Getting a grip on how DNA polymerases function

Premal H. Patel and Lawrence A. Loeb

DNA polymerases play a central role in maintenance of genetic information, and the structures of polymerases in complex with DNA and dNTP provide valuable insights into mechanisms utilized by DNA polymerases to achieve high fidelity. Comparison of several high resolution complexes of DNA polymerases bound with relevant substrates indicates that the two major families of polymerases function by very similar mechanisms.

Faithful replication of DNA from one generation to the next is crucial for long term species survival. Genomic integrity in prokaryotes, archaea and eukaryotes is dependent on efficient and accurate catalysis by multiple DNA polymerases. Following the discovery of DNA polymerase I (Pol I) of *Escherichia coli* by Kornberg and colleagues in 1957 (ref. 1), multiple DNA polymerases have been identified in prokaryotes and eukaryotes, including the recent discovery of several error-prone DNA polymerases[2]. Based on primary sequence similarity, these DNA polymerases can be categorized into families (Table 1). The most extensively studied polymerases include those in family A (found in prokaryotes, eukaryotes and bacteriophages) and those in family B (found in prokaryotes, eukaryotes, archaea and viruses). Collectively, polymerases from both families exhibit considerable sequence diversity and function during replication, recombination and repair. In contrast to this diversity of biological functions, increasing evidence indicates that many DNA polymerases exhibit an identical mechanism for DNA synthesis.

Table 1 Representative members of families of DNA polymerases

Family	Prokaryotic	Eukaryotic	Archaea	Viral
A	Pol I	Pol γ,θ		T3,T5,T7 pol
B	Pol II	Pol α,δ,ε,ζ	Pol BI, BII	Adenovirus, HSV, RB69, T4,T6 pol
C	Pol III(α)			
D			Pol D	
X		Pol β,λ,μ, Tdt		
RT		Telomerase		Reverse transcriptases
UmuC/DinB	Pol IV,V	Pol η, ι, κ		

Detailed kinetic studies have provided powerful tools to probe the catalytic mechanism of DNA polymerization[3] and the fidelity of DNA synthesis[4]. Nucleotide incorporation by DNA polymerases occurs in discrete steps: binding of the DNA template–primer, followed by binding to an incoming dNTP, phosphodiester bond formation, release of the pyrophosphate, and translocation to the next 3′ OH primer terminus. In this process, nucleotide selection within the active site of DNA polymerases is the major contributor to the fidelity of DNA synthesis[5]. Nucleotide selection encompasses correct Watson-Crick base pair formation and conformational changes within the protein and DNA during each nucleotide addition step. Together, these processes result in a 100,000-fold more frequent incorporation of the complementary nucleotide relative to the noncomplementary nucleotides. Inefficient extension of an incorporated terminal noncomplementary nucleotide allows time for removal by 3′-5′ exonuclease. The exonucleolytic (3′-5′) proofreading domain is an integral part of some DNA polymerases and contributes, on average, 10-fold to the overall mutation rate, although in some polymerases this contribution may be much greater.

Beginning in 1985 with the work by Steitz and coworkers on *E. coli* DNA pol I[6], each of the kinetic steps operative in nucleotide incorporation has been visualized using crystal structures of a variety of family A polymerases bound to relevant intermediates. Polymerase crystal structures in the presence of DNA and dNTP are especially revealing, as this conformation depicts potentially important amino acid interactions within the dNTP binding pocket that lead to high fidelity DNA replication. Structures of such productive replication intermediates (that is, polymerase–DNA–dNTP complexes) exist for family A polymerases (T7 pol[7], Bst pol I[8], and *Taq* pol I[9]). In a recent issue of *Cell*, Steitz and colleagues report the first productive family B polymerase in complex with DNA and incoming nucleotide triphosphate[10], capturing a crucial step just prior to nucleotide incorporation. These crystallography experiments were facilitated by the use of a mutant bacteriophage RB69 polymerase that lacked proofreading activity as a result of mutation of the exonuclease catalytic amino acids Asp 222 and Asp 327 to alanine. In addition, the primer was terminated with a complementary dideoxynucleotide at the 3′ primer terminus, thus preventing nucleotide incorporation, while stabilizing the ternary complex in a polymerizing mode[11]. Comparison of this structure with the previously published RB69 pol–DNA complex containing DNA at the 3′-5′ exonuclease active site (editing complex; Fig. 1)[12] and with similar structures of family A polymerases, provides key insights into how family B polymerases accurately incorporate nucleotides and excise mismatches.

