Available online at www.sciencedirect.com
SCIENCE DIRECT
DNA Repair 4 (2005) 1390–1398
DNA REPAIR
www.elsevier.com/locate/dnarepair

Mutability of DNA polymerase I: Implications for the creation of mutant DNA polymerases

Ern Loh, Lawrence A. Loeb *

Joseph Gottstein Memorial Cancer Laboratory, Department of Pathology, University of Washington, Seattle, WA 98195, USA

Available online 17 October 2005

Abstract

DNA polymerases of the Family A catalyze the addition of deoxynucleotides to a primer with high efficiency, processivity, and selectivity—properties that are critical to their function both in nature and in the laboratory. These polymerases tolerate many amino acid substitutions, even in regions that are evolutionarily conserved. This tolerance can be exploited to create DNA polymerases with novel properties and altered substrate specificities, using rational design and molecular evolution. These efforts have focused mainly on the Family A DNA polymerises – *Taq*, *E. coli* Pol I, and T7 – because they are widely utilized in biotechnology today. The redesign of polymerases often requires knowledge of the function of specific residues in the protein, including those located in six evolutionarily conserved regions. The most well characterized of these are motifs A and B, which regulate the fidelity of replication and the incorporation of nucleotide analogs such as dideoxynucleotides. Regions that remain to be more thoroughly characterized are motif C, which is critical for catalysis, and motifs 1, 2 and 6, all of which bind to DNA primer or template. Several recently identified mutants with abilities to incorporate nucleotides with bulky adducts have mutations that are not located within conserved regions and warrant further study. Analysis of these mutants will help advance our understanding of how DNA polymerases select bases with high fidelity.

Keywords: DNA polymerase; Replication fidelity; *Taq*; Pol I; Molecular evolution

1. Introduction

DNA polymerases are central to the maintenance of the genome. They catalyze the template-directed addition of deoxynucleotides onto a DNA primer and function in replication, repair, and recombination of DNA. Many DNA polymerases have evolved the ability to select substrates with exquisite stringency, and catalyze DNA synthesis with high efficiency and processivity. They can select against the incorporation of non-complementary bases and ribonucleotides, with an accuracy greater than 99.99%—all the while completing each reaction in a matter of milliseconds. Defects in this fidelity have been postulated to result in premature aging and an increased incidence of cancer, myopathies, and neuropathies [1–3].

The selectivity of DNA polymerases has been important not only to biologists but also to biotechnologists. Thermostable polymerases with high fidelity facilitate cloning and PCR-amplification while low fidelity polymerases are central to mutagenic PCR. DNA polymerases that more efficiently incorporate dideoxynucleotides or fluorescently labeled nucleotides have been developed for Sanger sequencing and microarray production, respectively. Retroviral DNA polymerases (reverse transcriptases) are used routinely to copy messenger RNAs. DNA polymers with modifications of the phosphate backbone that increases their nuclease resistance or neutralize their electronegativity have also been investigated for potential use in medicine [4]. Interest in these different DNA processes fuels efforts to create new polymerases with altered substrate specificities. Several DNA polymerases used today are in fact modified from their wildtype sequences—with the most notable being those used in mutagenic PCR, in vivo mutagenesis, and Sanger sequencing (Table 1). Moreover, efforts are currently underway to create polymerases with other novel properties such as greater accuracy, the ability to bypass bulky adducts efficiently, and the ability to incorporate an increasing array of nucleotide analogs.

DNA polymerases, like many other enzymes, are tolerant of amino acid substitutions that might otherwise compromise their activity [5–8]. This tolerance exists throughout the protein, even at highly conserved active site residues. Up to 40% of all substitutions at the *Taq* Pol I active site are tolerated without any

* Corresponding author. Tel.: +1 206 543 6015.
E-mail address: laloeb@u.washington.edu (L.A. Loeb).

1568-7864/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.dnarep.2005.09.006

Table 1
Select examples of mutant Family A DNA polymerases with altered substrate specificity

Altered property	Application	Mutation (s)	Reference
Increased efficiency of ddNTP incorporation	DNA sequencing	Motif B, *Taq* Pol I	[30]
Lower fidelity of dNTP incorporation	Mutagenic PCR	Motif A, *Taq* Pol I	[25]
Lower fidelity of dNTP incorporation	In vivo mutagenesis	Motif A and B, *E. coli* Pol I	[15]
Increased efficiency of fluorescently-labeled dNTP incorporation	DNA sequencing; production of fluorescent DNA probes	Multiple, *Taq* Pol I	[41,44]

significant loss of activity; yet some of these substitutes change catalytic specificity. These findings present exciting opportunities for biochemists to modify polymerases for novel purposes without abolishing their catalytic activity.

In the redesign of DNA polymerases or in efforts to understand their structure-function, two general approaches have been extensively utilized. The first is the rational mutation of polymerases based on structure, sequence conservation, and mechanism of catalysis. This approach has been highly successful in determining the function of individual residues, but when applied to creating polymerases for a highly specialized purpose, is limited by lack of rules for predicting the effects of multiple substitutions and substitutions at a distance from the active site. The second approach falls under the rubric of applied molecular evolution, the construction of large libraries from which one obtains mutant enzymes with desired properties either by screening or selection. Methods for the creation of libraries include substitutions of portions of genes by oligonucleotides containing random substitutions [9], phage libraries [10] chemical modification [11], copying genes via error-prone PCR [12], recombining segments of different genes by shuffling [13,14] and the continuous evolution of mutant molecules in vivo [15]. Methods have even been established to encapsulate DNA polymerases and their encoding genes in lipid vesicles and to create compartmentalized self-replicating units [16]. These technologies have yielded mutant enzymes with desired properties that never could have been predicted based on rational design. However, these techniques are also limited; only a small fraction of all possible substitutions can be sampled, and many of the most efficacious mutant polymerases we have so far analyzed harbor multiple amino acid substitutions. It has become increasingly apparent that in order to create proteins with new functions, both approaches need to be combined and iterated sequentially.

