Published in final edited form as:
Chem Rev. 2014 March 12; 114(5): 2759–2774. doi:10.1021/cr3005179.

Structural Comparison of DNA Polymerase Architecture Suggest a Nucleotide Gateway to the Polymerase Active Site

Sangwook Wu[†,§], **William A. Beard**[‡,§], **Lee G. Pedersen**[†,*], and **Samuel H. Wilson**[‡,*]

[†]Department of Chemistry, University of North Carolina, Chapel Hill, NC 27599-3290

[‡]Laboratory of Structural Biology, NIEHS, NIH, Research Triangle Park, NC 27709-12233

1. Introduction

1.1. Overview

DNA polymerases are essential for the repair and replication of genomic DNA. These enzymes must rapidly bind and incorporate the correct deoxynucleoside triphosphate (dNTP) from a pool of chemically and structurally similar molecules in a template-dependent manner to ensure high fidelity DNA synthesis. Appropriately, DNA binding precedes nucleotide selection.[1,2] In recent years, a wealth of structural information, primarily from crystallographic structures, characterizing the structure of DNA polymerases from diverse sources has hastened our molecular understanding of the critical and essential role they play in faithful genome replication and repair. Based on primary sequence, DNA polymerases are grouped into families.[3] This article seeks to assimilate structural information across DNA polymerases families to uncover molecular attributes that facilitate nucleotide binding and catalysis. We analyze structural similarities and differences between representative DNA polymerases from different families. By performing multiple structural alignments on fifteen polymerase ternary substrate complex crystal structures, the alignment confirms that the catalytic cores are conserved between families. It includes two key active site metal ions along with two critical bridging aspartate residues, and one variable acidic residue (aspartate or glutamate) that coordinate these metals. These elements, as well as the incoming nucleotide, are highly conserved in structural space between polymerase families. Several other charged residues are conserved among catalytic cores across polymerase families. Importantly, we describe a tubular “channel” leading from the enzyme boundary to the active site of the polymerases in three families; A, B, and X. The lack/presence of this channel suggest that the mechanism of mobilizing a dNTP for insertion may differ across polymerase families. In addition, domains other than the polymerase domain play a role in formation of this channel. The features of DNA synthesis may depend on how the molecular architecture of the channel affects substrate (right and wrong) access and product (PP_i) release. After a general description of DNA polymerase architecture and the active site for each family, we discuss dNTP diffusion into the active site.

[*]Corresponding author information: L.G.P. telephone, 919-962-1578; fax, 919-962-2388; lee_pedersen@unc.edu. S.H.W. telephone, 919-541-4701; fax, 919-41-4724; wilson5@niehs.nih.gov.
[§]These authors contributed equally to this work.

1.2. DNA Polymerase Architecture

Based on protein sequence homology, DNA polymerases are grouped into at least 5 families: A, B, C, X, and Y (Table 1).[3,4] Crystallographic structures of members of each family have been characterized in various liganded forms. These structures indicate that the proteins are multi-domain enzymes that include an accessory domain with an activity that facilitates their respective biological function (Table 1 and Figure 1). In the case of replicative DNA polymerases, an intrinsic 3′–5′-proofreading exonuclease activity often resides on a domain that improves overall DNA synthesis fidelity by removing misinserted nucleotides. In the case of a repair DNA polymerase, DNA polymerase (pol) β (X-family) includes an amino-terminal deoxyribose phosphate (dRP) lyase domain that removes the 5′-sugar-phosphate of an incised abasic site during the repair of simple DNA base lesions.[5] The polymerase domain of replicative and repair polymerases are composed of three subdomains, as illustrated in Figure 1. Additionally a group of DNA polymerases, primarily from the Y-family, are utilized to hasten DNA synthesis through damaged DNA. These polymerases often include protein subdomains of unknown function.

The DNA polymerase domain in various polymerases has been likened to a right-hand with fingers, palm, and thumb subdomains.[21] The palm subdomain coordinates two metals (Mg^{2+}) necessary for catalysis with three acidic amino acid side chains and is structurally homologous with members of the A-, B-, and Y-family polymerases (Figure 1). Although the C- and X-family polymerases bind substrates and metals in an analogous manner, their palm subdomain is not homologous with those of the other families. The subdomains of pol β are referred to as N (nucleotide binding/selection), C (catalytic), and D (DNA-binding) and are equivalent to the fingers, palm, and thumb, respectively, of right-handed polymerases (Figure 1).[22]

Long after the discovery of *Escherichia coli* DNA polymerase I by Kornberg and associates,[23,24] structural studies revealed that substrate binding to DNA polymerases induces structural rearrangements that facilitate selection of the correct incoming nucleotide. Comparing structures of binary DNA/polymerase complexes with those that include an incoming nucleotide (i.e., ternary complex) indicates that a subdomain (fingers or N-subdomain) often repositions itself to close upon the nascent base pair.[25,26] This results in a nascent base pair that is sandwiched between the primer terminus and polymerase. Critically, however, the architecture of the DNA-bound state of polymerases constrains the path that the nucleotide must traverse or diffuse to enter the polymerase active site. Thus, a feature of the molecular architecture of X-family members, that provided the original motivation for this study, is a well-defined “tubular” channel that leads to the active site from the surface of the enzyme (Figure 1). Since DNA restricts access to the active site, nucleotides must diffuse through this channel. Once the nucleotide gets to the active site, a specific geometry must be achieved with two divalent metal ions, DNA primer 3′-OH, and incoming nucleotide to ensure high efficiency DNA synthesis. We identified a similar channel in the structures of A- and B-family DNA polymerases, but a well-defined channel that serves as a gateway to the active site seems absent in the other families (Figure 1). It remains to be determined whether this limitation influences catalytic efficiency, fidelity, or biological function.

The identity, crystallographic structures, accessory domains, and biological function of the polymerases analyzed in this study are outlined in Table 1. In general, replicative DNA polymerases belong to the B- (eukaryotic) or C- (bacterial) families. Members of the X-family are primarily involved in gap-filling DNA repair while Y-family DNA polymerases are believed to be involved in translesional DNA synthesis. The A-family is composed of DNA polymerases that are involved in both DNA repair and replication.

2. Structural Homology

2.1. Polymerase Subdomains

As noted above, the polymerase domain of most DNA polymerases are comprised of three subdomains. The architectural arrangement of the subdomains of A-, B-, and Y-family DNA polymerases have been compared to a right-hand and are referred to as fingers, palm, and thumb. Although the structures of the fingers and thumb subdomains are distinct among these families, the palm subdomains are homologous.[27] The palm consists of two α-helices stacked against a β-sheet. The active site acidic residues that coordinate two divalent metals necessary for nucleotidyl transfer are found in the palm subdomain. In contrast, the equivalent subdomain of pol β (X-family) is structurally similar, but topologically distinct, from these other polymerase families. The strands of the β-sheet are anti-parallel in the palm subdomain of A-, B-, and Y-family DNA polymerases. For pol β, the β-sheet is mixed with essential acidic residues on parallel strands.[28] The structural ambiguity in the catalytic palm subdomain among the various polymerase families has confounded the subdomain nomenclature of X-family DNA polymerases. This is because the fingers and thumb subdomains can be defined by a structural alignment of the catalytic palm subdomains, ignoring topological differences, or a functional alignment of the catalytic participants (metals, dNTP, DNA).[29] A functional alignment of pol β with other DNA polymerases of known structure results in a consistent functional nomenclature of the subdomains; the fingers and thumb subdomains of different DNA polymerases are functionally equivalent. In contrast, the nomenclature based on the original structural alignment of the catalytic subdomain of pol β defines the subdomains opposite to that of the functional alignment. Since the topology of the catalytic subdomain of X-family members is unique, an alternative is to consider members of the X–family as left-handed, rather than the original right-hand analogy.[30] This approach highlights the non-homologous nature of the catalytic palm subdomains, but requires prior knowledge of the architectural origin of the nomenclature. A functional alignment would be the simplest approach, but the hand-like analogy offers no functional insight. Accordingly, a functionally-based nomenclature where the subdomains are referred to as C– (catalytic), D– (DNA binding), and N–subdomains (nucleotide binding/selection) to highlight their intrinsic function.[22] These correspond to the palm, thumb, and fingers subdomains, respectively, for polymerases that utilizes the architectural analogy to a right hand.

Even though replicative C-family DNA polymerases do not show primary sequence similarity with members of the X-family, their catalytic subdomain exhibits strikingly similarity to that of pol β and not to the eukaryotic B-family replicative enzymes.[12,13] Based on the structure of the catalytic subdomain of polymerases from the various families, there

appears to be two superfamilies of DNA polymeraes that can be separated by the topology of the β-sheet in the catalytic subdomain: anti-parallel (A-, B-, and Y-family) or mixed (C- and X-family). Although members of these alternate superfamilies appear to have evolved separately, they share many general structural and mechanistic features with polymerases from these other families.

2.2. Structural Alignment of DNA Polymerases Within Families

To structurally align DNA polymerase structures, the structural alignment program STAMP,[31] a component of the VMD package,[32] was employed that minimizes C_α distance between aligned residues of each crystal structure by rigid-body rotation and translation. The multiple structure alignments (MSA) of ternary complex crystal structures from representative members of A-, B-, C-, X- and Y-families were analyzed (Table 1). The crystal structures within each family were aligned according to a C_α distance minimization criterion. After MSA, pseudo-colors were assigned to each residue according to Q_{res},[33] from blue (Q_{res} = 1) to white (Q_{res}= 0). Q_{res} is a similarity order parameter of the aligned residues:

$$Q_{res}^{(i,n)} = N \sum_{(m \neq n)}^{proteins} \sum_{(j \neq i-1,j,j+1)}^{residues} \exp\left[-\frac{\left(r_{ij}^{(n)} - r_{i'j'}^{(m)} \right)^2}{2\sigma_{ij}{}^2} \right] \quad (1)$$

where $Q^{i,n}{}_{res}$ is the structural similarity of the i^{th} residue in the n^{th} protein, $r^{(n)}{}_{ij}$ is the C_α distance between residues i and j of the protein, N is a normalization constant and σ_{ij} is the variance, indicating sequence separation of the structural alignment between residues i and j: $\sigma_{ij}{}^2 = |i-j|^{0.15}$.[33]

Multiple structure alignment was initially performed for three crystallographic structures of members of the X-family of DNA polymerases. Three high-resolution ternary complex structures (pol/DNA/dNTP) were compared: pol β (PDB ID 2FMS),[14] pol λ (PDB ID 2PFO),[15] and pol μ (PDB ID 2IHM).[16] These structures include two essential metal ions at the catalytic core making them excellent candidates for comparison. Figure 2 shows the resultant overlay of the three structures after MSA. The ribbon representations are colored according to Q_{res}, a similarity order parameter of each aligned residue. The alignment indicates Leu551 of pol λ has the largest value of Q_{res} (0.86). A functional role for this residue has not been identified suggesting that its conservation is due to a structural role. Not surprisingly, the active site aspartates that coordinate catalytically essential divalent metal ions exhibit high Q_{res} values (≥0.77). The dark blue ribbon representation indicates that many of the residues of these polymerases are structurally well conserved (Figure 2). However, localized elements shown in white are also observed, highlighting where structural conservation is absent.