Mechanism of polymerization

Crystal structures of DNA polymerases resemble, in overall morphology, a cupped human right hand, with fingers (which bind an incoming nucleotide and interact with the single-stranded template), palm (which harbors the catalytic amino acid residues and also binds an incoming dNTP) and thumb (which binds double-stranded DNA) subdomains (Fig. 1). During the DNA binding step in family A polymerase structures, the thumb subdomain rotates towards the palm subdomain; this allows the conserved amino acid residues at the tip of the thumb domain to 'grip' the DNA

along the minor groove. DNA bound to the polymerase is bent into an S-shaped conformation; the first bend results from contacts in the minor groove and the second bend forms as the template strand is kinked at the polymerase active site[9]. We have postulated that the latter contacts cause the template base to be flipped away by >90° from the axis of the double helix[13]. Steitz and coworkers now report that binding of DNA causes similar conformation changes within RB69 polymerase, with the tip of the thumb forming nonspecific interactions with the sugar-phosphate backbone along the minor groove. Of note, DNA is in B-form throughout, whereas the conformation is in an A-form within the active site of family A polymerases and HIV-1 reverse transcriptase (RT)[14]. Analogous to polymerases in family A, the single-strand portion of the template is rotated 180° outside the helix axis in RB69 polymerase (Fig. 2).

Nucleotide incorporation by polymerases occurs through an orchestrated sequence of steps in both family A and B DNA polymerases. These two families exhibit significant sequence diversity, but both contain amino acid motifs A, B and C[15], which are well conserved within a family and only partially conserved between families A and B. Structurally, these motifs form the incoming dNTP-binding cleft and adopt nearly identical folding patterns (Fig. 2). The crystal structures of family A and B polymerases reveal that binding of the incoming dNTP is associated with the polymerase adopting a 'closed conformation', characterized by closing of the nucleotide binding site and produced by the rotation of helices found in the fingers subdomain (Fig. 2; green helices). In addition, this protein conformational change is accompanied by 'flipping' of the complementary template base (>90°) by rotation of the base around the phosphodiester bond towards the DNA helix axis[13]. Following these conformational changes upon binding of complementary nucleotide, hydrophobic amino acid residues located in motif B (green in Fig. 2) and motif A (yellow in Fig. 2) form a hydrophobic pocket that surrounds the base and ribose portions of the incoming dNTP. In addition, a hydrophilic pocket surrounds the triphosphate portion of the dNTP, which is coordinated by divalent metal ions (Fig. 2). Thus, faithful incorporation of each nucleotide involves precise positioning of hydrophobic and hydrophilic interactions with the dNTP, two divalent cations, the 3′-primer hydroxyl group, and Watson-Crick base pairing. Together, these events stabilize the 'closed conformation', restrict conformations and structures of the incoming nucleotide, promote the efficiency of correct nucleotide incorporation, and facilitate phosphodiester bond formation between the 3′ primer hydroxyl group and the α-phosphate of the incoming dNTP[7,9–11,13,14]. It is interesting that family A and B polymerases, despite considerable sequence diversity (even within the polymerase active site), bind and incorporate nucleotides by nearly identical mechanisms.