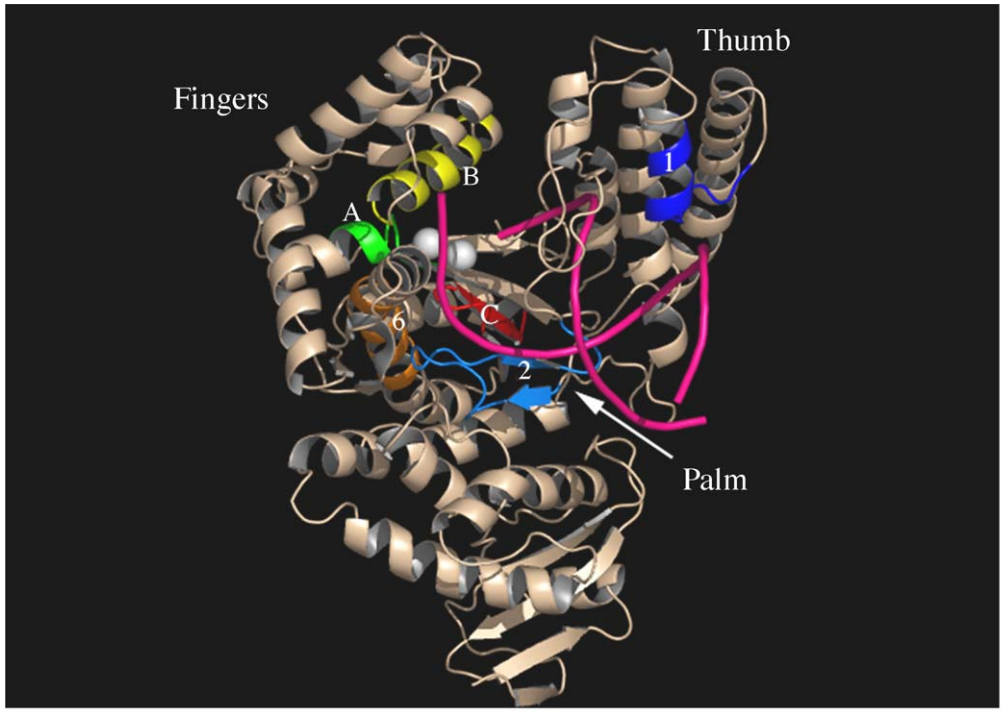
Efforts to create new DNA polymerases have focused mainly on the Family A, whose members includes *E. coli* DNA Pol I, *Taq* Pol I, and T7 DNA polymerase. These enzymes are investigated extensively because of their simplicity in subunit composition, their ease of purification, their voluminous biochemical characterization, and in some cases, their stability in extreme conditions. In this paper, we will review how this family of DNA polymerases selects its substrates and ways in which this selectivity has been successfully altered. We will begin by looking at the overall structure of the enzyme and known mechanisms that it uses for accurate replication. We then progress to discuss the regions of the polymerase that contact its substrates – the six conserved motifs found in Family A (Fig. 1) – and how each affects substrate selectivity. Finally, we consider amino acid residues outside these motifs and their role in altering the activity and fidelity of polymerases.

This manuscript is dedicated to the memory of Dale Mosbaugh, who approached science as a biochemist, a reductionist and an incisive experimenter. He projected a joyous and even childlike curiosity to discover how enzymes function both at a structural level and within a cell. His focus was on uracil glycosylase, the enzyme that removes misincorporated or spontaneously generated uracil residues from DNA, and in a sense, monitors errors by DNA polymerases. His approach was to redesign or inhibit uracil glycosylase and to probe its function in cells. He always wanted to use applied molecular evolution to change the substrate specificity of uracil glycosylase and thus create new and unique uracil glycosylases. His unanticipated departure will deprive us of the many discoveries he would have attained in pursuing this uncharted path. In this manuscript, we will focus on DNA polymerases, a companion enzyme that has been probed extensively by mutagenesis. Our focus on DNA polymerase is based on their centrality in the DNA synthetic process and the value of this enzyme for biotechnology.

2. Overall structure of DNA polymerase I

The structure of Family A DNA polymerases has been compared to a right human hand, with domains akin to the palm, thumb, and fingers (Fig. 1a) [17]. X-ray diffraction studies of three Family A DNA polymerases co-crystallized with DNA and nucleotide substrates highlight the role of each domain in catalysis [18–20]. In these snapshots, the enzyme uses its palm and thumb domains to grip the double-stranded DNA primer-template, while its fingers curl over the palm to form the binding pocket for the nascent base pair. The primer-template is oriented with the primer 3′ end near the fingers and the 5′ end near the thumb. The binding sites for the primer terminus, incoming nucleotide, and two magnesium ions required for catalysis are located on the inner face of the fingers and on surface of the palm near the fingers. Some members of the Family A DNA polymerases have additional domains that contain exonuclease activities.

These polymerase domains undergo two major structural rearrangements during catalysis [18–20]. The first is the opening of the thumb-palm cavity upon binding of the DNA substrate. This is accomplished by the movement of the polymerase thumb away from the palm. Simultaneously, the tip of the thumb moves in the opposite direction toward the palm to form contacts with the DNA minor groove. The second conformational change is the rotation of the fingers toward the palm concomitant with



(a)

(b)

	Motif 1	Motif 2
E.coli PolI	578 FNLSSTKQLQTIL 590	660 HQAVTATGRLSSTDPNLQ 677
Taq PolI	482 FNLNSRDQLERVL 494	565 NQTATATGRLSSSDPNLQ 582
Bst PolI	526 FNINSPKQLGTVL 538	607 NQALTQTGRLSSVEPNLQ 624
T7 Pol	334 FNPSSRDH IQKKL 346	421 NPNGAVTGRATHAFPNLAQ 439