The residues in the amino-terminal 8-kDa lyase domain of pol β are highly conserved among X-family members. Although the domain exhibits high structural identity, the deoxyribose phosphate lyase activity exhibited with pol β and pol λ is absent in pol μ.[16] A particular lysine residue in the 8-kDa domain (Lys72, pol β; Lys312, pol λ) is the primary nucleophile in the dRP lyase activity while valine sits at this position in pol μ.[16,34,35] In

addition to its lyase activity, another critical role of this domain in X-family polymerases is targeting the enzyme to the 5′-phosphate in gapped DNA during replication/repair.[36–38] The three polymerase subdomains, along with the lyase domain, form a donut-like structure that creates a tubular channel (Figures 1 and 2). Since DNA enters the active site from one side of the polymerase structure, nucleoside triphosphates must enter the active site from the opposite side. This channel is observed in all three ternary complex structures from X-family DNA polymerases (e.g., pol β; Figure 1) and is described in detail below. Despite catalytic and sequence differences in the 8-kDa domains of these three polymerases, the 8-kDa domain is an essential component of this channel that influences the path of the incoming nucleotide.

Figure 2 illustrates that the polymerase domains of three A-family ternary complex structures analyzed have very similar molecular architectures. The largest Q_{res} for an A-family member was 0.88 for Asn580 of Klentaq DNA polymerase. Yet, as with the X-family enzymes, there were regions that lacked homology and are illustrated in white (Figure 2). In addition, the analysis revealed an apparent channel where dNTPs could diffuse into the active site (e.g., pol β; Figure 1). The T7 structure also includes thioredoxin, an extrinsic accessory domain believed to act as a processivity subunit. In the crystal structure, it is distant from the short duplex required for crystallization and does not form part of the channel. Accordingly, it is not shown in Figure 2.

We found that B-family DNA polymerases have a more diverse molecular architecture than the corresponding enzymes in the A- and X-families (Figure 2). In this case, the highest Q_{res} is only 0.71. Interestingly, a tubular channel is not easily recognized in the absence of the 3′-5′ exonuclease domain in B-family polymerases (Figure 1). As observed with the yeast pol δ ternary complex structure, the 3′–5′ proofreading exonuclease domain is essential for the structural integrity of the channel that provides access to dNTPs (Figure 3). This is similar to the structural role provided by the 8-kDa domain of X-family enzymes where an intrinsic accessory domain restricts nucleotide access to the active site (e.g., lyase domain of pol β, Figure 1).

The aligned structures after MSA for C- and Y-family representatives are also shown in Figure 2. The alignments illustrate relatively conserved structures (blue ribbons) except for some elements that are unique to each polymerase (white regions). A channel that might restrict nucleotide access to the active site is not clearly observed in the Y-family structures (e.g., pol ι in Figure 1). The fingers and thumb subdomains are small compared to those in other families and the active sites are solvent exposed. The small size of the fingers and thumb subdomains may be related to the low-fidelity of Y-family DNA polymerases.[39] The C-family enzymes also do not have a clear channel (Figures 1). However, in one member of the C-family, the 3′–5′ exonuclease domain of PolC (PDB ID 3F2B) was deleted for structural characterization and may provide a component of the channel.[12] Considering that the 3′–5′ exonuclease domain is structurally integral to the channel in B-family DNA polymerases, the possibility of a channel for nucleotide entry to the active site should not be excluded in C-family enzymes.

2.3. Catalytic Core

2.3.1. Structural Alignment Within Families—The channels noted in several families of DNA polymerases lead from the surface into the active site or catalytic core. The catalytic core for X-family DNA polymerases is illustrated for pol β (Figure 4), and it exhibits high conservation among the family members. Essential elements of the core are the bound incoming dNTP, two octahedral-coordinated divalent metal ions (Mg^{2+}), the 3′-DNA primer-terminus, and active site aspartates (190, 192, and 256) (Figure 4A). We examined the relative positions of the pro-R_P non-bridging oxygen of the α-phosphate of the bound dNTP and two divalent metal ions. These three elements play a critical role in chemistry for nucleotide insertion. The relative positions of these three atoms after MSA for different X-family DNA polymerases are tabulated in Table 2. The results indicate that these atoms are conserved in geometric space. By performing MSA for the three ternary complexes of pol β (PDB ID 2FMS), pol λ (PDB ID 2PFO), and pol μ (PDB ID 2IHM), we observe that the three aspartate residues and metals are conserved in structural space as well as sequential space (Figures 4A and 5; Table 2).

The conservation in the positioning of the pro-R_P oxygen of the α-phosphate of the bound dNTP and two divalent metal ions in structural space is also observed in A-family DNA polymerases. However, the active site acidic residues that coordinate the metal ions include a glutatmate residue (Asp475, Asp654 and Glu655 in T7). The glutamate in T7 DNA polymerase shows a somewhat different position than the corresponding aspartate (e.g., Asp256, pol β) of X-family and the other members of the A-family. An analogous glutamate in the catalytic core is also observed in members of the Y-family. The relative positioning of these atoms with respect to other A-family members is tabulated in Table 3.

For B-family members, the position of catalytic residues and the three aspartate residues are conserved in both structural and sequential space. However, there is some variation in the divalent metal ion composition of several of the crystallographic structures. The pol δ (PDB ID 3IAY) active site has three Ca^{2+} ions (Figure 5).[10] The third divalent metal in pol δ has been suggested to facilitate product pyrophosphate dissociation.[10] The other Ca^{2+} ions in pol δ are positioned as other divalent metals in other DNA polymerase families. While there is some variation in the number and position of the divalent metal ions in B-family enzymes, the relative position of the catalytic participants appears to be conserved. Structural and sequential space conservation of the three aspartate positions of the B-family structures is also observed. However, unlike the X-family, the three aspartate residues (Asp249, Phi29; Asp608, pol δ; Asp411, RB69) located in the catalytic core after MSA, exhibit a somewhat different orientation in each structure (not shown). Despite the variations in acidic residue identity and metal ion, the conservation of three negatively charged residues (aspartates and/or glutamates) and two divalent metal ions in both structural and sequential space are common features in A-, B- and X-families irrespective of species.

The ternary complex structures analyzed above represent pre-catalytic complexes that have been trapped using a dideoxy-terminated primer terminus (i.e., O3′ absent), inert divalent metal (i.e., Ca^{2+}), or non-hydrolysable incoming nucleotide analog (i.e., bridging oxygen between P_α and P_β substituted with carbon or nitrogen). Recently, ternary complex

crystallographic structures of intermediate complexes undergoing nucleotidyl transfer have been captured with natural substrates and metals. This was achieved by generating crystals of pre-catalytic substrate complexes in the presence of Ca^{2+} and initiating the reaction through ion exchange with Mg^{2+}. The reaction is stopped by freezing the crystal as the reaction progresses (i.e., after a defined time period) and the structure is determined. This approach has successfully trapped ternary product complex structures of pol η (Y-family)[40] and pol β (X-family).[41] Interestingly, an additional transient divalent metal site is observed bridging oxygen atoms on the products (Figure 4B). This metal appears to be associated with the product state and postulated to be involved in pyrophosphorolysis, the reverse reaction of DNA synthesis.[41] It is observed during correct, but not incorrect nucleotide insertion. Additionally, it is expected to be unique to X- and Y-family DNA polymerases since other DNA polymerases have a conserved basic side chain that occupies this position.[25,41]

2.3.2. Structural Alignment of Catalytic Cores From Various Families— Although it is not unexpected that a structural conservation was observed among family members or that the active site of polymerases from various families is conserved, it is of functional interest to examine whether their might be a structural conservation beyond the active site for DNA polymerases belonging to different families. Investigating the structurally conserved assembly units between different families is especially challenging. Unfortunately, the STAMP alignment algorithm[31] could not be directly applied to DNA polymerases in different families, because of their dissimilar architectures and gene sequences. Thus, in an attempt to compare the chemical environment of the active site in various family members, we selected the three catalytic atoms described above (two metals and the pro-R_P oxygen of the α-phosphate of the bound dNTP) to define an alignment platform. This permitted us to examine the conservation of residues in the vicinity of the active site of members from different polymerase families. Figure 6 shows the relative positions of the divalent metal ions, and the bound dNTP and DNA primer terminus after the alignment of three crystallographic structures from three different families (Klentaq, A-family; pol δ, B-family; pol β, X-family). Overall, the figure illustrates that the triphosphate groups, metals and primer terminus are structurally conserved between families. The aspartate (Asp762) that may be involved in primer O3′ deprotonation is somewhat displaced in the B-family representatives (Figure 6B).

2.3.3. Active Site Architecture—Since substrates and cofactors are highly charged, it is reasonable to expect that electrostatic interactions provide the dominant force that assembles catalytic participants (i.e., incoming dNTP, metals, active site acidic residues and primer terminus). As indicated previously for pol β,[42] charged residues beyond the active site (i.e., second shell), may contribute to stabilization of the transition state. These residues would also be expected to contribute to the dynamics involved in the conformational adjustments that must occur to achieve the geometry necessary for catalysis. By performing MSA for the three X-family closed ternary complexes, i.e., for pol β (PDB ID 2FMS), pol λ (PDB ID 2PFO), and pol μ (PDB ID 2IHM), we find that two arginine residues likely contribute to substantial stabilization of the active site; Arg254 and Arg183 (pol β), Arg488 and Arg420 (pol λ) and, Arg418 and Arg323 (pol μ). Figure 7 illustrates the relative positions of these

residues with respect to the catalytic core of pol β. The equivalent residues in the other X-family DNA polymerases superimpose with those from pol β highlighting their structural and sequence conservation; Gly189 (pol β), Gly426 (pol λ), and His329 (pol μ) also structurally align. The histidine in pol μ is within hydrogen bonding distance to the phosphate backbone of the primer terminus and the γ-phosphate of the incoming nucleotide. These interactions may facilitate template-independent polymerization.[16]

We also performed an alignment with nine polymerase structures from other families with respect to the pro-R_P oxygen of the α-phosphate and two divalent metal ions. The aligned structures are: T7 (PDB ID 1T7P),[8] Klentaq (PDB ID 3KTQ),[6] and BF (PDB ID 3EZ5)[7] from the A-family; Phi29 (PDB ID 2PYJ)[9] and pol δ (PDB ID 3IAY) from the B-family; pol β (PDB ID 2FMS) and pol λ (PDB ID 2PFO) from the X-family; pol ι (PDB ID 3H4D)[18] and pol η (PDB ID 3MR2)[20] from the Y-family. Several structures only included a single divalent metal and were excluded: PolC (PDB ID 3F2B, C-family), PolIII (PDB ID 3E0D, C-family)[13] and pol (PDB ID 3IN5, Y-family)[19] and pol μ (PDB ID 2IHM, X-family). Additionally, the two active site Ca^{2+} ions in Dpo4 (PDB ID 2R8H, Y-family)[17] appear to distort the active site as judged by large deviations in the positions of active site residues relative to the other structures in the alignment, and this structure also was not used.