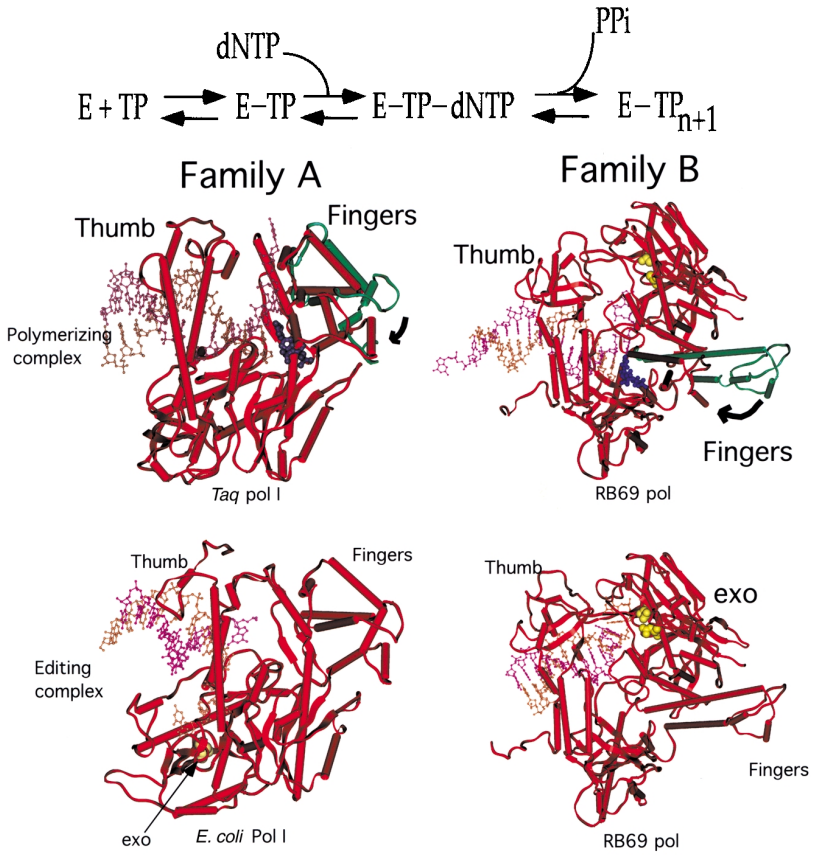

Fig. 1 Structural comparison of family A and B polymerases. Nucleotide incorporation occurs in discrete steps and involves polymerase (E) binding with DNA (TP), followed by binding to dNTP to form the E–TP–dNTP 'polymerizing complex'. During the dNTP-binding step within family A (PDB code 3KTQ) and family B (PDB code 1IG9), the fingers subdomain in the open conformation (shown in green) rotate towards (black arrow) the palm subdomain to produce a closed ternary complex. Amino acids within the thumb subdomain interact along the DNA minor groove during nucleotide incorporation and translocation steps. Noncomplementary basepairs are efficiently excised by polymerases. Partitioning to the proofreading domain (distance of 30–40 Å) involves rotation of the thumb tip and DNA. The location of the exonuclease site (yellow) relative to the polymerase active site differs in the two families of polymerase (accession codes 1CLQ and 1KLN for RB69 pol and *E. coli* pol I, respectively). Family A and B polymerases exhibit similar mechanisms for nucleotide incorporation, as well as for base excision.

Mechanism of proofreading

Incorporation of an incorrect nucleotide at the primer terminus in family A and B polymerases slows extension and allows time for the 3′ primer terminus to partition into the 3′ exonuclease site. Comparison of the bacteriophage RB69 structures in polymerizing[10] and in editing modes[12] shows that partitioning of DNA from the polymerase active site into the exonuclease site (separated by 30–40 Å) is accompanied by a rotation of the DNA along the helix axis and melting of several base pairs at the 3′ primer terminus axis[10] (Fig. 1). The tip of the thumb subdomain rotates with the DNA and maintains the only protein–DNA interactions that are conserved between the two conformations. Thus, it is thought that the movement of the thumb tip directs the partitioning. Comparison of family A polymerases in polymerizing and editing modes yields largely similar results, where the DNA and the thumb tip solely maintain interactions while