	Motif A	Motif B
E.coli PolI	702 VSA**D**YSQI**E**LR 712	754 **R**RSA**K**A INF GL IY 766
Taq PolI	607 VAL**D**YSQI**E**LR 617	659 **R**RAA**K**T INF GY LY 671
Bst PolI	650 FAA**D**YSQI**E**LR 660	701 **R**RQA**K**AVNF G IV Y 713
T7 Pol	472 AGI**D**ASGL**E**LR 482	518 **R**DNA**K**T F IYG FL Y 530

	Motif 6	Motif C
E.coli PolI	843 AINA PMQGTAA D IIK 857	877 IMQV**HD E**LVFE 887
Taq PolI	748 AFNM PVQ G TAA DLMK 762	780 LLQV**HD E**LVLE 790
Bst PolI	791 AMNT PIQ G SAA D IIK 805	825 LLQV**HD E**L ILE 835
T7 Pol	609 ALNTLLQ SA GA L ICK 623	649 MAW V**HD E**IQVG 659

Fig. 1. Structure of DNA polymerase I. (a) Crystal structure of *Taq* DNA polymerase I in closed conformation. Palm, thumb, fingers and exonuclease domains are labeled. Motifs 1 (purple), 2 (blue), A (green), B (yellow), 6 (orange), and C (red) are highlighted. Primer-template duplex DNA is colored pink. The two catalytic magnesium ions bound to the palm are colored white. Figure was generated from PDB coordinates 3KTQ [19] using MacPymol. (b) Evolutionarily conserved motifs of DNA polymerase Family A: motifs 1, 2, A, B, 6, and C from E. *coli*, *Taq*, *Bst*, and T7 are aligned according to amino acid conservation. Numbers bracketing each sequence denote the first and last residue. Bolded letters are evolutionarily conserved residues throughout all members of Family A. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

nucleotide binding. This transitions – from the "open" to the "closed" form of the enzyme – restricts the size of the active site, forms important contacts between the fingers and the nucleotide, and helps to establish the required geometry for substrate recognition and incorporation.

Extensive work on Family A DNA polymerases over the past 30 years have carved the path to understanding how DNA polymerases synthesize DNA with exceptionally high accuracy [21]. Two types of enzyme-to-substrate interaction have been postulated to account for this selectivity. The first are bonding interactions formed between residues of the active site and the substrates. These interactions consist of direct ionic bonds, hydrogen bonds, or magnesium-mediated bonds between residues and charged groups on the DNA duplex and incoming deoxynucleotide. These interactions not only position the substrates correctly but also select the correct and exclude incorrect nucleotide substrates at the active site. The second type of interaction for establishing substrate selectivity is non-bonded and relies on steric repulsive forces that prevent molecules from clashing [22]. The active site has a precise geometry, which allows complementary basepairs to fit but disfavors non-complementary basepairs that typically have incorrect size and shapes. This steric hindrance in some instances accentuates the direct bonding mechanism, by sandwiching substrates in an ori-

entation in which their charged groups are forced into contact with charged residues on the polymerase.

The sum of all these interactions, bonding and non-bonding, contributes to three checkpoints that determine replication fidelity [21]. The first of these is the preferential binding of the complementary nucleotide to the polymerase—with a 10- to 100-fold greater affinity than non-complementary nucleotides. The second checkpoint for selection occurs when the enzyme transitions from the open to the closed conformation. The rate for this transition is up to 10,000 times faster for the correct nucleotide than for an incorrect one. The closed conformation is induced by the binding of a nucleotide and is stabilized by interactions between the polymerase and correctly paired bases. The final checkpoint for correct nucleotide incorporation occurs at the chemical step. The rate for catalysis slows for incorrect nucleotides because the bonds that stabilize the transition state are perturbed from their ideal lengths and angles when a non-complementary nucleotide occupies the active site.

To copy a DNA template correctly, polymerases must also avoid inserting or deleting bases. Several models for how frameshifts arise predict that misalignment of the primer and template strands during DNA synthesis creates unpaired bases that are eventually fixated as insertions or deletions [21]. Because these strand misalignments are more likely to occur when DNA is unbound, a polymerase that dissociates frequently from the DNA also has lower frameshift fidelity. Hence, the binding affinity of a polymerase to its DNA substrates is important for not only its processivity but also its accuracy.

Family A DNA polymerases share six motifs that are evolutionarily conserved (Fig. 1b). These motifs contact the substrates and form the active site (Table 2). Motifs A, B, and C are the most highly conserved. In fact, motifs A and C exist in all known DNA and RNA polymerases. Motifs 1, 2 and 6 are conserved structurally but vary more in amino acid sequence, suggesting that these regions may have fewer bonding interactions with the substrates.

2.1. Motif A

Located in the palm domain, motif A is one of two conserved motifs present in all DNA and RNA polymerases. Residues in motif A make contact with the primer strand bases, sugar-phosphate backbone, and the catalytic magnesium ions [18–20]. The binding of the magnesium ions by Asp705 (*E. coli*) is critical for catalysis. All known mutations of this residue abolish polymerase activity. In an analysis of 300 active mutants of *Taq* DNA polymerase I constructed with random substitutions in motif A, the only amino acid residue that was not substitutable was this aspartic acid [7].

The other highly conserved residue within motif A is Glu710 (*E. coli*), which appears to participate in the stabilization of the closed form of the polymerase. Joyce and co-workers, however, found that this residue can be replaced without abolishing polymerase activity [23]. One novel property obtained from the mutation of Glu710 is the ability to incorporate ribonucleotides. The E710A mutant can incorporate rNTPs up to 1000-fold more efficiently than wildtype, but this substitution decreases dNTP incorporation by as much as 40-fold. Based on crystallographic evidence, the hydrophobic chain of Glu710 forms half of the pocket surrounding the 2′ position of the deoxyribose ring. This hydrophobic pocket occludes ribonucleotides from the active site the polymerase. It has also been suggested that the Glu710 could also bind to a magnesium ion and serve to occlude the 2′ hydroxy group of the ribose ring. However, structural evidence for this hypothesis needs to be established.