The structural alignment of the catalytic core and conserved active site residues in these nine structures is shown in Figures 8 and 9. The imaginary plane, composed of the pro-R_P oxygen of the α-phosphate and two divalent metal ions (Mg^{2+}) in pol β, is shown as a reference plane in Figure 8 (semi-transparent gray disk). The two metal ions from the other structures are projected relative to the plane after the alignment. Figure 8 illustrates that all of the metal ions from the nine structures, except pol ι (Y-family), superimpose well. The positions of two of the metal ions (orange) of pol ι are displaced slightly from the consensus. Interestingly, whereas the phosphates of the incoming nucleotides superimpose well, the sugar and bases do not suggesting that the nascent base pair geometry in these structures differ somewhat. Following this approach, we traced the positions of key active site residues in the catalytic core and other conserved charged residues. In Figure 9, the nitrogen atoms of basic side chains within hydrogen bonding distance to either of the phosphates of the incoming nucleotide are shown. Other atoms (i.e., oxygen and nitrogen) that might influence the active site environment are shown in Figure 10C. There is a cluster of lysine residues near the α-phosphate of the dNTPs (Figure 9, top panel) and a histidine cluster, A-family, around the β-phosphate of the incoming dNTP (Figure 9, middle panel). The cluster of lysine residues near the α-phosphate could easily coordinate the pro-S_P oxygen, but in the case of the α-phosphate in X- and Y-family DNA polymerases, there is a lack a basic side chain that might coordinate the pro-S_P oxygen.[25] Arginines from X-family polymerases are clustered around the β-phosphate (Figure 9, middle panel). Lysines from Y-family and arginines from A-family polymerases are found in clusters near the γ-phosphate (Figure 9, lower panel). These arginines and lysines are distributed in a very broad region enveloping the scaffold of the phosphate backbones of the incoming dNTPs; along with the divalent metals, they neutralize the charge on the triphosphate as well as play a critical role in catalysis. It is likely that the basic side chains of these residues also play an important role in targeting the incoming nucleotide to the polymerase active site.[43] In addition, the structural

alignment indicates that subtle electrostatic differences exist between polymerases from different families and this may impact substrate specificity (i.e., ability to select right from wrong; fidelity). In addition to a proofreading activity, replicative DNA polymerases exhibit high fidelity due to their efficient insertion of the correct nucleotide.[44] The high fidelity and correct insertion efficiency exhibited by replicative DNA polymerases is due to their rapid insertion of a correct nucleotide, while low fidelity enzymes exhibit slow rates of insertion. Since both low and high fidelity DNA polymerases bind the correct nucleotide with similar efficiencies, fidelity is linked to the rate of correct nucleotide insertion. It is expected that the basic side chain interactions observed with the pro-S_P oxygen of the α-phosphate of the incoming nucleotide with replicative polymerases noted above will facilitate correct nucleotide insertion.[25]

From the structural alignment, we observed that the two active site aspartates that coordinate both metals are highly conserved in position (Figure 10A and B). The oxygens of these aspartates converged tightly around the metal ions (Figure 10B). The lower efficiency and fidelity of DNA synthesis exhibited by Y-family DNA polymerases is most likely reflected in the positioning of charged active site residues in the catalytic core. As noted above, X- and Y-family DNA polymerases do not exhibit a conserved basic side chain interacting with the pro-S_P oxygen of the α-phosphate. The lack of this interaction would be expected to diminish the rate of correct nucleotide insertion thereby decreasing fidelity.[25] Likewise, the aspartate or glutamate residues located near the 3′-OH primer terminus (Figure 10C) exhibit a broad range of positions in contrast to the bridging aspartate residues (Figure 6B). Structurally conserved residues near the active site are tabulated for each polymerase family in Table 4. The role of this Asp/Glu near the primer terminus is of immense interest. Site-directed mutagenesis of this residue in pol β (Asp256) coupled with crystallographic, activity pH-profiles, and computational studies indicate that this residue plays a central role in nucleotidyl transfer.[45] In addition to coordinating the catalytic metal, Asp256 coordinates O3′ of the primer terminus and serves as a general base upon O3′ activation; i.e., $O3'^-$ formation.[42] Thus, the pK_as of both the donor (primer O3′) and the acceptor group (OD2 of Asp256) are regulated by the catalytic metal ion.

3. DNA Polymerase Channels

3.1. Nucleotide Access

An electrostatic surface representation of DNA polymerases often identifies a cleft where DNA binds; basic protein side chains interact with the negatively charged sugar-phosphate backbone of the DNA. As noted above, structures of ternary substrate complexes indicate that the path accessible for an incoming nucleotide is restricted to an apparent channel leading from the surface into the polymerase active site. Previous calculations have suggested that nucleotide diffusion into the active site of RNA polymerase is severely restricted.[49] Before the characteristics of these channels in DNA polymerases are described, it is useful to compare the relative total solvent-excluded volume of each DNA polymerase. From available crystal structures, it is evident that pol β (X-family) is the smallest polymerase examined and those from the replicative C-family the largest. However, since domains or flexible loops are often removed to facilitate crystallization, the calculated

volumes should be considered a minimum estimate. We have listed the ratio of residues present in the crystallized structure (C) to the total number of residues (T) for representative members of the various polymerase families in Table 5. Thus, the size of the crystallized enzymes are: C-family (Pol C) > B-family (Phi29) > A-family (Klentaq) > Y-family (pol ι) > X-family (pol β).

When the surface of a ternary substrate complex structure is viewed from an appropriate angle, an apparent pore or channel that leads to the active site is observed (e.g., with pol β; Figure 11). Figure 12 illustrates the calculated channel for the binary DNA and ternary (+dNTP) complex structures of pol β. In these illustrations, the pink opaque surface designates the surface of the calculated channel. In Figure 12, the protein in the vicinity of the channel is semi-transparent to enable illustration of how the protein influences the topology of the channel. For the binary DNA complex, the channel can be viewed as the environment through which the dNTP must traverse to reach the active site/primer-terminus. For the ternary substrate complex structures, the channel represents a constrained path for the departure of pyrophosphate. We quantify and compare, for most cases, the shapes and volume of these entrance and exit channels. These estimates provide boundaries and insights into nucleotide binding and strategies used by different polymerases to limit or trap substrates/products necessary for DNA repair and replication synthesis.

3.2. Depth

To quantify features of the nucleotide access channel, three amino acid C_α atoms define an imaginary plane perpendicular at the channel entrance. The distances from this plane to the C3′ of the DNA primer terminus and the center of mass of the bound incoming dNTP defines the depth of the channel. For instance, in the case of pol β, we selected C_α (Asp17, lyase domain; Arg102, D-subdomain; and Arg149, C-subdomain) to define the imaginary plane at the entrance of the channel. Although the catalytic event actually occurs at O3′ of the primer terminus, this atom is often missing in crystallographic structures, so that chemistry is abrogated. The distances from the pol β primer terminus C3′ and the center of mass of bound dNTP to the channel entrance described above are 13.3 and 13.1 Å, respectively. For Klentaq (A-family), the distances are 12.5 and 10.8 Å, respectively. Similarly, for Phi29 (B-family), the distances are 22.4 and 24.9 Å, respectively. These values provide an estimate for the depth of the respective channels that a nucleotide must traverse as it approaches the polymerase active site (Table 6). Thus, a nucleotide must diffuse through a confined space considerably further with Phi29 than with the other polymerases examined.

3.3. Volume

The volume of the channel can be estimated using the 3V algorithm.[50] This approach adopts two rolling probe spheres (large and small). The larger rolling probe (r = 8 Å; shell volume) scans the entire surface of the DNA polymerase (only protein) while a smaller probe (r = 3 Å; solvent-excluded volume) is used to determine the space that is “unoccupied” in a cavity or channel by subtracting the solvent excluded volume from the shell volume. We define the dNTP access channel volume, by estimating the volume of a channel in the absence of dNTP and DNA strands (A) minus the volume of the dNTP and DNA that penetrate the

channel (B) + volume of free dNTP (C); i.e., (A)−(B)+(C). There is some difficulty in selecting the sizes of the large/small probes. For example, 9 and 3 Å was shown to be a reasonable pair for drug-binding pocket of P-glycoprotein.[50,51] In another example, 10 and 3 Å was a reasonable pair for the ribosomal exit tunnel.[50,52,53] After exploration of alternate probe sizes, 8 and 3 Å seem to be appropriate for DNA polymerases.

The channel volume corresponds to the space available to a nucleotide as it accesses the active site from the surface of the enzyme. This information and the visualization of the overall shape of the channel in various DNA polymerase are useful. First, the channel volumes for binary DNA and ternary substrate complexes for pol β are illustrated in Figure 12. For the ternary complex of pol β, the measured channel volume is 3,338 $Å^3$. We corrected this value for DNA penetrating the channel by 751 $Å^3$; and finally, the volume of the dNTP, 389 $Å^3$, was added so as to estimate the channel volume utilized by the incoming dNTP, 2,976 $Å^3$ (Table 4). This volume, 3,338 $Å^3$ is encompassed by the pink surfaces in Figures 11 and 12. The estimated channel volumes for other families are given in Table 7.

3.4. Function

3.4.1. Influence of Conformational Changes on Active Site Access—From kinetic and structural analysis of binary DNA and ternary (+dNTP) complexes of DNA polymerases from several families, the incoming dNTP interacts with the fingers or N subdomain and is positioned in the active site. The precise positioning of the incoming dNTP induced by the fingers subdomain in the active site is thought to be coupled to a conformational change from an “open” to “closed” form of the DNA polymerase.[25,26] This repositioning relies on rotation of the fingers subdomain ~41° toward the primer-template in T7 (PDB ID 1T7P)[8] and 46° in Klentaq (PDB ID 3KTQ)[6] to achieve the “closed” conformation (A-family members). In contrast, kinetic data suggests that the dNTP undergoes free diffusion through the channel with RB69 (B-family).[54] As tabulated in Table 6, the distance between the center of mass of the bound incoming dNTP and the imaginary plane at the entrance of the channel is ~25 Å for Phi29 DNA polymerase (B-family).

Since many DNA polymerases undergo conformational adjustments upon binding a nucleotide, it is informative to examine the impact of the conformational change on the channel. Where both binary and ternary complex structures are available, we computed the associated volumes for each pair: binary (PDB ID 3EYZ) and ternary (PDB ID 3EZ5) for BF; binary (PDB ID 4KTQ) and ternary (PDB ID 3KTQ) for Klentaq; binary (PDB ID 2PZS) and ternary (PDB ID 2PYJ) for Phi29; binary (PDB ID 3ISB) and ternary (PDB ID 2FMS) for pol β as well as a ternary complex with an active site mismatch (PDB ID 3C2M); and binary (PDB ID 1XSL) and ternary (PDB ID 1XSN)) for pol λ. The channel volumes of the ternary complex structures are shown for representative members of A, B, and X-families (Figure 13A). The channel volumes for ternary complex structures of A and X-families are similar. However, the comparison of channel volumes of the ternary complex structures for B-family members exhibits a wider distribution, 5,937–7,980 $Å^3$.

The ratio of channel volumes for the paired binary and ternary complex structures (V_{bin}/V_{ter}) can be calculated to determine the fractional volume change due to subdomain closing upon nucleotide binding: $F_{vol}=(V_{bin}-V_{ter})/(V_{bin})$ (Figure 13B). This quantity is useful for

understanding the channel response upon binding the incoming dNTP. The measured F_{vol} values were as follows: Phi29 (0.19) ~ Klentaq (0.18) > BF (0.11) > pol λ (0.04). For pol β structures, we also calculated the fractional volume change with an active site mismatched nascent base pair (PDB ID 3C2M) and compared it with that for a correct base pair (PDB ID 2FMS); these were −0.01 and 0.08 respectively. For the ratio of channel volumes (V_{binary}/$V_{ternary}$), Phi29 (1.23) ~ Klentaq (1.21) > BF (1.13) > pol λ (1.04) (Table 4). These data are consistent with the concept that the movement of the dNTP through the channel in members of the B-family occurs primarily by diffusion. The data also indicate that pol β exhibits a small fractional volume change upon binding the correct *or* incorrect nucleotide.