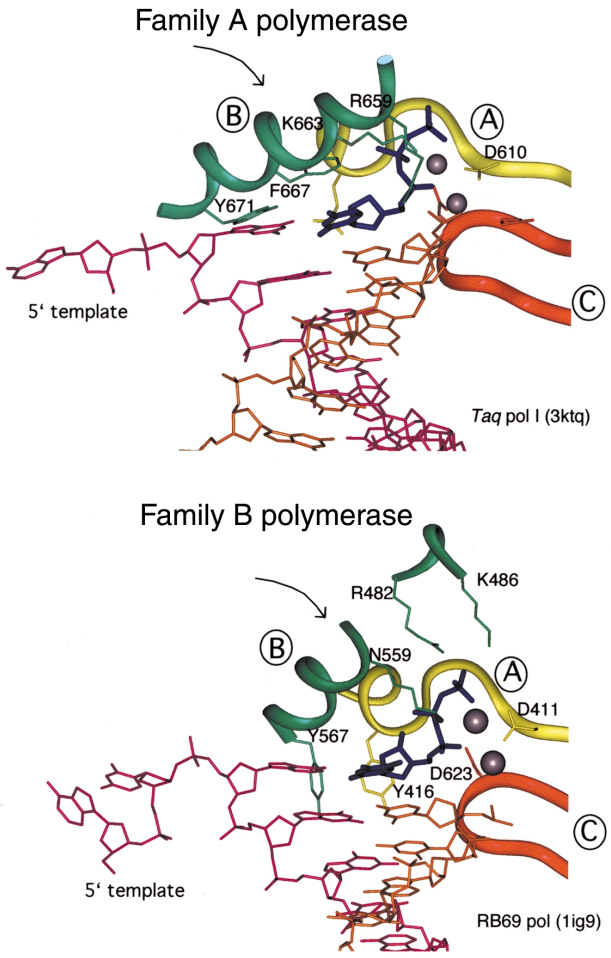

Fig. 2 Comparison of the active sites of family A and B polymerases. The view is from the solvent-exposed major groove and focuses on the dNTP-binding step. Amino acids of motif B (green helix) in the fingers subdomain rotate (black arrow) towards the catalytic palm amino acids in motifs A (yellow) and C (red). The incoming dNTP (blue) is Watson-Crick base-paired with template; the base and ribose portion pack in a hydrophobic pocket and the triphosphate portion is in a hydrophilic environment that includes divalent cations. The remaining single-stranded portion of the template within the active site is flipped 180° out of the helix axis. Despite considerable sequence diversity, family A and B polymerases contain active sites that are structurally superimposible and bind the incoming dNTP in a similar fashion.

rotating ~25° during partitioning the primer terminus 30–40 Å to the exo site (Fig. 1). Notably, the direction of rotation of the thumb tip differs in the two families of polymerases (Fig. 1). In family A polymerases, the enzymes must traverse backward along the DNA and unwind 3–4 base pairs in order for the primer to reach the exonuclease site. In RB69 pol, partitioning to the exonuclease site is preceded by partial melting of the 3′ primer terminus (by >2–3 bases), without polymerase translocation[10,12]. In summary, the polymerase and exonuclease domains in family A and B are separated by a large distance, and in both families the tip of the thumb subdomain is used to partition DNA to a specific active site.

Future directions

Considering the high fidelity of DNA synthesis and the importance of the active site in binding correctly paired incoming nucleotides, it is unexpected that so few active site amino acids in the crystal structure form contacts with the substrates. With family A polymerases, important functional roles can be ascribed to only a few amino acid residues within the catalytic cleft. Random mutagenesis by our laboratory has shown that most active site amino acids (within *E. coli* pol I[16] and *Taq* pol I[17,18]) can be substituted without significant alterations in activity and substrate selectivity. Similarly, RB69 pol has only a very few direct contacts with the incoming nucleotide. These residues include: Asp 411, Asp 623 (both coordinate divalent metal cations), Tyr 416 (interacts with the ribose portion of the incoming dNTP), Arg 482 and Lys 486 (both interact with the triphosphate portion of the incoming dNTP in the closed structure; Fig. 2). The roles for many other amino acids located within these conserved motifs have not been established. We surmise that family B polymerases contain an active site that is also highly mutable.