Our group has also examined residues in motif A that contributed to rNTP exclusion by substituting random oligonucleotides for the nucleotides that encode motif A and studied the incorporation of ribonucleotides by mutant DNA polymerases [24]. We created over 100,000 variants of *Taq* polymerase by randomizing the sequence within motif A, selected active clones from this library using genetic complementation of a *polA*$^{-}$ *E. coli* strain, and then tested 300 of these for their ability to incorporate ribonucleotides. Selected substitutions at both Glu615 and Ile614 (equivalent to Glu710 and Ile709 in *E. coli* Pol I) increased the rate of rNTP incorporation. Interestingly, several mutations at Ile614, as well as others in motif A, act as mutator alleles—they reduce the base substitution fidelity of the enzyme by as much as 300-fold [5,15,25]. How do these two changes coincide? One possibility is that substitutions at Ile614 increase the enzyme's preference for specific non-Watson–Crick basepairs. Alternatively, these substitutions are likely to widen the active site and relax the substrate specificity of the enzyme. Ile614 makes no direct contacts with the incoming nucleotide, and its mutational effects are likely to be indirect, either by

Table 2
Conserved motifs of Family A DNA polymerases

Motif	Conservation[a]	Contacts to substrate	Select mutant(s) and property	Reference
A	Family A, B, X, Y, and reverse transcriptase	DNA primer and Mg^{2+}	*E. coli* E710A: increased ribonucleotide incorporation	[23]
B	Family A, B, and X	Incoming nucleotide and template base	*Taq* F667Y: increased dideoxynucleotide incorporation	[30]
C	Family A, B, X, Y, and reverse transcriptase	DNA primer terminus and Mg^{2+}	*E. coli* H881A: increased replication fidelity	[35]
1	Family A	DNA template and primer backbone	*E. coli* deletion of residues 590-613: decreased DNA binding, processivity, and frameshift fidelity	[36]
2	Family A	DNA minor groove and template backbone	*E. coli* R668A: decreased DNA binding	[35,37,38]
6	Family A	DNA minor groove and template backbone	*E. coli* Q849A: decreased DNA binding	[35,37,38]

[a] See Refs. [45–48].

distorting the size or shape of the active site or by altering the geometry of its neighboring residues. Ile614 mutations may expand the active site pocket, thereby increasing the repertoire of tolerable nucleotide substrates. The hypothesis of widening the active site is supported by the crystal structures of several error-prone DNA Family Y polymerases, each containing a larger active site pockets and incorporating non-complementary nucleotides at high frequency [26–28]. The ability of Y family polymerases to incorporate ribonucleotides has not yet been reported.

The creation of an RNA polymerase from DNA polymerase by Joyce and co-workers was only partially successful because the enzymes identified still maintained a preference for incorporating dNTPs over rNTPs. Romesberg and coworkers went one step further by using applied molecular evolution to select for a more efficient RNA polymerase [10]. They constructed a phage-conjugated library of 10^7 *Taq* variants by randomizing the sequence within specific motifs and then used four rounds of phage display selection to recover those variants that incorporate rNTPs. They identified at least 3 *Taq* enzymes with greater rNTP than dNTP incorporation (as much as a 10,000-fold increase in rNTP incorporation) and catalytic efficiencies that were comparable to the wildtype enzyme. These mutants contained 2, 5, and 8 amino acid substitutions. This finding suggests that multiple mutations in DNA polymerase may be necessary for conversion into an efficient RNA polymerase. Interestingly, all the mutations in these new enzymes mapped to either the Motif A or its vicinity, despite the presence of mutations in Motifs B and C in the library. One could speculate that Motif A is perhaps a more important determinant of ribonucleotide selection than the other motifs, or that mutations in Motif A are better tolerated in terms of activity.

Substitutions in Motif A of *Taq* also allow for the incorporation of nucleotides with bulky modifications at the 2′ position of the sugar. The desire to create these specialized polymerases stems from the need in industry for the synthesis of polymers with novel functionality. To this end, the 2′ nucleotide position of DNA serves as a convenient location for tethering desired compounds. Romesberg and co-workers attempted to evolve polymerases that could incorporate nucleotides containing a 2′ ether using phage display [29]. Although they randomized residues in several regions across the gene, the mutants they selected contained only motif A mutations. These mutations mapped to Ile614 and Glu615. Although further work is needed to compare the activity and fidelity of these mutants to wildtype *Taq*, motif A appears to be a strong determinant of selectivity at the 2′ position.

Attempts to change DNA polymerase substrate specificity by mutating motif A residues have been reasonably successful, especially in the realm of ribonucleotide incorporation. Mutations at residues Ile709 and Glu710 (*E. coli*) have demonstrated the greatest potency. These successes, however, have introduced additional questions as to how these residues determine fidelity. Ile709, in particular, contacts no substrates directly, and thus offers few clues to its mechanism of action. It remains to be seen whether each of these mutations expand the active site, as we have hypothesized.

2.2. Motif B

The residues of motif B, located in the fingers domain, form the O-helix in *Taq* Pol I. This helix contacts the nascent basepair during polymerization and hydrogen bonds with the triphosphate of the incoming nucleotide. Four residues are evolutionarily conserved in this motif: Arg659, Lys663, Phe667, and Tyr671, all of which locate to the helical face that contacts the newly formed basepair. To determine which residues can be replaced without compromising polymerase activity, Suzuki et al. randomized the motif B sequence to generate a library containing 10^4 variants and selected for active variants by complementation to a *polA*$^-$ *E. coli* strain [8]. Their results revealed that only Arg659 and Lys663 are essential for activity, which is in concordance with crystallographic data demonstrating their binding to the phosphates of the incoming nucleotide.

