p. 160-166.

79. Wu, E.Y. and L.S. Beese, *The structure of a high fidelity DNA polymerase bound to a mismatched nucleotide reveals an "ajar" intermediate conformation in the nucleotide selection mechanism.* J Biol Chem, 2011. **286**(22): p. 19758-67.

80. Bebenek, K., L.C. Pedersen, and T.A. Kunkel, *Replication infidelity via a mismatch with Watson-Crick geometry.* Proc Natl Acad Sci U S A, 2011. **108**(5): p. 1862-7.

81. Vaisman, A., et al., *Fidelity of Dpo4: effect of metal ions, nucleotide selection and pyrophosphorolysis.* EMBO J, 2005. **24**(17): p. 2957-67.

82. Astatke, M., N.D.F. Grindley, and C.M. Joyce, *How E. coli DNA polymerase I (Klenow fragment) distinguishes between deoxy- and dideoxynucleotides.* J. Mol. Biol., 1998. **278**(1): p. 147-165.

83. Traut, T.W., *Physiological concentrations of purines and pyrimidines.* Mol Cell Biochem, 1994. **140**(1): p. 1-22.

84. Ferraro, P., et al., *Quantitation of cellular deoxynucleoside triphosphates.* Nucleic Acids Res, 2010. **38**(6): p. e85.

85. Nick McElhinny, S.A., et al., *Abundant ribonucleotide incorporation into DNA by yeast replicative polymerases.* Proc Natl Acad Sci U S A, 2010. **107**(11): p. 4949-54.

86. Nick McElhinny, S.A., et al., *Genome instability due to ribonucleotide incorporation into DNA.* Nat Chem Biol, 2010. **6**(10): p. 774-81.

87. Astatke, M., et al., *A single side chain prevents Escherichia coli DNA polymerase I (Klenow fragment) from incorporating ribonucleotides.* Proc Natl Acad Sci U S A, 1998. **95**(7): p. 3402-3407.

88. Minnick, D.T., et al., *Side chains that influence fidelity at the polymerase active site of Escherichia coli DNA polymerase I (Klenow fragment).* J. Biol. Chem., 1999. **274**(5): p. 3067-3075.

89. Kasiviswanathan, R. and W.C. Copeland, *Ribonucleotide discrimination and reverse transcription by the human mitochondrial DNA polymerase.* J Biol Chem, 2011. **286**(36): p. 31490-500.

90. Cavanaugh, N.A., et al., *Molecular insights into DNA polymerase deterrents for ribonucleotide insertion.* J Biol Chem, 2011. **286**(36): p. 31650-60.

91. Joyce, C.M., et al., *Fingers-closing and other rapid conformational changes in DNA polymerase I (Klenow fragment) and their role in nucleotide selectivity.* Biochemistry, 2008. **47**(23): p. 6103-16.

92. Santoso, Y., et al., *Conformational transitions in DNA polymerase I revealed by single-molecule FRET.* Proc Natl Acad Sci U S A, 2010. **107**(2): p. 715-20.

93. Mizrahi, V. and P. Huberts, *Deoxy- and dideoxynucleotide discrimination and identification of critical 5' nuclease domain residues of the DNA polymerase I from Mycobacterium tuberculosis.* Nucleic Acids Res, 1996. **24**(24): p. 4845-52.

94. Tabor, S. and C.C. Richardson, *A single residue in DNA polymerases of the Escherichia coli DNA polymerase I family is critical for distinguishing between deoxy- and dideoxyribonucleotides.* Proc. Natl. Acad. Sci. U. S. A., 1995. **92**: p. 6339-6343.

95. Topal, M.D. and J.R. Fresco, *Complementary base pairing and the origin of substitution mutations.* Nature, 1976. **263**(5575): p. 285-289.

96. Sowers, L.C., et al., *DNA base modification: ionized base pairs and mutagenesis.* Mutat Res, 1987. **177**(2): p. 201-18.