Structures of Klentaq (A-family) transitioning between binary and ternary complexes are illustrated in Figure 14. A large protein backbone change occurs in the vicinity of the incoming nucleotide upon ternary complex formation, presumably trapping the incoming nucleotide (Figure 14). The conformational change results in a large change in channel volume (from 3,220 Å^3 to 2,651 Å^3). Kinetic analysis of PP_i release from T7 DNA polymerase[55] and the Klenow fragment for *E. coli* DNA polymerase I[56] suggests that it dissociates after a post-chemistry conformational change such as subdomain opening. This is also consistent with structural characterization of ternary product complexes recently reported for the X-family pol β; i.e., the PP_i product is only observed in the closed protein conformation. [41]

Finally, the shape of the channel entrance for DNA polymerases from different families is illustrated in Figures 15 (with protein), and the overall shape of the channel is illustrated in Figure 16 (without protein). Figure 16 shows the diverse topologies of the channels for both binary and ternary complexes. These observations are consistent with properties of the subdomain motion. For A- and B-family polymerases, the hinging motion occurs in the vicinity of the templating base, whereas for pol β this occurs in the vicinity of the incoming nucleotide.[25] Accordingly, opening and closing motions with pol β have a smaller impact on the volume near the dNTP-binding pocket. This is consistent with the report that pol λ does not show a significant protein conformational change between the ternary and binary complxes.[15] Thus, conformational changes with pol λ are limited upon binding an incoming nucleotide and the volume of the channel would not be expected to be altered in the transition between binary and ternary complexes.

3.4.2. Proofreading Exonuclease Domain—The impact of the dNTP-induced conformational change on channel volume is different for A- and B-family members. Since these polymerases often include a proofreading exonuclease domain, the efficiency of proofreading could be impacted by the position of this accessory domain relative to the channels described here. The 3′–5′ proofreading exonuclease domain removes incorrect nucleotides that have undergone catalysis resulting in a mismatched primer terminus. For the A-family polymerases, for instance, it has been proposed that a misinsertion could cause instability in the ternary “closed” complex, leading to a quick transition to the “open” conformation.[6] Thus, misinsertion could be coupled to the large conformational change of the DNA polymerase. The 3′–5′ proofreading exonuclease domain of the A-family polymerases is positioned distal to the incoming dNTP access channel (Figure 1). In contrast, this domain forms part of the channel for the B-family polymerases (Phi29, Figure

1; pol δ, Figure 3). Thus, since the exonuclease domain forms a portion of the channel in the B-family polymerases, it might influence substrate and product binding. In this case, the polymerase domain conformational changes may not be a prerequisite for proofreading, as it might be for the A-family. However, it has recently been demonstrated that the stability of the closed complex is dramatically lower after misinserting an incorrect nucleotide with pol β that does not have a proofreading activity.[41] It was postulated that the open conformation would provide an opportunity for an extrinsic proofreading enzyme access to the mismatched primer terminus. Accordingly, it is not unexpected that subdomain motions could also influence proofreading with B-family DNA polymerases.

4. Concluding Remarks

From structural alignments of fifteen DNA polymerases crystal structures across five DNA polymerase families, a common catalytic core in all of these enzymes include the incoming dNTP, two metal ions, two bridging aspartate residues and a third variable acidic residue (aspartate and glutamate). These elements are conserved in structural space across the families, and of special note is the conserved architecture of the triphosphate group of the incoming nucleotide. Charged residues in the vicinity of the catalytic core are conserved among, but not between, polymerase families.

We also observed well-defined dNTP access channels leading into the catalytic core in DNA polymerases of the A-, B-, and X-families. With the A-family polymerases, the distance from the DNA primer terminus to the channel entrance is the shortest among the three families. On the other hand, the A- and B-family polymerases exhibit a large relative change in channel volume upon dNTP binding. In contrast, the distance from the DNA primer terminus to the channel entrance in B-family polymerases is greater than that calculated for A- and X-family members. The smaller polymerase subdomains and solvent accessible active site of the Y-family DNA polymerases suggest that access to their active sites are not restricted for these specialized DNA polymerases. Finally, the channels in the X-family polymerases are not appreciably different in DNA binary or ternary (+dNTP) liganded states. It remains to be determined how trafficking of substrate binding and product release is modulated for the divergent polymerase channels described here.

Acknowledgments

LGP acknowledges NIH grant HL-06350. This research was supported by Research Project Numbers Z01-ES050158 and Z01-ES050161 to S.H.W. by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences and was in association with the National Institutes of Health Grant 1U19CA105010. Structural images were produced using the Chimera package[57] from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIH P41 RR-01081).

References

1. McClure WR, Jovin TM. J Biol Chem. 1975; 250:4073. [PubMed: 1092683]
2. Tanabe K, Bohn EW, Wilson SH. Biochemistry. 1979; 18:3401. [PubMed: 465481]
3. Braithwaite DK, Ito J. Nucleic Acid Res. 1993; 21:787. [PubMed: 8451181]
4. Burgers PMJ, Koonin EV, Bruford E, Blanco L, Burtis KC, Christman MF, Copeland WC, Friedberg EC, Hanaoka F, Hinkle DC, Lawrence CW, Nakanishi M, Ohmori H, Prakash L, Prakash

S, Reynaud CA, Sugino A, Todo T, Wang Z, Weill JC, Woodgate R. J Biol Chem. 2001; 276:43487. [PubMed: 11579108]

5. Beard WA, Wilson SH. Chem Rev. 2006; 106:361. [PubMed: 16464010]
6. Li Y, Korolev S, Waksman G. EMBO J. 1998; 17:7514. [PubMed: 9857206]
7. Golosov AA, Warren JJ, Beese LS, Karplus M. Structure (Camb). 2010; 18:83. [PubMed: 20152155]
8. Doublié S, Tabor S, Long AM, Richardson CC, Ellenberger T. Nature. 1998; 391:251. [PubMed: 9440688]
9. Berman AJ, Kamtekar S, Goodman JL, Lazaro JM, de Vega M, Blanco L, Salas M, Steitz TA. EMBO J. 2007; 26:3494. [PubMed: 17611604]
10. Swan MK, Johnson RE, Prakash L, Prakash S, Aggarwal AK. Nat Struct Mol Biol. 2009; 16:979. [PubMed: 19718023]
11. Wang M, Xia S, Blaha G, Steitz TA, Konigsberg WH, Wang J. Biochemistry. 2011; 50:581. [PubMed: 21158418]
12. Evans RJ, Davies DR, Bullard JM, Christensen J, Green LS, Guiles JW, Pata JD, Ribble WK, Janjic N, Jarvis TC. Proc Natl Acad Sci USA. 2008; 105:20695. [PubMed: 19106298]
13. Wing RA, Bailey S, Steitz TA. J Mol Biol. 2008; 382:859. [PubMed: 18691598]
14. Batra VK, Beard WA, Shock DD, Krahn JM, Pedersen LC, Wilson SH. Structure (Camb). 2006; 14:757. [PubMed: 16615916]
15. Garcia-Diaz M, Bebenek K, Krahn JM, Pedersen LC, Kunkel TA. DNA Repair (Amst). 2007; 6:1333. [PubMed: 17475573]
16. Moon AF, Garcia-Diaz M, Bebenek K, Davis BJ, Zhong X, Ramsden DA, Kunkel TA, Pedersen LC. Nat Struct Mol Biol. 2007; 14:45. [PubMed: 17159995]
17. Wang Y, Musser SK, Saleh S, Marnett LJ, Egli M, Stone MP. Biochemistry. 2008; 47:7322. [PubMed: 18563918]
18. Jain R, Nair DT, Johnson RE, Prakash L, Prakash S, Aggarwal AK. Structure (Camb). 2009; 17:974. [PubMed: 19604477]
19. Vasquez-Del Carpio R, Silverstein TD, Lone S, Swan MK, Choudhury JR, Johnson RE, Prakash S, Prakash L, Aggarwal AK. PLoS ONE. 2009; 4:e5766. [PubMed: 19492058]
20. Biertümpfel C, Zhao Y, Kondo Y, Ramón-Maiques S, Gregory M, Lee JY, Masutani C, Lehmann AR, Hanaoka F, Yang W. Nature. 2010; 465:1044. [PubMed: 20577208]
21. Ollis DL, Brick P, Hamlin R, Xuong NG, Steitz TA. Nature. 1985; 313:762. [PubMed: 3883192]
22. Beard WA, Shock DD, Yang XP, DeLauder SF, Wilson SH. J Biol Chem. 2002; 277:8235. [PubMed: 11756435]
23. Kresge N, Simoni RD, Hill RL. J Biol Chem. 2005; 280:e46.
24. Lehman IR, Bessman MJ, Simms ES, Kornberg A. J Biol Chem. 1958; 233:163. [PubMed: 13563462]
25. Beard WA, Wilson SH. Structure (Camb). 2003; 11:489. [PubMed: 12737815]
26. Doublié S, Sawaya MR, Ellenberger T. Structure (Camb). 1999; 7:R31. [PubMed: 10368292]
27. Steitz TA. J Biol Chem. 1999; 274:17395. [PubMed: 10364165]
28. Sawaya MR, Pelletier H, Kumar A, Wilson SH, Kraut J. Science. 1994; 264:1930. [PubMed: 7516581]
29. Steitz TA, Smerdon SJ, Jager J, Joyce CM. Science. 1994; 266:2022. [PubMed: 7528445]
30. Beard WA, Wilson SH. Mutat Res. 2000; 460:231. [PubMed: 10946231]
31. Russell RB, Barton GJ. Proteins: Struct, Funct, Genet. 1992; 14:309. [PubMed: 1409577]
32. Humphrey W, Dalke A, Schulten K. J Mol Graph. 1996; 14:33. [PubMed: 8744570]
33. Eastwood MP, Hardin C, Luthey-Schulten Z, Wolynes PG. IBM J Res Dev. 2001; 45:475.
34. Garcia-Diaz M, Bebenek K, Gao G, Pedersen LC, London RE, Kunkel TA. DNA Repair (Amst). 2005; 4:1358. [PubMed: 16213194]
35. Prasad R, Beard WA, Chyan JY, Maciejewski MW, Mullen GP, Wilson SH. J Biol Chem. 1998; 273:11121. [PubMed: 9556598]

36. Moon AF, Garcia-Diaz M, Batra VK, Beard WA, Bebenek K, Kunkel TA, Wilson SH, Pedersen LC. DNA Repair (Amst). 2007; 6:1709. [PubMed: 17631059]
37. Prasad R, Beard WA, Wilson SH. J Biol Chem. 1994; 269:18096. [PubMed: 8027071]
38. Sawaya MR, Prasad P, Wilson SH, Kraut J, Pelletier H. Biochemistry. 1997; 36:11205. [PubMed: 9287163]
39. Rothwell PJ, Waksman G. Adv Protein Chem. 2005; 71:401. [PubMed: 16230118]
40. Nakamura T, Zhao Y, Yamagata Y, Hua Y-j, Yang W. Nature. 2012; 487:196. [PubMed: 22785315]
41. Freudenthal BD, Beard WA, Shock DD, Wilson SH. Cell. 2013; 154:157. [PubMed: 23827680]
42. Lin P, Pedersen LC, Batra VK, Beard WA, Wilson SH, Pedersen LG. Proc Natl Acad Sci USA. 2006; 103:13294. [PubMed: 16938895]
43. Freudenthal BD, Beard WA, Wilson SH. Structure (Camb). 2012; 20:1829. [PubMed: 22959623]
44. Beard WA, Shock DD, Vande Berg BJ, Wilson SH. J Biol Chem. 2002; 277:47393. [PubMed: 12370169]
45. Batra VK, Perera L, Lin P, Shock DD, Beard WA, Pedersen LC, Pedersen LG, Wilson SH. J Am Chem Soc. 2013; 135:8078. [PubMed: 23647366]
46. Beard WA, Shock DD, Batra VK, Pedersen LC, Wilson SH. J Biol Chem. 2009; 284:31680. [PubMed: 19759017]
47. Garcia-Diaz M, Bebenek K, Krahn JM, Kunkel TA, Pedersen LC. Nat Struct Mol Biol. 2005; 12:97. [PubMed: 15608652]
48. Batra VK, Beard WA, Shock DD, Pedersen LC, Wilson SH. Mol Cell. 2008; 30:315. [PubMed: 18471977]
49. Batada NN, Westover KD, Bushnell DA, Levitt M, Kornberg RD. Proc Natl Acad Sci USA. 2004; 101:17361. [PubMed: 15574497]
50. Voss NR, Gerstein M. Nucleic Acids Res. 2010; 38:555.
51. Aller SG, Yu J, Ward A, Weng Y, Chittaboina S, Zhuo R, Harrell PM, Trinh YT, Zhang Q, Urbatsch IL, Chang G. Science. 2009; 323:1718. [PubMed: 19325113]
52. Frank J, Zhu J, Penczek P, Li Y, Srivastava S, Verschoor A, Radermacher M, Grassucci R, Lata RK, Agrawal RK. Nature. 1995; 376:441. [PubMed: 7630422]
53. Voss NR, Gerstein M, Steitz TA, Moore PB. J Mol Biol. 2006; 360:893. [PubMed: 16784753]
54. Yang G, Franklin M, Li J, Lin TC, Konigsberg W. Biochemistry. 2002; 41:2526. [PubMed: 11851399]
55. Patel SS, Wong I, Johnson KA. Biochemistry. 1991; 30:511. [PubMed: 1846298]
56. Dahlberg ME, Benkovic SJ. Biochemistry. 1991; 30:4835. [PubMed: 1645180]
57. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. J Comput Chem. 2004; 25:1605. [PubMed: 15264254]