All other motif B residues tolerate substitutions without loss of genetic complementability, but substitutions in these positions alter the enzyme's substrate specificity. The most drastic example is observed at one residue, Phe667 in *Taq* and Phe762 in *E coli* Pol I, that discriminates against dideoxynucleotides during DNA synthesis. In both polymerases, mutation of the phenylalanine to a tyrosine eliminates their ability to discriminate against dideoxynucleotides and allows them to incorporate these analogues preferentially over natural deoxynucleotides [30]. This result is underscored by the observation in T7 DNA polymerase where mutation of the native residue Tyr526 to a phenyalanine produces the opposite result—it increases the enzymes discrimination against dideoxynucleotides by over 1000-fold. This bi-directional reversal of substrate specificity is one of the most elegant demonstrations of the effect of a single residue in DNA polymerases. Tabor and Richarson hypothesized that because the residue and the 2′ hydroxy group of the incoming nucleotide are in the same vicinity, they may coordinate a catalytic metal ion or water during catalysis. Moreover, loss of either the 2′ hydroxy or the tyrosine hydroxy could be compensated by the gain of the other. This hypothesis is not in accord with the crystal structure, which shows the two chemical substituents neighboring each other, but fails to show that either one coordinates a metal or water [18]. Nevertheless, another possible mechanism – that of electronegativity compensation – could also explain these results. The enhancement in the incorporation of dideoxynucleotides have been instrumental in developing methods for rapid sequencing of DNA. Today, the F667Y *Taq* polymerase is extensively used in DNA Sanger sequencing.

The other highly conserved but mutable residue in motif B is Tyr671 (*Taq*) [8]. This residue, with its aromatic ring, stacks with the newly formed basepair and stabilizes it. Mutation of the corresponding Tyr766 in *E. coli* Pol I increases the enzymes ability to incorporate ribonucleotides, but decreases the fidelity of DNA replication [23,31]. This increase in ribonucleotide incorporation is, however, exclusive to purines, suggesting that base structure may also contribute to sugar discrimination in the polymerase.

As in the case of motif A, motif B contains residues that do not contact the substrate but can nevertheless alter its specificity. Perturbations of the active site, either sterically or geometrically,

is sometimes invoked to explain changes in substrate specificity [25]. However, in the case of motif B, an additional mechanism is offered by Suzuki and co-workers. They suggest that one such mutant, *Taq* A661E, with increased rate of rNTP incorporation achieves this phenotype by stabilizing the enzyme in the closed conformation through intra-protein interactions [32]. Movement of motif B is a major component of the open to closed transition, and if this motif can be stabilized in the closed form, the enzyme may have a longer time to incorporate substrates, even those that are normally disfavored. Results with *E. coli* Pol I also suggest that separate mechanisms to discriminate bases exists at motifs A and B. When I709N and A759R error-prone alleles from motifs A and B are combined, they produce synergistic rather than additive increases in mutation frequency [15]. One would expect that if these alleles function through the same mechanism, their combination would yield only additive results at best. The biochemical verification of this hypothesis remains to be performed, but stabilizing the closed form of the polymerase may represent a powerful strategy for decreasing the ability of the polymerase to discriminate between different substrates.

Amino acid residues in motif B may also confer nucleotide discrimination through bonds with select bases. In the case of *Taq* polymerase, Waksman and co-workers contend that the enzyme is uniquely bonded to ddGTP through Arg660, which increases its incorporation relative to the three other dideoxynucleotides [33]. An R660D mutant of this polymerase displayed decreased ddGTP incorporation and produced sequencing reactions where ddGTP terminations do not predominate. So far the crystal structure has failed to provide a mechanism to account for this base-specific enhancement of incorporation.

The dramatic changes seen in motif B mutants highlight its importance in determining substrate specificity. The most prominant of these changes, seen in the switch in preference from deoxynucleotides to dideoxynucleotides illustrates how one residue alone can impact selectivity. Most determinants of substrate specificity, however, are more subtle. Finally, residues that stabilize a catalytically competent form of the enzyme may widen the substrate specificity of an enzyme by allowing more time for chemical steps involving disfavored reactions.

2.3. Motif C

Motif C is located in the palm domain, and like motif A, is conserved in all DNA and RNA polymerases; it contains the second aspartic residue (Asp882 in *E. coli*) that coordinates the catalytic magnesium ions and is crucial for polymerase activity [18–20,34]. This motif also contains the evolutionary conserved Val880 and His881, which interact with the nucleotide sugar of the primer terminus, and Glu883, which coordinates with Asp882 via a water molecule.

Several residues in motif C have been assayed for their role in determining DNA replication fidelity, and of these, one of the more interesting is His881. The H881A mutant demonstrates a base substitution antimutator effect when assayed on a gapped plasmid [35]. This effect is due to a decreased ability to extend mismatches rather than to changes in the ability to misinsert bases. Such a phenotype is not unique to this mutant, but it illustrates an important concept: the ability of a DNA polymerase to utilize a nucleotide depends on its efficiency at both the insertion and extension steps of polymerization. The efficiencies for these processes are sometimes disconnected, but both must be considered when creating a successful polymerase.

It appears that motif C plays a role in both the insertion and extension steps of polymerization, but the specific contribution of many motif C residues to replication fidelity remains largely unexplored, as no extensive mutagenesis analysis has been reported, compared to motifs A and B. This region provides a rich target for future investigations into polymerase specificity determinants.

2.4. Motif 1

Motif 1 consists of a helix and loop located at the tip of the thumb domain, where it interacts with the sugar-phosphate backbone of the DNA template and primer, four to seven base-pairs upstream from the active site [18–20]. Of the conserved regions, motif 1 residues are the most distant from the active site. Relatively few mutations in this motif have been studied in detail. Kunkel and co-workers deleted the entire motif, including residues 590 to 613, in *E. coli* Pol I, and observed a 100-fold increase in K_D to DNA, a decrease in processivity, and a decrease in frameshift fidelity when the mutant was assayed on homopolymeric runs of bases [36]. No large changes in the dNTP binding or catalytic rate occurred. Because such a large region was deleted, it is difficult to conclude which residues are responsible for the phenotype. Later work showed that N579A and S582A mutants actually had slightly increased frameshift fidelity [35]. Because individual mutations in motif 1 have only slight effects on frameshift fidelity, the authors of this later study suggest that there may be some functional redundancy in the DNA binding ability of these residues. Further mutational analysis of this region will be necessary to resolves this issue and to determine the role of each residue in greater detail. Nevertheless, motif 1 appears to be promising target if one wanted to change the binding affinity of the polymerase to the DNA backbone.