97. Lawley, P.D. and P. Brookes, *Ionization of DNA Bases or Base Analogues as a Possible Explanation of Mutagenesis, with Special Reference to 5-bromodeoxyuridine.* J. Mol. Biol., 1962. **4**: p. 216-219.

98. Goodman, M.F., et al., *Biochemical basis of DNA replication fidelity.* Crit Rev Biochem Mol Biol, 1993. **28**(2): p. 83-126.

99. Joyce, C.M. and N.D.F. Grindley, *Identification of two genes immediatly downstream from the polA gene of Escherichia coli.* J.Biol.Chem., 1982. **257**: p. 1958-1964.

100. Lawyer, F.C., et al., *Isolation, characterization, and expression in Escherichia coli of DNA polymerase gene from Thermus aquaticus.* J.Biol.Chem., 1989. **264**: p. 6427-6437.

101. Wang, W., H.W. Hellinga, and L.S. Beese, *Structural evidence for the rare tautomer hypothesis of spontaneous mutagenesis.* Proc Natl Acad Sci U S A, 2011. **108**(43): p. 17644-8.

102. Weymouth, L.A. and L.A. Loeb, *Mutagenesis during in vitro DNA synthesis.* Proc Natl Acad Sci U S A, 1978. **75**(4): p. 1924-8.

103. Sirover, M.A. and L.A. Loeb, *Infidelity of DNA synthesis in vitro: Screening for potential metal mutagens or carcinogens.* Science, 1976. **194**: p. 1434-1436.

104. Kabsch, W., *Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell dimensions.* J. Appl. Cryst., 1993. **26**: p. 795-800.

105. Otwinowski, Z. and W. Minor, *Processing of X-ray diffraction data collected in oscillation mode.* Methods Enzymol., 1997. **276A**: p. 307-326.

106. No.4, C.C.P., *The CCP4 Suite: programs for protein crystallography.* Acta Crystallographica, 1994. **50**: p. 760-763.

107. Murshudov, G.N., A.A. Vagin, and E.J. Dodson, *Refinement of Macromolecular Structures by the Maximum-Likelihood Method* Acta Cryst. D, 1997. **53**: p. 240-255.

108. Afonine, P.V., Grosse-Kunstleve, R.W. & Adams, P.D., *The Phenix refinement framework.* CCP4 Newsletter on protein crystallography, 2005. **42** (contribution 8).

109. Emsley, P. and K. Cowtan, *Coot: Model-Building Tools for Molecular Graphics.* Acta Cryst. D, 2004. **60**: p. 2126-2132.

110. Brünger, A.T., et al., *Crystallography & NMR System: A New Software Suite for Macromolecular Structure Determination.* Acta Cryst. D, 1998. **54**: p. 905-921.

111. Kunkel, T.A. and K. Bebenek, *DNA replication fidelity.* Annu Rev Biochem, 2000. **69**: p. 497-529.

112. Echols, H. and M.F. Goodman, *Fidelity mechanisms in DNA replication.* Annu Rev Biochem, 1991. **60**: p. 477-511.

113. Harris, V.H., et al., *The effect of tautomeric constant on the specificity of nucleotide incorporation during DNA replication: support for the rare tautomer hypothesis of substitution mutagenesis.* J Mol Biol, 2003. **326**(5): p. 1389-401.

114. Hunter, W.N., et al., *Structure of an adenine-cytosine base pair in DNA and its implications for mismatch repair.* Nature, 1986. **320**(6062): p. 552-555.

115. Boulard, Y., et al., *The pH dependent configurations of the C.A mispair in DNA.* Nucleic Acids Res, 1992. **20**(8): p. 1933-41.

116. Steitz, T.A., *DNA- and RNA-dependent DNA polymerases.* Curr. Opin. Struct. Biol., 1993. **3**: p. 31-38.

117. Vaguine, A.A., J. Richelle, and S.J. Wodak, *SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model.* Acta Crystallogr D Biol Crystallogr, 1999. **55**(Pt 1): p. 191-205.

118. Dickerson, R.E., et al., *Definitions and nomenclature of nucleic acid structure parameters.* J. Mol. Biol., 1989. **205**: p. 787-791.