Biographies

Sangwook Wu received a B. A. in Biochemistry from Yonsei University (Korea) in 1990. After working as a Scientist at Samsung Display Devices (1995–1999), he obtained his Ph.D. in Theoretical Condensed Matter Physics (Theory of Glass Transition) from Iowa State University in 2005. From 2005 to the present, he has been a research associate at UNC-CH, first in the laboratory of Garegin Papoian and more recently in the laboratory of Lee Pedersen, focusing on the computational dynamics of biological macromolecules.

William A. Beard received a Ph.D. from Purdue University in 1986 where he studied the coupling of electrochemical gradients with ATP formation. He did postdoctoral training at St. Jude Children's Research Hospital characterizing wild-type and mutant derivatives of human dihydrofolate reductase by steady-state, pre-steady-state, and transient-state kinetic methods. In 1991, he joined Samuel H. Wilson at the National Cancer Institute, NIH, to study the enzymology of HIV-1 reverse transcriptase and DNA repair enzymes, primarily mammalian DNA polymerase β. Through a combination of site-directed mutagenesis, kinetic analysis, molecular modeling and structure determination, these systems serve as ideal models for uncovering basic catalytic mechanisms and strategies that influence genome stability. He is currently a Staff Scientist at the National Institute of Environmental Health Sciences, NIH.

Lee Pedersen is a Professor of Physical Chemistry at UNC-CH. He received his B.Ch. from the University of Tulsa in 1961 and Ph.D. from the University of Arkansas in 1965. After work at Columbia University (NSF Postdoc '65–'66) and Harvard University (NIH Postdoc '66–'67), he joined the faculty of chemistry at UNC-CH. His research has spanned quantum chemical theory and applications to dynamical theory and applications on biological systems. He retired from teaching in 2006 as M. A. Smith Professor, but continues to maintain an active research program supported by NIH. Since 1986, he has maintained a fruitful association with the nearby National Institute of Environmental Health Sciences.

Samuel H. Wilson received a bachelor's degree from the University of Denver and did his graduate and postdoctoral training in medicine and biochemistry, respectively, at Harvard Medical School and the NIH, and conducted two years of graduate training in organic synthesis at the University of Denver Research Institute. He began his career in 1970 at the National Cancer Institute (NCI), National Institutes of Health (NIH) and remained there until late 1991 when he moved to the University of Texas, Galveston, to found a center

focused in the areas of DNA repair, structural biology, and functional genomics. He moved to the National Institute of Environmental Health Sciences (NIEHS), NIH, in 1996. As head of the DNA Repair and Nucleic Acid Enzymology Section in the Laboratory of Structural Biology, he has had a long-term research interest in mammalian DNA metabolism. Over the past several years, he and his associates have focused their efforts on the mammalian base excision DNA repair pathway and the role of DNA polymerase β. He has authored and co-authored many research publications and has been editor of four reference volumes. His recent activities include membership on numerous advisory groups involved in biological and medical research. He was chair of the Mammalian DNA Repair Gordon Research Conference and co-chair of the biannual US-Japan and US-EU international conferences on DNA Repair. He is Editor-in-Chief of the journal *DNA Repair* and is a member of several editorial boards.

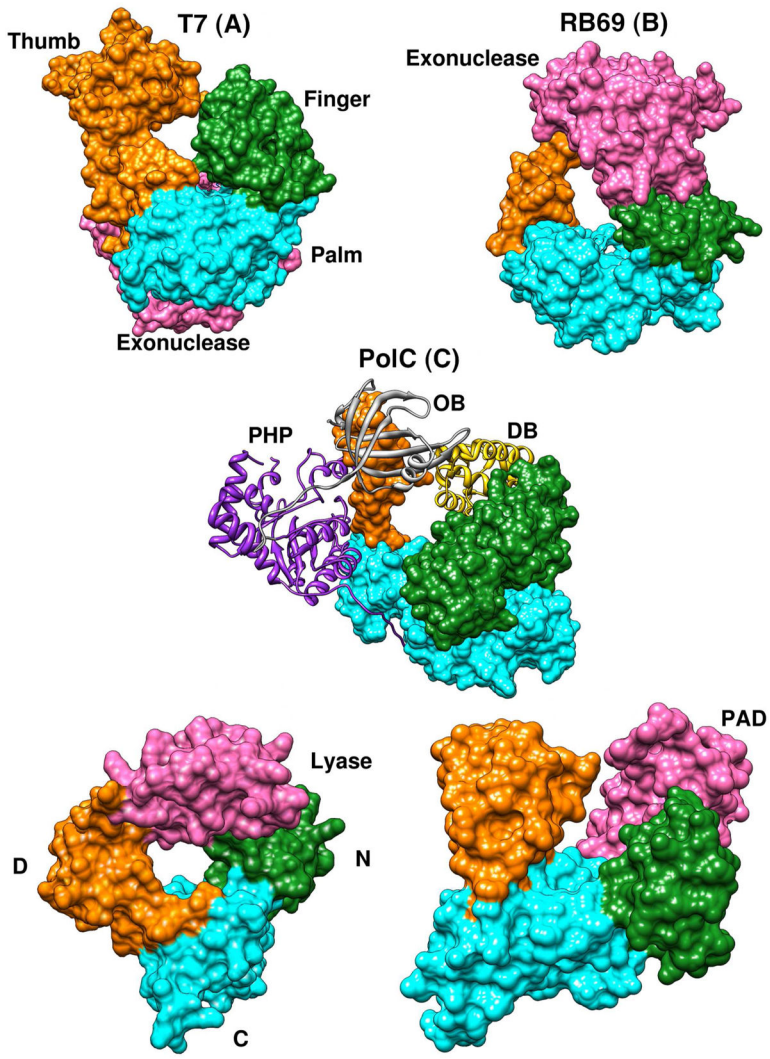

Figure 1.

DNA polymerase domain and subdomain organization. The polymerase domain of representative family members are colored according to subdomain. The DNA polymerases of the A- (T7), B- (Phi29), and Y-families (pol ι) are colored according to an architectural analogy to a right-hand: thumb, orange; palm, light blue; and fingers, green. The C- (PolC) and X-family (pol β) DNA polymerase subdomains are colored similarly but their nomenclature often uses a functional designation: D (DNA binding), orange; C (catalytic), light blue; and N (nucleotide selection), green. The accessory domains are colored pink (e.g., exonuclease, lyase, or PAD). Also indicated in a ribbon representation are the PHP (polymerase and histidinol phosphate domain), DB (duplex binding domain) and OB (oligonucleotide/oligosaccharide-binding domain). The exonuclease domain of T7 and Phi29 DNA polymerases are positioned differently with respect to the polymerase active site. For pol ι, the PAD (polymerase associated domain) lies near the catalytic core. The DNA is omitted for clarity, but enters the active site from the back of the structures in this viewpoint. A tubular channel where nucleotides must enter the active site is indicated in the pol β structure. The PDB ID codes for the illustrated structures are given in Table 1.

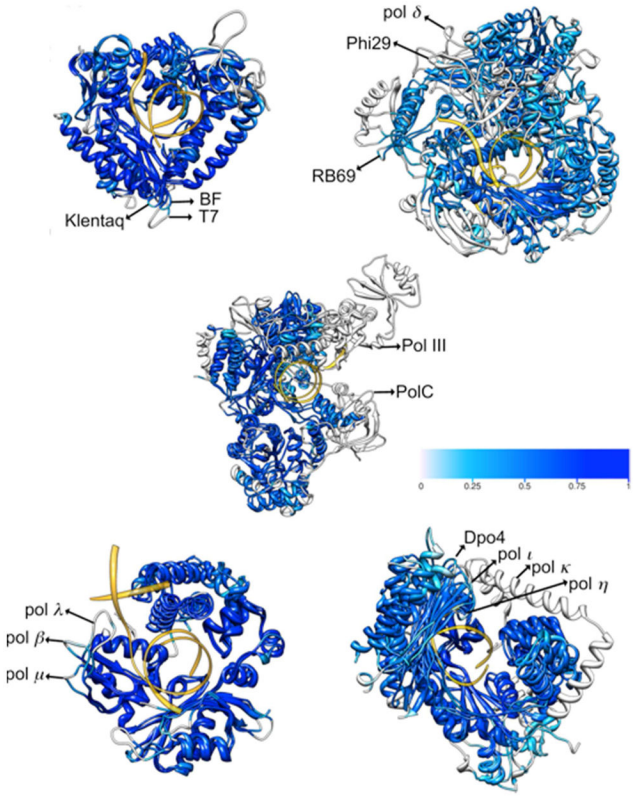

Figure 2.

DNA polymerase family-dependent structure conservation. Ribbon representation of crystallographic ternary complex structures of alternate DNA polymerase families after multiple structural alignments (MSA). The viewpoint is opposite to that shown in Figure 1 (i.e., rotated 180º) so that the view is down the duplex DNA (semi-transparent gold ribbons). Accordingly, the fingers, or N, subdomain is on the left in each structure (opposite to that shown in the Figure 1) while the palm, or C, subdomain is positioned at the bottom of each structure. The ribbons are colored according to the calculated MSA for each C_α. A color key illustrating backbone structural conservation is shown: dark blue as the MSA value approaches 1 (i.e., conserved), and unique regions for each structure are colored white (i.e., MSA = 0). The PDB ID ID codes for the illustrated structures are given in Table 1.

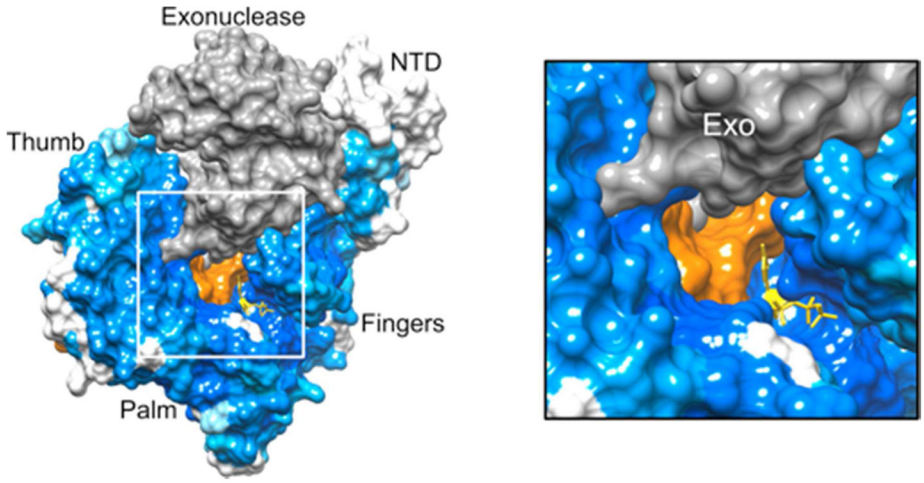

Figure 3.