2.5. Motif 2

Motif 2 consists of two beta strands located in the palm domain. Residues in this domain interact with the DNA minor groove and template sugar-phosphate backbone [18–20]. One of the most important residues in this motif is Arg668 (*E. coli*), which binds to the primer and template terminal bases, through the N-3 position of purines and the O-3 of pyrimidines within the minor groove. Mutation of this residue decreases catalytic activity by up to 300-fold and increases K_D to DNA by up to 23-fold [34,37,38]. Arg668 may especially be important for allowing the enzyme to extend mispairs and bypass lesions such as abasic sites and oxidative damage; Arg668 mutants lose these abilities and therefore appear more accurate when tested on templates containing base adducts [39,40]. When assayed on undamaged DNA substrates, however, the R668A mutant exhibits a two-fold decrease in base substitution fidelity [35].

Other residues in motif 2 bind to the template sugar-phosphate backbone have also been analyzed. N675A and N678A single mutants do not have significantly altered frameshift fidelity [35]. These results, like those seen in motif 1, suggest redundancy in the sugar-phosphate binding ability of the polymerase and imply that in order to create polymerases that bind strongly to DNA with modified sugar-phosphate backbones, multiple mutations may be required. Rational designs of polymerases in the future should consider the contacts that motif 2 makes with the minor groove and template backbone.

2.6. Motif 6

Motif 6 consists of a helix running parallel to the DNA template strand in the palm, at the base of the fingers. It too participates in binding the DNA minor groove. Gln849 in this motif is especially important because it hydrogen bonds to the first base of the template strand [18–20]. Loss of Gln849 decreases DNA binding by up to 40-fold while virtually preserving dNTP binding [34,37,38]. Fidelity assays based on in vitro synthesis across a gapped plasmid show increases in replication accuracy of three- to 10-fold for the Q849A mutant, depending upon the sequence that is tested [35]. Surprisingly, this result contrasts those seen with the R668A mutant, which is also a minor grove binding residue, located in motif 2. Why these two minor groove bonding residues produce opposite phenotypes when mutated will be an interesting issue to pursue in the future.

2.7. Residues outside the conserved motifs

The evolutionary conserved motifs represent structural elements of the protein where several residues involved in substrate binding, conformational change, or catalysis are clustered. However, residues that are outside of these clusters also play significant roles in enzyme substrate specificity. Kunkel and co-workers showed that mutations outside the conserved motifs can have up to a two-fold effect on altering replication fidelity [35]. To be sure, the mutated residues are often evolutionarily conserved, but these findings nevertheless point to regions that are several angstroms from the active site as potential targets for enzyme redesign. Mutations at these residues may produce changes in volume and shape that are transmitted to the active site due to neighbor-to-neighbor amino acid interactions. A larger amino acid substitution, for example, could push neighboring residue into the active site, whereas, a smaller amino acid substitution could create a cavity that could draw other residues away from the active site. Crystal structures of these mutants may reveal whether this concept is indeed correct.

Other examples of the importance of mutations in non-conserved regions appears in a study by Holliger and co-workers, who evolved *Taq* DNA polymerase to become efficient at extending mismatches [41]. These enzymes have an increased ability to extend nucleotides labeled with biotin, fluorescein, or bulky dyes. One of their mutants has substitutions that map to only areas outside the conserved motifs, and only one of these mutations, N583S, makes a contact with a substrate, which is a primer base. The other mutant they report has one mutation in motif A, but five other mutations outside of conserved motives. That these mutants have eight and six substitutions, respectively, suggests that multiple mutations may be necessary for the extension assays they have tested. Why would multiple mutations be necessary to confer a desired trait? One possibility is that mutations that do not contact substrates directly have minute effects on the enzyme's selectivity; in order for these mutations to have a considerable effect, many of them must be present and they must be additive.

An alternative reason why multiple mutations may be required for drastic changes in substrate specificity is that the majority of these serve a compensatory role to balance the detrimental effect of one mutation. In this scenario, one mutation may be sufficient to alter the substrate specificity, but it is also detrimental to the stability or activity of the enzyme. Other mutations would then be necessary to compensate for this detrimental effect, such as by stabilizing the folding of the protein. It will be interesting to verify these hypotheses as novel polymerases with multiple mutations are dissected and studied further.

Mutations outside conserved motifs often have no contact with substrates. In such cases, the effects of the mutations may be indirect and mediated through mechanisms previously mentioned. These include changes in steric exclusion of the active site, distortion of the reaction geometry, and retention of the polymerase in a catalytically competent conformation. Multiple mutations and mutations distributed outside the active site may be necessary to compensate for other mutations that destabilize the polymerase.

3. Conclusion

The structure of DNA polymerase can be divided into domains and motifs with functions assigned to each based on crystallographic information. Experimental data indicate that multiple regions of polymerase often influence the properties of the enzyme. Regions of the polymerase that make unmistakable contacts with substrates are easily designated with a function, and in a few cases, these functions have borne out through experimentation. However, much of the enzyme does not make clear contacts with substrates, or in some cases, these contacts contradict experimental data. Moreover, the crystal structures of polymerases are clouded by uncertainty as to which conformations are the most important for activity. Recent structures showing catalytically competent crystals of *Bst* DNA polymerase I has begun to verify the biological significance of currently available structures [42,43].

While the rational approach to redesigning polymerases has worked in a few instances, and only when residue-to-substrate contacts are clear, the popular strategy for moving ahead with future polymerase redesign has been through molecular evolution. This approach has produced a wealth of polymerase mutants that need to be fully characterized. Many of these mutants have only mutations that are outside of the conserved motifs. What new mechanisms can we uncover about how these mutations function?