119. Lu, X.J. and W.K. Olson, *3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures.* Nucleic Acids Res, 2003. **31**(17): p. 5108-21.

120. Frederico, L.A., T.A. Kunkel, and B.R. Shaw, *Cytosine deamination in mismatched base pairs.* Biochemistry, 1993. **32**(26): p. 6523-6530.

121. Fonseca Guerra, C., et al., *Adenine tautomers: relative stabilities, ionization energies, and mismatch with cytosine.* J Phys Chem A, 2006. **110**(11): p. 4012-20.

122. Fogarasi, G., *Water-mediated tautomerization of cytosine to the rare imino form: An ab initio dynamics study.* Chemical Physics, 2008. **349**(1-3): p. 204-209.

123. Hobza, P. and J. Sponer, *Structure, energetics, and dynamics of the nucleic acid base pairs: Nonempirical ab initio calculations.* Chemical Reviews, 1999. **99**(11): p. 3247-3276.

124. Garcia-Diaz, M., et al., *A closed conformation for the Pol lambda catalytic cycle.* Nat Struct Mol Biol, 2005. **12**(1): p. 97-8.

125. Wang, J., *DNA polymerases: Hoogsteen base-pairing in DNA replication?* Nature, 2005. **437**(7057): p. E6-7; discussion E7.

126. Adams, P.D., et al., *PHENIX: a comprehensive Python-based system for macromolecular structure solution.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 213-21.

127. Crick, F., *Central dogma of molecular biology.* Nature, 1970. **227**: p. 561-563.

128. Henry, A.A. and F.E. Romesberg, *The evolution of DNA polymerases with novel activities.* Curr Opin Biotechnol, 2005. **16**(4): p. 370-7.

129. Holmberg, R.C., A.A. Henry, and F.E. Romesberg, *Directed evolution of novel polymerases.* Biomol Eng, 2005. **22**(1-3): p. 39-49.

130. Patel, P.H. and L.A. Loeb, *Multiple amino acid substitutions allow DNA polymerases to synthesize RNA.* J Biol Chem, 2000. **275**(51): p. 40266-72.

131. Yang, G., et al., *A conserved Tyr residue is required for sugar selectivity in a Pol alpha DNA polymerase.* Biochemistry, 2002. **41**(32): p. 10256-61.

132. Bonnin, A., et al., *A single tyrosine prevents insertion of ribonucleotides in the eukaryotic-type phi29 DNA polymerase.* J Mol Biol, 1999. **290**(1): p. 241-51.

133. Brown, J.A., et al., *A novel mechanism of sugar selection utilized by a human X-family DNA polymerase.* J Mol Biol, 2010. **395**(2): p. 282-90.

134. Cases-Gonzalez, C.E., M. Gutierrez-Rivas, and L. Menendez-Arias, *Coupling ribose selection to fidelity of DNA synthesis. The role of Tyr-115 of human immunodeficiency virus type 1 reverse transcriptase.* J Biol Chem, 2000. **275**(26): p. 19759-67.

135. Gao, G., et al., *Conferring RNA polymerase Activity to a DNA polymerase: A single residue in reverse transcriptase controls substrate selection.* Proc.Natl.Acad.Sci.USA., 1997. **94**: p. 407-411.

136. Sherrer, S.M., et al., *Kinetic basis of sugar selection by a Y-family DNA polymerase from Sulfolobus solfataricus P2.* Biochemistry, 2010. **49**(47): p. 10179-86.

137. DeLucia, A.M., N.D. Grindley, and C.M. Joyce, *An error-prone family Y DNA polymerase (DinB homolog from Sulfolobus solfataricus) uses a 'steric gate' residue for discrimination against ribonucleotides.* Nucleic Acids Res, 2003. **31**(14): p. 4129-37.

138. Kirouac, K.N., Z. Suo, and H. Ling, *Structural mechanism of ribonucleotide discrimination by a y-family DNA polymerase.* J Mol Biol, 2011. **407**(3): p. 382-90.