Surface representation of pol δ (B-family) colored according to MSA. The 3′–5′-exonuclease domain (residues 316–531; gray surface) sterically restricts and defines the dNTP access channel for this DNA polymerase. A surface representation of the DNA is colored orange. The domains and polymerase subdomains are indicated. The DNA duplex penetrates the channel and occludes nucleotide access from the opposite side in this viewpoint. The image on the right shows a detailed view of the vicinity of the channel (outlined in white in the global view). The incoming nucleotide (yellow stick representation) is sitting in the active site at the end of the channel. The PDB ID code is 3IAY.[10]

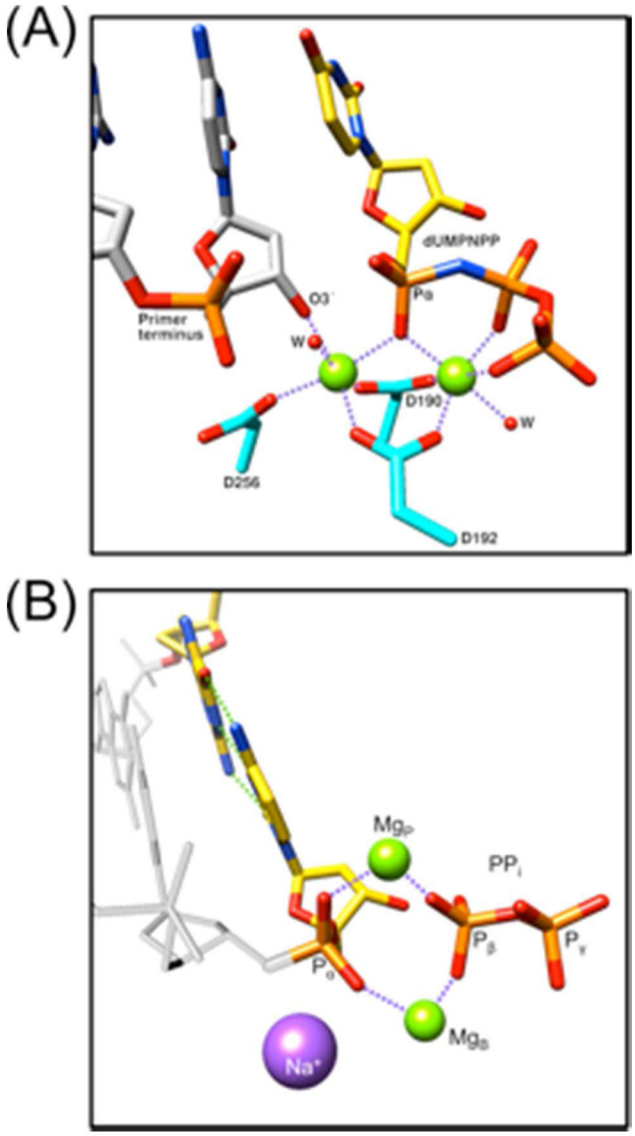

Figure 4.

DNA polymerase β active site. (A) Two divalent metal ions (green, Mg^{2+}) are octahedrally coordinated by substrates (non-bridging oxygens of each phosphate of the incoming dNTP and DNA primer terminus O3′), two water molecules (W) and three acidic enzyme side chains (Asp190, Asp192, and Asp256; D190, D192, and D256, respectively). A non-hydrolyzable analog dUMPNPP, (2′-deoxyuridine-5′-(α,β)-imido triphosphates; yellow carbons), was used to trap the ternary substrate pre-catalytic complex (PDB ID 2FMS).[14] The catalytic metal that coordinates O3′ of the primer terminus induces a 3′-endo sugar pucker at the primer terminus. The α-phosphate (P_α) of the incoming nucleotide is indicated. (B) Crystallographic characterization of intermediates during correct nucleotide insertion has identified a product associated divalent metal (Mg_P) that coordinates oxygens on phosphates of the product molecules (nascent primer terminus and PP_i; PDB ID 4KLG).[41] Four water molecules complete the octahedral coordination for this transient

metal (not shown). The nucleotide associated divalent metal B has lost its coordination with P_γ as PP_i begins to dissociate. A sodium ion (Na^+) is observed occupying the metal A site in this product ternary complex. In this view, the preceding base pair (previous primer terminus) is light gray.

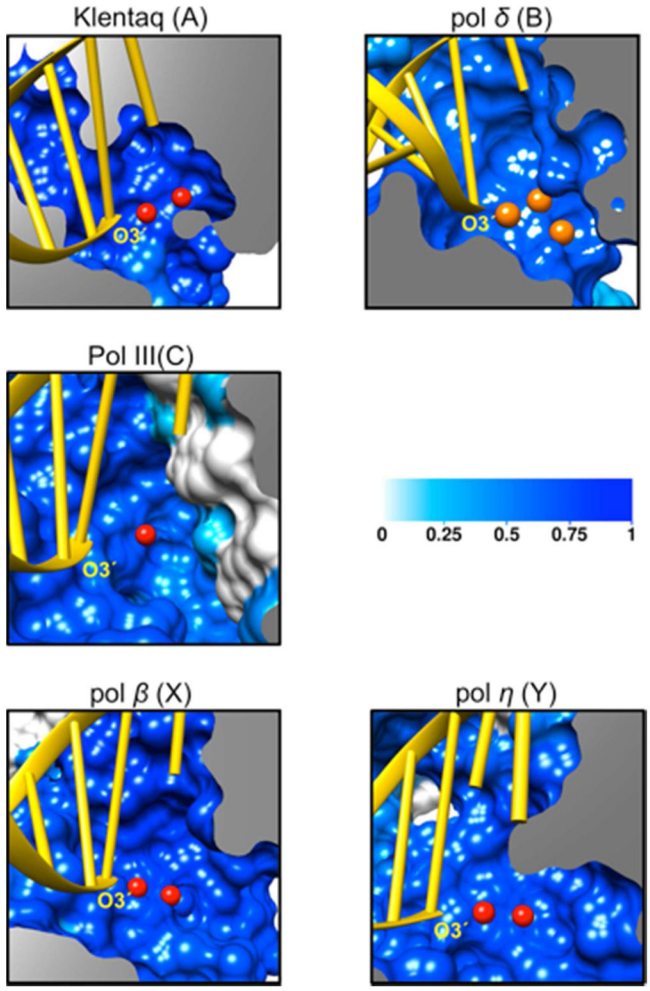

Figure 5.

DNA polymerase family-dependent active site conservation. Surface representation of crystallographic ternary complex structures of DNA polymerases from alternate families after multiple structural alignments (MSA). The surfaces are colored according to the calculated MSA for each C_α. A color key illustrating backbone structural conservation is shown: dark blue as the MSA value approaches 1 (i.e., conserved), and unique regions for each structure are colored white (i.e., MSA = 0). The focused view is on the polymerase active site and indicates that the active sites are highly conserved (i.e., colored dark blue). The 3′-end of the primer terminus (O3′) in each structure is indicated. The red and orange spheres represent the active site divalent metals (red, Mg^{2+}; orange, Ca^{2+}). The gray surface represents the clipping plane necessary to view the active site. The PDB ID codes for the illustrated structures are given in Table 1.

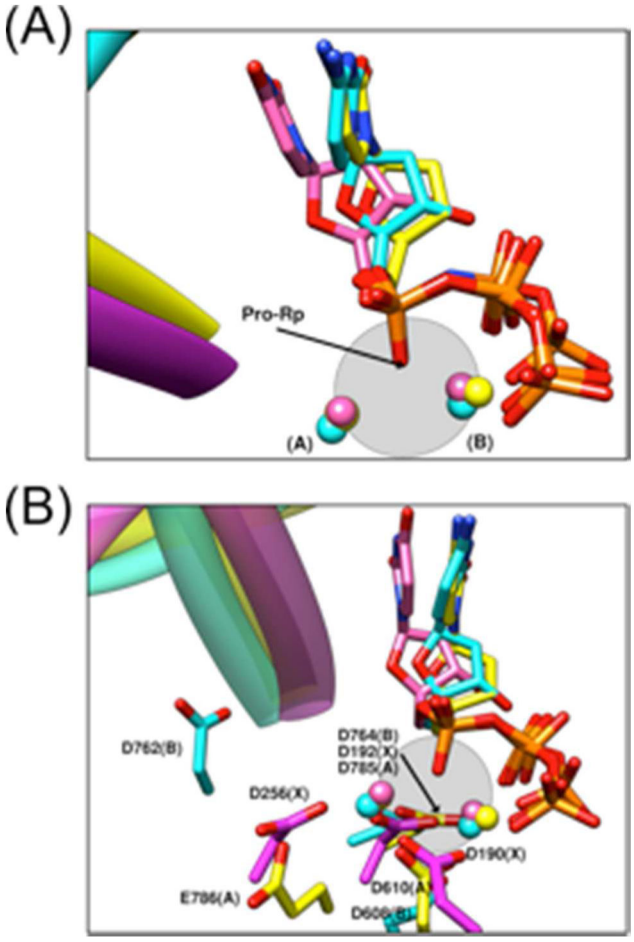

Figure 6.

Overlay of the catalytic core of three DNA polymerases from different families. (A) Klentaq (A-family, yellow), pol δ (B-family, cyan), and pol β (X-family X, pink) were superimposed with respect to the R_p-oxygen of the a-phosphate of the incoming dNTP and the two active site metal ions (A and B). These three atoms form an imaginary plane (semi-transparent gray circle). The oxygens of the α-, β-, and γ-phosphates of the incoming dNTPs and metals align well. (B) DNA polymerase active site acidic residues. The oxygens of two aspartates (Asp608, Asp764; pol δ) are hidden and not clearly seen in this view. Three residues, Asp785 (Klentaq), Asp764 (pol δ), and Asp192 (pol β), are well conserved behind the plane. The PDB ID codes for the illustrated structures are given in Table 1.

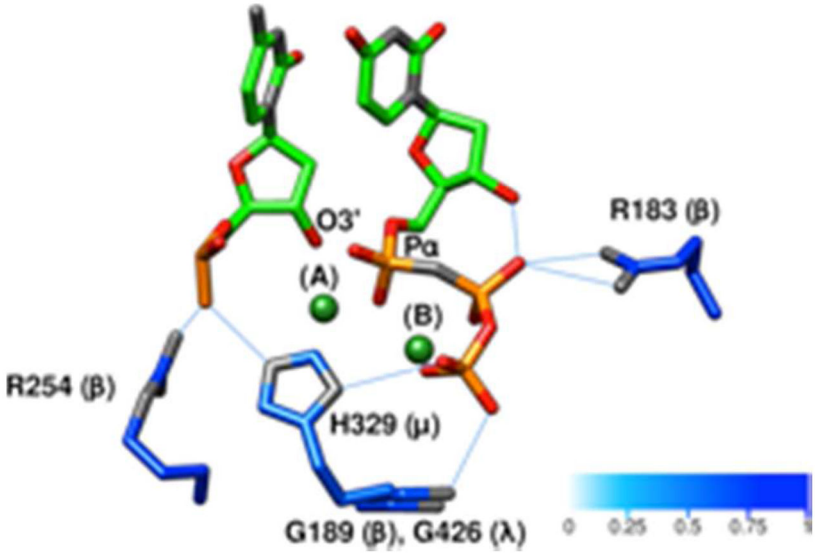

Figure 7.