While these studies have described the involvement of individual amino acids in catalysis they have not, so far, presented

a comprehensive structure of how DNA polymerases change at each nucleotide-binding event. Is base-selection primarily governed by the geometry at the active site or by kinetic differences in the incorporation of complementary and non-complementary nucleotides? If DNA polymerases are dynamic structures that change conformation throughout the molecule at each catalytic step, do we have the tools to define a dynamic conformation? Can one label DNA polymerases with fluorescent analogs and measure changes in conformation throughout the molecule as a function of nucleotide additions? The dream of many "polymerologists" is to have a movie that portrays conformation as a function of nucleotide additions. Static structures even those involving ternary complexes of polymerase, template-primer and deoxynucleotide substrates may not be sufficient to establish the mechanism of base-selection which is central to the fidelity of DNA replication.

Acknowledgements

We would like to thank Manel Camps and Nate Elliott for their critical reading of this manuscript and Cole Bower for help in preparing the manuscript. Work in the authors' laboratory was supported by NIH grants, CA102029 and CA78885, and the Poncin Scholarship Fund.

References

[1] R.E. Goldsby, N.A. Lawrence, L.E. Hays, E.A. Olmsted, X. Chen, M. Singh, B.D. Preston, Defective DNA polymerase-6 proofreading causes cancer susceptibility in mice, Nat. Med. 7 (2001) 638–639.
[2] M.J. Longley, M.A. Graziewicz, R.J. Bienstock, W.C. Copeland, Consequences of mutations in human DNA polymerase gamma, Gene 354 (2005) 125–131.
[3] A. Trifunovic, A. Wredenberg, M. Falkenberg, J.N. Spelbrink, A.T. Rovio, C.E. Bruder, Y.M. Bohlooly, S. Gidlof, A. Oldfors, R. Wibom, J. Tornell, H.T. Jacobs, N.G. Larsson, Premature ageing in mice expressing defective mitochondrial DNA polymerase, Nature 429 (2004) 417–423.
[4] J. Micklefield, Backbone modification of nucleic acids: synthesis, structure and therapeutic applications, Curr. Med. Chem. 8 (2001) 1157–1179.
[5] A. Shinkai, P.H. Patel, L.A. Loeb, The conserved active site motif A of *Escherichia coli* DNA polymerase I is highly mutable, J. Biol. Chem. 276 (2001) 18836–18842.
[6] H.H. Guo, J. Choe, L.A. Loeb, Protein tolerance to random amino acid change, Proc. Natl. Acad. Sci. U.S.A. 101 (2004) 9205–9210.
[7] P.H. Patel, L.A. Loeb, DNA polymerase active site is highly mutable: evolutionary consequences, Proc. Natl. Acad. Sci. U.S.A. 97 (2000) 5095–5100.
[8] M. Suzuki, D. Baskin, L.E. Hood, L.A. Loeb, Random mutagenesis of *Thermus aquaticus* DNA polymerase I: concordance of immutable sites in vivo with the crystal structure, Proc. Natl. Acad. Sci. U.S.A. 93 (1996) 9670–9675.
[9] M.S. Horwitz, L.A. Loeb, Promoters selected from random DNA sequences, Proc. Natl. Acad. Sci. U.S.A. 83 (1986) 7405–7409.
[10] G. Xia, L. Chen, T. Sera, M. Fa, P.G. Schultz, F.E. Romesberg, Directed evolution of novel polymeerase activities: mutation of a DNA polymerase into an efficient RNA polymerase, Proc. Natl. Acad. Sci. U.S.A. 99 (2002) 6597–6602.
[11] J.B. Sweasy, L.A. Loeb, Mammalian DNA polymerase *p* can substitute for DNA polymerase I during DNA replication in *Escherichia coli*, J. Biol. Chem. 267 (1992) 1407–1410.
[12] R.C. Cadwell, G.F. Joyce, Randomization of Genes by PCR Mutagenesis, vol. 2, Cold Spring Harbor Laboratory, 1992, 28–33.
[13] W.P.C. Stemmer, Rapid evolution of a protein in vitro by DNA shuffling, Nature 370 (1994) 389–391.
[14] F.H. Arnold, J.C. Moore, Optimizing industrial enzymes by directed evolution, Adv. Biochem. Eng. Biotechnol. 58 (1997) 1–14.
[15] M. Camps, J. Naukkarinen, B.P. Johnson, L.A. Loeb, Targeted gene evolution in *Escherichia coli* using a highly error-prone DNA polymerase I, Proc. Natl. Acad. Sci. U.S.A. 100 (2003) 9727–9732.
[16] F.J. Ghadessy, J.L. Ong, P. Holliger, Directed evolution of polymerase function by compartmentalized self-replication, Proc. Natl. Acad. Sci. U.S.A. 98 (2001) 4552–4557.
[17] D.L. Ollis, P. Brick, R. Hamlin, N.G. Xuong, T.A. Steitz, Structure of large fragment of *Escherichia coli* DNA polymerase I complexed with dTMP, Nature 313 (1985) 762–766.
[18] S. Doublie, S. Tabor, A.M. Long, C.C. Richardson, T. Ellenberger, Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 A resolution, Nature 391 (1998) 251–258.
[19] Y. Li, S. Korolev, 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. 17 (1998) 7514–7525.
[20] J.R. Kiefer, C. Mao, J.C. Braman, L.S. Beese, Visualizing DNA replication in a catalytically active Bacillus DNA polymerase crystal, Nature 391 (1998) 304–307.
[21] T.A. Kunkel, K. Bebenek, DNA replication fidelity, Annu. Rev. Biochem. 69 (2000) 497–529.
[22] E.T. Kool, Active site tightness and substrate fit in DNA replication, Annu. Rev. Biochem. 71 (2002) 191–219.
[23] M. Astatke, K. Ng, N.D.F. Grindley, C.M. Joyce, A single side chain prevents *Escherichia coli* DNA polymerase I (Klenow fragment) from incorporating ribonucleotides, Proc. Natl. Acad. Sci. U.S.A. 95 (1998) 3402–3407.
[24] P.H. Patel, L.A. Loeb, Multiple amino acid substitutions allow DNA polymerises to synthesize RNA, J. Biol. Chem. 275 (2000) 40266–40272.
[25] P.J. Patel, H. Kawate, E. Adman, M. Ashbach, L.A. Loeb, A single highly mutable catalytic site amino acid is critical for DNA polymerase fidelity, J. Biol. Chem. 276 (2001) 5044–5051.
[26] H. Ling, F. Boudsocq, R. Woodgate, W. Yang, Crystal structure of a Y-family DNA polymerse in action: a mechanism for error-prone and lesion-bypass replication, Cell 107 (2001) 91–102.
[27] J. Trincao, R.E. Johnson, C.R. Escalante, S. Prakash, L. Prakash, A.K. Aggarwal, Structure of the catalytic core of *S. cerevisiae* DNA polymerase eta: implications for translesion DNA synthesis, Mol. Cell. 8 (2001) 417–426.
[28] B.L. Zhou, J.D. Pata, T.A. Steitz, Crystal structure of a DinB lesion bypass DNA polymerase catalytic fraagment reveals a classic polymerase catalytic domain, Mol. Cell. 2 (2001) 427–437.
[29] M. Fa, A. Radeghieri, A.A. Henry, F.E. Romesberg, Expanding the substrate repertoire of a DNA polymerase by directed evolution, J. Am. Chem. Soc. 126 (2004) 1748–1754.
[30] S. Tabor, 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. 92 (1995) 6339–6343.
[31] S.S. Carroll, M. Cowart, S.J. Benkovic, A mutant of DNA polymerase I (Klenow fragment) with reduced fidelity, Biochemistry 30 (1991) 804–813.
[32] M. Ogawa, A. Tosaka, Y. Ito, S. Yoshida, M. Suzuki, Enhanced ribonucleotide incorporation by an O-helix mutant of Thermus aquaticus DNA polymerase I, Mutat. Res. 485 (2001) 197–207.
[33] Y. Li, V. Mitaxov, G. Waksman, Structure-based design of *Taq* DNA polymerases with improved properties of dideoxynucleotide incorporation, Proc. Natl. Acad. Sci. U.S.A. 96 (1999) 9491–9496.
[34] A.H. Polesky, T.A. Steitz, N.D. Grindley, C.M. Joyce, Identification of residues critical for the polymerase activity of the Klenow fragment of DNA polymerase I from *Escherichia coli*, J. Biol. Chem. 265 (1990) 14579–14591.