In summary, Steitz and coworkers describe a 2.6 Å resolution structure of RB69 polymerase in complex with DNA and incoming dNTP. This structure shows significant conformational changes in the fingers and thumb domains relative to the structure in the editing mode[12]. A detailed analysis of the dNTP-binding site identified amino acids with central roles during DNA synthesis. Comparison of the polymerizing mode structure with the previous editing mode structure presents a view of the conformational change that occurs during partitioning between the polymerase and exonuclease sites. Overall, the multistep mechanism of polymerization and base excision by RB69 (family B) appears to be strikingly similar to DNA polymerases in family A. This is especially interesting because many family A polymerases with solved structures are primarily involved in DNA repair (and thus synthesize 1–30 nucleotide stretches of DNA), whereas RB69 polymerase replicates a genome (and thus incorporates stretches of 1,000s of nucleotides). Disadvantages of this study include the fact that the polymerase is bound with an inhibiting cation, Ca^{2+} (whereas the natural metal cofactor is Mg^{2+}), thus detracting from a definitive model for phosphodiester bond formation for this enzyme. Additional structures of RB69 polymerase bound to DNA at the active site in the presence of noncomplementary nucleotides but in the absence of dNTP should further delineate the conformational changes following DNA binding and provide novel insights into the participation of the enzyme in base selection.

In the past decade, a variety of DNA polymerase structures in complex with physiologically relevant intermediates have provided insights into the various steps of catalysis by these enzymes. Much of the effort has focused on a handful of polymerases, including Pol I[19], HIV-1 RT[14,20] and DNA pol β[11]. We learned that these polymerases use an identical phosphoryl transfer mechanism. However, it is unclear how polymerase structure relates to cellular function. Why do polymerases function predominantly in either replica-

tion or repair? Why can some polymerases use either RNA or DNA templates? How are mispairs formed and extended and what governs processivity, damaged template bypass and so forth? In the near future, other polymerase structures will be determined and their mechanism compared. Ultimately, we await the structure of entire replication complexes, including processivity factors, helicases and single-strand binding proteins to yield clues regarding how these proteins work in concert in different cellular DNA transactions.

Acknowledgments
We are very grateful to C. Verlinde for help with structural illustrations.

Premal H. Patel and Lawrence A. Loeb are at the Joseph Gottstein Memorial Cancer Laboratory, University of Washington School of Medicine, Seattle, Washington, USA, 98195-7705. Correspondence should be addressed to P.H.P. email: ppatel@u.washington.edu or L.A.L. email: laloeb@u.washington.edu

1. Kornberg, A. *Science* **131**, 1503 (1960).
2. Tang, M. *et al. Proc. Natl. Acad. Sci. USA* **95**, 9755–9760 (1998).
3. Kuchta, R.D., Mizrahi, V., Benkovic, P.A., Johnson, K.A. & Benkovic, S.J. *Biochemistry* **26**, 8410–8417 (1987).
4. Kunkel, T.A. & Loeb, L.A. *Science* **213**, 765–767 (1981).
5. Kunkel, T.A. & Bebenek, K. *Annu. Rev. Biochem.* **69**, 497–529 (2000).
6. Ollis, D.L., Brick, P., Hamlin, R., Xuong, N.G. & Steitz, T.A. *Nature* **313**, 762–766 (1985).
7. Doublie, S., Tabor, S., Long, A.M., Richardson, C.C. & Ellenberger, T. *Nature* **391**, 251–258 (1998).
8. Kiefer, J.R., Mao, C., Braman, J.C. & Beese, L.S. *Nature* **391**, 304–307 (1998).
9. Li, Y., Korolev, S. & Waksman, G. *EMBO J.* **17**, 7514–7525 (1998).
10. Franklin, M.C., Wang, J. & Steitz, T.A. *Cell* **105**, 657–667 (2001).
11. Pelletier, H., Sawaya, M.R., Kumar, A., Wilson, S.H. & Kraut, J. *Science* **264**, 1891–1903 (1994).
12. Shamoo, Y. & Steitz, T.A. *Cell* **99**, 155–166 (1999).
13. Patel, P.H., Suzuki, M., Adman, E., Shinkai, A. & Loeb, L.A. *J. Mol. Biol.* **308**, 823–837 (2001).
14. Patel, P.H. *et al. Biochemistry* **34**, 5351–5363 (1995).
15. Delarue, M., Poch, O., Tordo, N., Moras, D. & Argos, P. *Protein Eng* **3**, 461–467 (1990).
16. Shinkai, A., Patel, P.H. & Loeb, L.A. *J. Biol. Chem.* **276**, 18836–18842 (2001).
17. Patel, P.H. & Loeb, L.A. *Proc. Natl. Acad. Sci. USA* **97**, 5095–5100 (2000).
18. Suzuki, M., Baskin, D., Hood, L.E. & Loeb, L.A. *Proc. Natl. Acad. Sci. USA* **93**, 9670–9675 (1996).
19. Beese, L.S., Derbyshire, V. & Steitz, T.A. *Science* **260**, 352–355 (1993).
20. Huang, H., Chopra, R., Verdine, G.L. & Harrison, S.C. *Science* **282**, 1669–1675 (1998).