Overlay of the catalytic core of three X-family DNA polymerases. Two conserved active site arginines are shown (183 and 254 in pol β, R183 and R254, respectively). Only the metal ions of pol β (dark green spheres, Mg^{2+}) are shown. Protein carbon atoms are colored according to their C_α Q_{res} values (0–1), and oxygen and nitrogen protein atoms are colored red and gray, respectively. His329 (H329) of pol μ is also shown corresponding to Gly189 (G189) and Gly426 (G426) of pol β and pol λ, respectively. This histidine hydrogen bonds to phosphate oxygens on the primer-terminus and γ-phosphate of the incoming nucleotides facilitating template-independent DNA synthesis.[16] This histidine is also observed in terminal deoxynucleotide transferase, another template-independent DNA polymerase. The primer-terminus O3′ and P_α of the incoming nucleotide of pol β are also indicated. The PDB ID codes for the illustrated structures are given in Table 1.

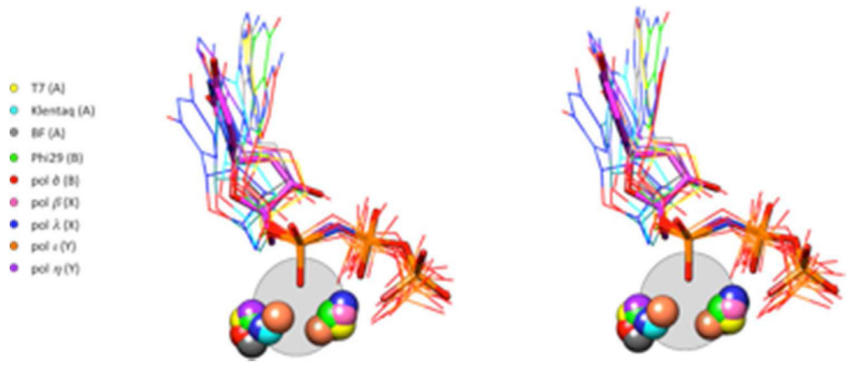

Figure 8.

Stereo view of an overlay of the catalytic core of ternary complex DNA polymerase structures from different families (A, T7, Klentaq, and BF; B, Phi29 and pol δ; X, pol β and pol λ; Y, pol ι and pol η). The nine structures (dNTP and metal ions) are aligned with respect to the pro-R_P oxygen of the α-phosphate of the incoming dNTP of pol β (thick pink carbon bonds), and the two metal ions. The plane defined by these three atoms is shown as reference (semi-transparent gray disk). The PDB ID codes for the illustrated structures are given in Table 1.

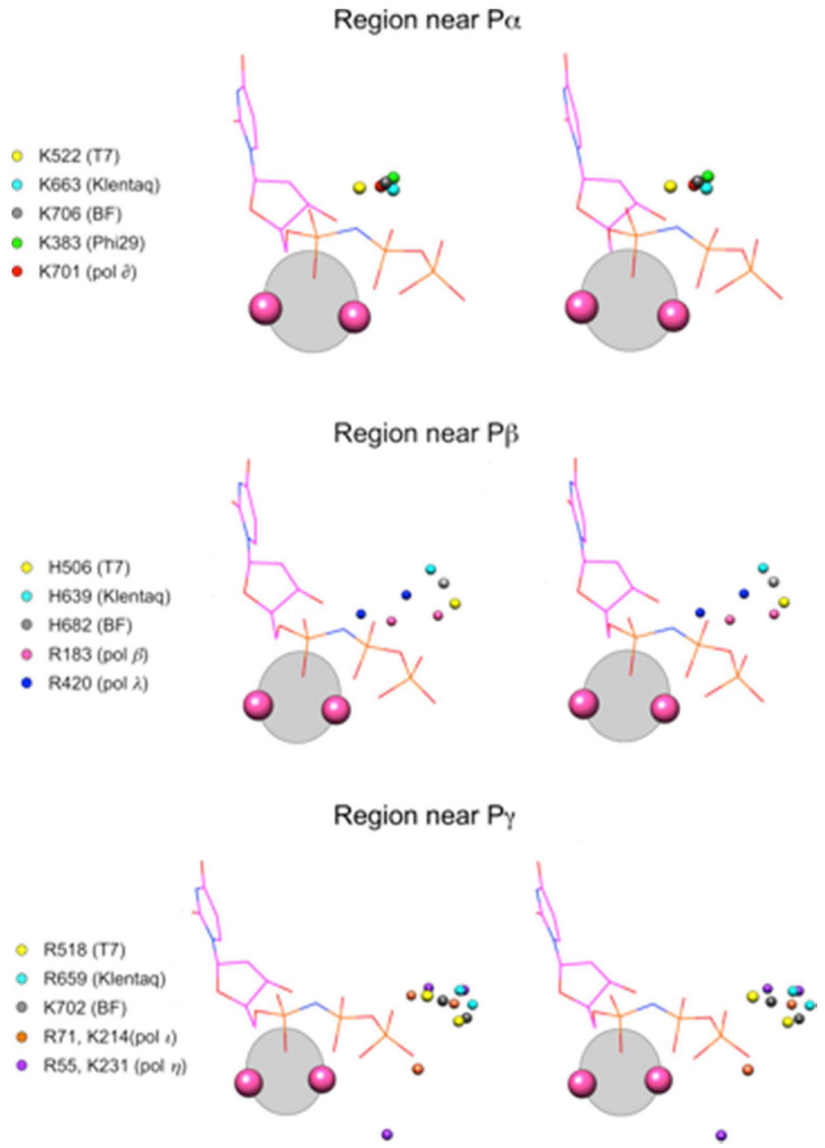

Figure 9.

Stereo view of an overlay of DNA polymerase ternary complex structures illustrating potential hydrogen bonding donors (arginine, lysine, and histidine) in the vicinity of the phosphates of the incoming nucleotide for polymerases from various families. The imaginary plane (semi-transparent gray), active site metals, and incoming nucleotide from pol β are shown as reference. These atoms are within 5 Å of the corresponding phosphates. The PDB ID codes for the illustrated structures are given in Table 1.

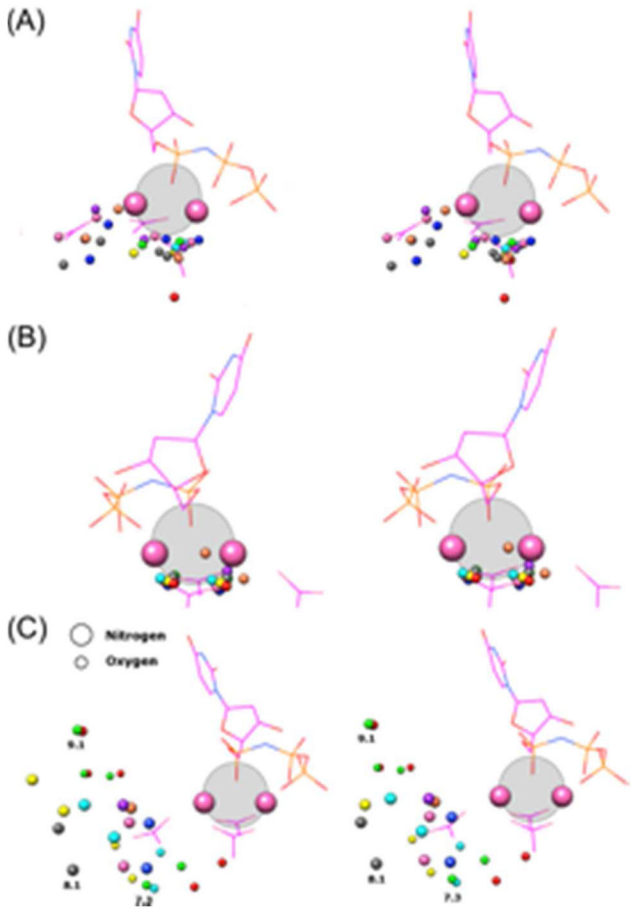

Figure 10.

Stereo view of an overlay of DNA polymerase ternary complex structures illustrating atoms in the vicinity of the active site metals ligands. The atoms are colored according to the identity of the DNA polymerase as illustrated in Figures 8 and 9. The imaginary plane (semi-transparent gray), active site metals, and incoming nucleotide from pol β are shown as reference. Acidic side chain oxygens are shown that are within 4.93 Å of metal A (panel A) and 2.67 Å of metal B (panel B). Since the catalytic metal site is not always occupied, a liberal cutoff threshold was chosen to identify potential coordinating acidic side chains for metal A. Since site B is always occupied with a divalent metal, a more conservative threshold was employed. (C) Side chain nitrogen (large spheres) and oxygen (small spheres) atoms within 9.1 Å of metal A are shown (excluding metal liagnds). The PDB ID codes for the illustrated structures are given in Table 1. The distances denoted are measured from the catalytic ion (metal A).

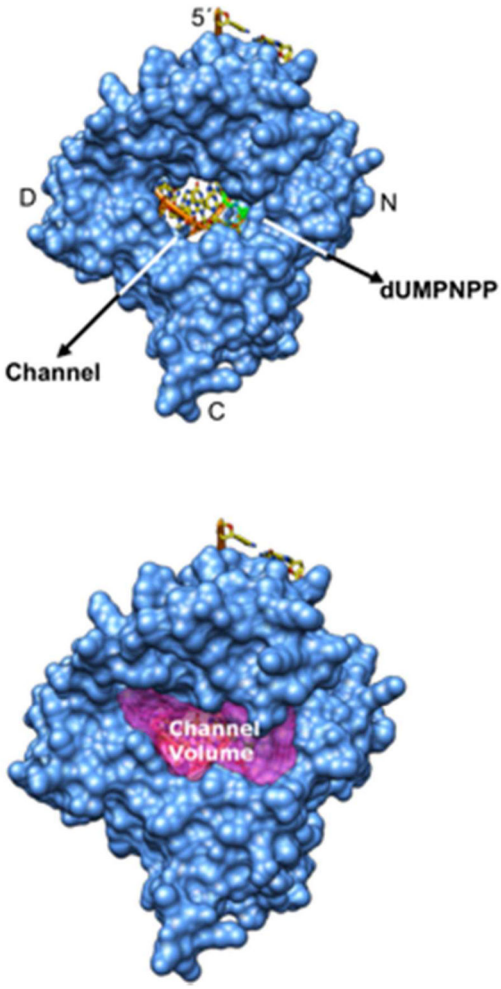

Figure 11.

Nucleotide access channel for pol β (X-family; PDB ID 2FMS).[14] The view is the similar to that in Figure 1 (DNA enters the active site from the opposite side in this view). The top panel illustrates the donut-like protein architecture creating a path that the incoming nucleotide must diffuse to access the active site. The 5′-terminus of the template strand is indicated, as well as the relative position of the C-, D-, and N-subdomains of the polymerase domain. The green incoming nucleotide is indicated. The bottom panel superimposes the surface of the calculated channel (pink) estimated by 3V algorithm software.[50]

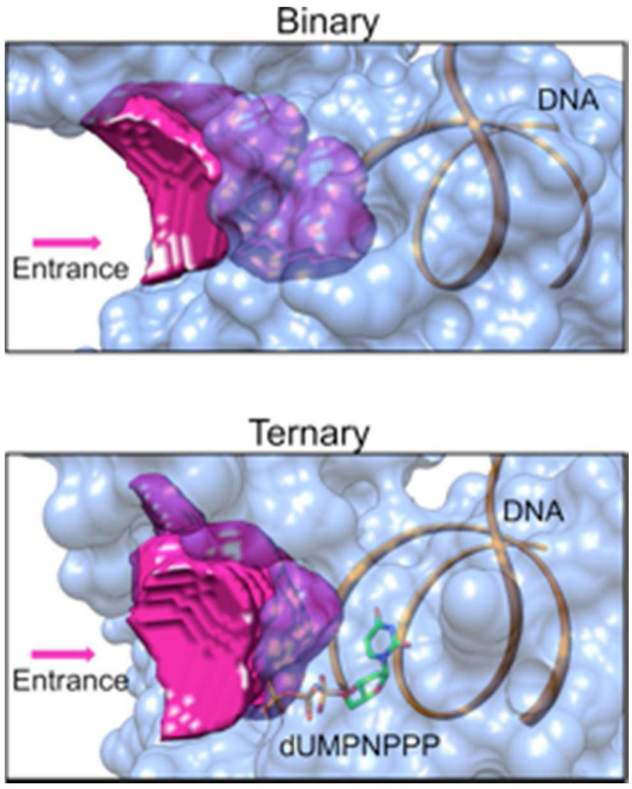

Figure 12.