139. Van de Sande, J.H., P.C. Loewen, and H.G. Khorana, *Studies on polynucleotides. 118. A further study of ribonucleotide incorporation into deoxyribonucleic acid chains by deoxyribonucleic acid polymerase I of Escherichia coli.* J Biol Chem, 1972. **247**(19): p. 6140-8.

140. McCulloch, S.D. and T.A. Kunkel, *The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases.* Cell Res, 2008. **18**(1): p. 148-61.

141. Petruska, J., L.C. Sowers, and M.F. Goodman, *Comparison of nucleotide interactions in water, proteins, and vacuum: model for DNA polymerase fidelity.* Proc Natl Acad Sci U S A, 1986. **83**(6): p. 1559-62.

142. Smock, R.G. and L.M. Gierasch, *Sending signals dynamically.* Science, 2009. **324**(5924): p. 198-203.

143. Benkovic, S.J., G.G. Hammes, and S. Hammes-Schiffer, *Free-energy landscape of enzyme catalysis.* Biochemistry, 2008. **47**(11): p. 3317-21.

144. Nashine, V.C., S. Hammes-Schiffer, and S.J. Benkovic, *Coupled motions in enzyme catalysis.* Curr Opin Chem Biol, 2010. **14**(5): p. 644-51.

145. Goodey, N.M. and S.J. Benkovic, *Allosteric regulation and catalysis emerge via a common route.* Nat Chem Biol, 2008. **4**(8): p. 474-82.

146. Reynolds, K.A., R.N. McLaughlin, and R. Ranganathan, *Hot spots for allosteric regulation on protein surfaces.* Cell, 2011. **147**(7): p. 1564-75.

147. McCoy, A.J., et al., *Phaser crystallographic software.* J Appl Crystallogr, 2007. **40**(Pt 4): p. 658-674.

148. Davis, I.W., et al., *MolProbity: all-atom contacts and structure validation for proteins and nucleic acids.* Nucleic Acids Res, 2007. **35**(Web Server issue): p. W375-83.

149. Leontis, N.B., J. Stombaugh, and E. Westhof, *The non-Watson-Crick base pairs and their associated isostericity matrices.* Nucleic Acids Res, 2002. **30**(16): p. 3497-531.

150. Saenger, W., *Principles of Nucleic Acid Structure.* 1984, New York, New York:: Springer-Verlag.

Biography

Weina Wang

Date & Place of Birth: March 8, 1985
Qingjian, Shaanxi, China

Education

Ph.D. Biochemistry, Duke University, Durham, North Carolina, May 2012

B.S. Biological Science, Tsinghua University, Beijing, China, July 2006

Publications

1. Wang, W., Hellinga, H.W., and Beese, L.S. (2011). Structural evidence for the rare tautomer hypothesis of spontaneous mutagenesis. *Proc. Natl. Acad. Sci. U. S. A.* 108(43): 17644-8.
2. Wang, W., Wu, E.Y., Hellinga, H.W., and Beese, L.S. (2012). Structural factors that determine selectivity of a high-fidelity DNA polymerase for deoxy-, dideoxy-, and ribo-nucleotides. *In submission*.
3. Wang, W., Hellinga, H.W., and Beese, L.S. (2012). Structural basis for ensembles of incorporation pathways of a high-fidelity DNA polymerase. *In preparation*.
4. Wang, W., Grimley, J.S., Beese, L.S., and Hellinga H.W. (2012). The diversity of engineered binding sites in fluorescent proteins reveals principles of halide recognition. *In submission*.
5. Grimley, J.S., Li L., Wang, W., Beese, L.S., Augustine, G.J., and Hellinga H.W. (2012). Visualization of synaptic inhibition with an optogenetic sensor optimized by protein engineering automation. *In submission*.

Awards & honors

Linus Pauling Prize, American Crystallographic Association Annual Meeting (2011)

Invited Oral Presentation, American Crystallographic Association Annual Meeting (2010)

Rigaku Poster Award, Mid-Atlantic Macromolecular Crystallography Meeting (2009)

Kamin Poster Award, Department of Biochemistry, Duke University (2008)