Unfolded, yes, but random? Never!

Kevin W. Plaxco and Michael Gross

NMR experiments with partially aligned protein molecules in strongly denaturing conditions suggest that the unfolded state is less chaotic than is widely believed. Hence protein folding is probably less paradoxical than Levinthal originally thought.

Unfolded proteins are seen as a heterogeneous mess. Each amino acid contributes two freely rotating bonds to the backbone of the polypeptide chain, and thus even a small protein (100 amino acid residues) has an astronomical number of configurations it can adopt when unfolded (~10^{100}). This conformational diversity is thought to greatly complicate the process of finding the unique native structure, and the loss of its entropy is the largest single component of the free energy of folding. The surprising fact is that despite this complexity, most proteins still manage to fold up in a biologically relevant time, an apparent paradox first attributed to Levinthal. A great deal of research in protein folding has been aimed at finding the kinetic tricks by which protein chains speed up this search, with paradigms including first pathways, then folding nuclei, and most recently smooth energy landscapes biased in favor of the native state. But are we really certain that unfolded states are astronomically complex? Recent results suggest there may be a surprising simplicity to this seemingly heterogeneous mess.

Experimental characterization of the ensemble of states populated by an unfolded protein has proven extremely difficult. Traditional approaches, such as small angle X-ray or neutron scattering (SAXS[1] and SANS[2], respectively) provide only an ensemble average measure of global physical properties (in this case the radius of gyration, R_g). They provide nothing even remotely approaching a three-dimensional model of the state with atomic detail. As reported in a recent issue of *Science*[3], by addressing this problem with an NMR technique recently developed to refine solution structures of folded proteins to higher resolution, David Shortle and Michael Ackerman at Johns Hopkins University have provided the first tantalizing glimpse of the detailed structure of a 'highly' unfolded state. What they see is a state of perhaps surprising simplicity.

Shortle and Ackerman analyzed dipolar couplings, which measure the angle between the vector separating two magnetic nuclei and the direction of the applied magnetic field. As a protein tumbles in solution, this angle normally assumes all possible values and the dipolar couplings average to zero. In contrast, if this symmetry is broken, say by the interaction of the protein with an asymmetric environment, the resulting net anisotropy can produce measurable couplings. In order to create an anisotropic environment for the protein to tumble in, Shortle and Ackerman diffused their sample into a polyacrylamide gel contained in NMR tubes[3]. By simply squashing or stretching the gel they deformed the normally spherical microscopic cavities of the gel to ellipsoids ('UFOs' or 'cigars', respectively). This, they demonstrated, creates a small net alignment of the protein molecules relative to the external magnetic field, which is sufficient to produce measurable dipolar couplings. With folded proteins, it is even possible to calculate a backbone structure from this information[4].

Using this methodology, Shortle and Ackerman analyzed the 131-residue variant of staphylococcal nuclease resulting from truncation of both termini, known as Δ131Δ. Previous spectroscopic studies of this fragment of the well-characterized α-β protein demonstrated that, although it is unfolded in the absence of denaturants (for example, its R_g is 50% greater than that of the folded protein, obtainable in the presence of ligands), it assumes a