[35] D.T. Minnick, K. Bebenek, W.P. Osheroff, R.M.J. Turner, M. Astatke, L. Liu, T.A. Kunkel, C.M. Joyce, Side chains that influence fidelity at the polymerase active site of *Escherichia coli* DNA polymerase I (Klenow fragment), J. Biol. Chem. 274 (1999) 3067–3075.
[36] D.T. Minnick, M. Astatke, C.M. Joyce, T.A. Kunkel, A thumb subdomain mutant of the large fragment of *Escherichia coli* DNA polymerase I with reduced DNA binding affinity, processivity, and frameshift fidelity, J. Biol. Chem. 271 (1996) 24954–24961.
[37] K. Singh, M.J. Modak, Presence of 18-A long hydrogen bond track in the active site of *Escherichia coli* DNA polymerase I (Klenow fragment). Its requirement in the stabilization of enzyme-template-primer complex, J. Biol. Chem. 278 (2003) 11289–11302.
[38] T.E. Spratt, Identification of hydrogen bonds between *Escherichia coli* DNA polymerase I (Klenow fragment) and the minor groove of DNA by amino acid substitution of the polymerase and atomic substitution of the DNA, Biochemistry 40 (2001) 2647–2652.
[39] M.D. McCain, A.S. Meyer, S.S. Schultz, A. Glekas, T.E. Spratt, Fidelity of mispair formation and mispair extension is dependent on the interaction between the minor groove of the primer terminus and Arg668 of DNA polymerase I of *Escherichia coli*, Biochemistry 44 (2005) 5647–5659.
[40] E.E. Gestl, K.A. Eckert, Loss of DNA minor groove interactions by exonuclease-deficient Klenow polymerase inhibits O6-methylguanine and abasic site translesion synthesis, Biochemistry 44 (2005) 7059–7068.
[41] F.J. Ghadessy, N. Ramsay, F. Boudsocq, D. Loakes, A. Brown, S. Iwai, A. Vaisman, R. Woodgate, P. Holliger, Generic expansion of the substrate spectrum of a DNA polymerase by directed evolution, Nat. Biotechnol. 22 (2004) 755–759.
[42] S.J. Johnson, J.S. Taylor, 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. 100 (2003) 3895–3900.
[43] S.J. Johnson, L.S. Beese, Structures of mismatch replication errors observed in a DNA polymerase, Cell 116 (2004) 803–816.
[44] J.P. Anderson, B. Angerer, L.A. Loeb, Incorporation of reporter-labeled nucleotides by DNA polymerases, Biotechniques 38 (2005) 257–264.
[45] M. Delarue, O. Poch, N. Tordo, D. Moras, P. Argos, An attempt to unify the structure of polymerases, Protein Eng. 3 (1990) 461–467.
[46] P.H. Patel, M. Suzuki, E. Adman, A. Shinkai, L.A. Loeb, Prokaryotic DNA polymerase I: evolution, structure, and "base flipping" mechanism for nucleotide selection, J. Mol. Biol. 308 (2001) 823–837.
[47] C.M. Joyce, T.A. Steitz, Function and structure relationships in DNA polymerases, Annu. Rev. Biochem. 63 (1994) 777–822.
[48] C.M. Kondratick, M.T. Washington, S. Prakash, L. Prakash, Acidic residues critical for the activity and biological function of yeast DNA polymerase eta, Mol. Cell Biol. 21 (2001) 2018–2025.