Binary and ternary pol β complex channels. The volume of the apparent channel leading to the active site is calculated for the binary (PDB ID 3ISB, top panel)[46] and ternary (PDB ID 2FMS, bottom panel)[14] substrate complex structures of pol β. It is defined by estimating the volume of the channel in the absence of dNTP and DNA strands (A) minus the volume of dNTP and DNA strands which overlap with the channel (B) and adding back the volume of free dNTP (C); (A)−(B)+(C). The dNTP is part of the channel in the ternary substrate complex. Structurally, the surface of the enclosed calculated channel volume is colored pink. The surface of the polymerase is a semi-transparent light blue. The dNTP is shown in a stick representation (green carbons) in the ternary substrate complex. The volumes of the channels are estimated employing the two-rolling sphere method (8 and 3 Å radii).[50]

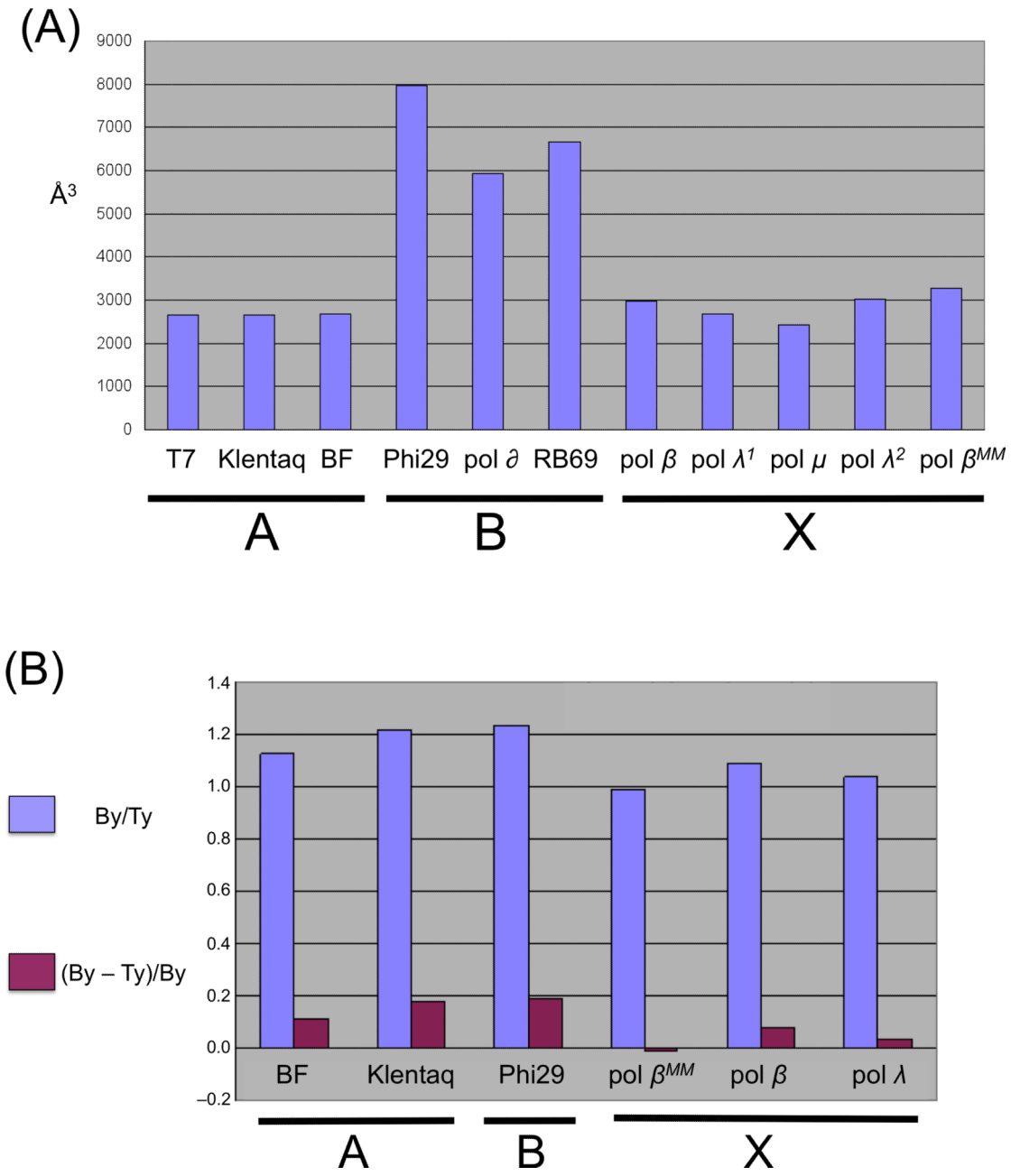

Figure 13.

(A) The calculated channel volumes for ternary substrate complex structures from various families. Additionally, the calculated channel volume for a ternary complex structure of pol β with an active site mismatch (dAMPCPP–dG, incoming nucleotide–templating nucleotide; pol β^{MM}), and pol λ with one (pol λ^1) or two (pol λ^2) active site metals are also shown. (B) The ratio of the channel volumes calculated for binary (By) DNA complexes relative to that determined upon formation of the ternary (Ty) complex (i.e., +dNTP). Relative volume difference between the binary DNA complex of pol β and the mismatched ternary complex (pol β^{MM}) is also shown. The PDB ID codes for the structures used in these calculations are given in Tables 1 and 4.

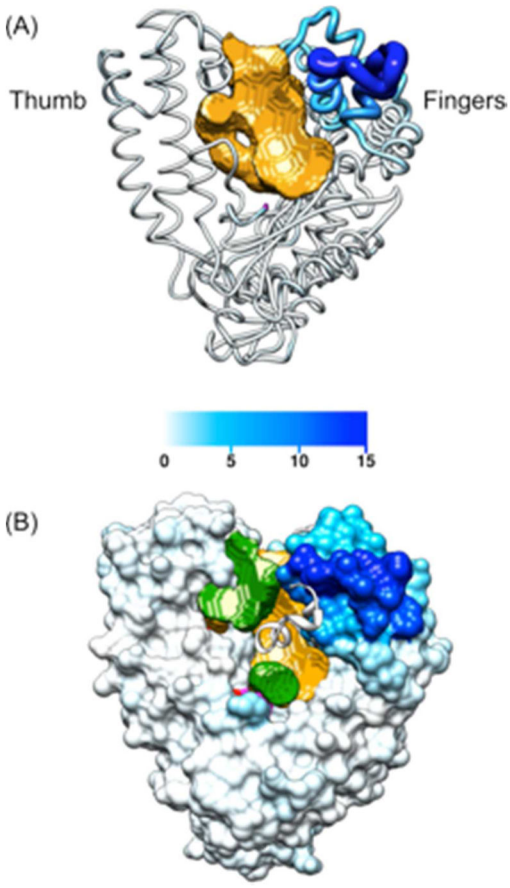

Figure 14.

(A) Altered backbone conformation of an A-family DNA polymerase (Klentaq) upon binding a nucleotide. The polymerase backbone of the binary DNA complex (PDB ID 4KTQ) is shown in a worm representation. The backbone is colored according to the C_α displacement when binding an incoming dNTP (PDB ID 3KTQ).[6] A color scale for the C_α displacement (Å) is shown below the structure. Additionally, the diameter of the worm is proportional to the magnitude of backbone displacement. The largest conformational adjustment upon binding the incoming nucleotide is in the Fingers subdomain. The gold surface encompasses the estimated channel volume of the binary complex employing two-rolling sphere method.[50] (B) A surface representation of the binary DNA complex colored according to the backbone displacement described above is superimposed with a ribbon representation of the ternary complex. The viewpoint is the same as that shown in panel A. Only a short portion of α-helix O (ternary complex) that repositions itself upon binding a nucleotide is visible. The calculated channels for the binary (gold) and ternary complexes (green) are also shown.

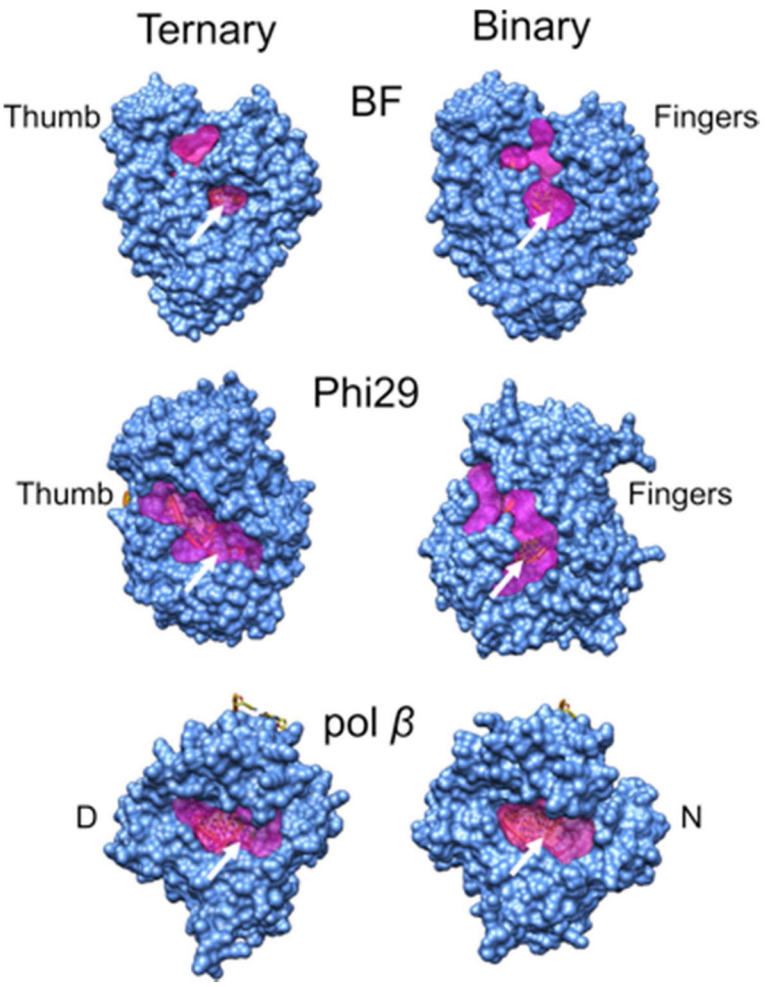

Figure 15.

Altered channel topology when transitioning from the binary DNA complex to a ternary complex. The primer terminus in each structure is indicated (white arrow) and the subdomains responsible for DNA binding (Thumb or D-subdomain) and dNTP binding (Fingers or N-subdomain) are indicated. The channel volume (semi-transparent pink surface) of each complex was calculated employing a two-rolling sphere method.[50] The PDB ID codes for the ternary and binary complexes are given in Tables 1 and 4, respectively.

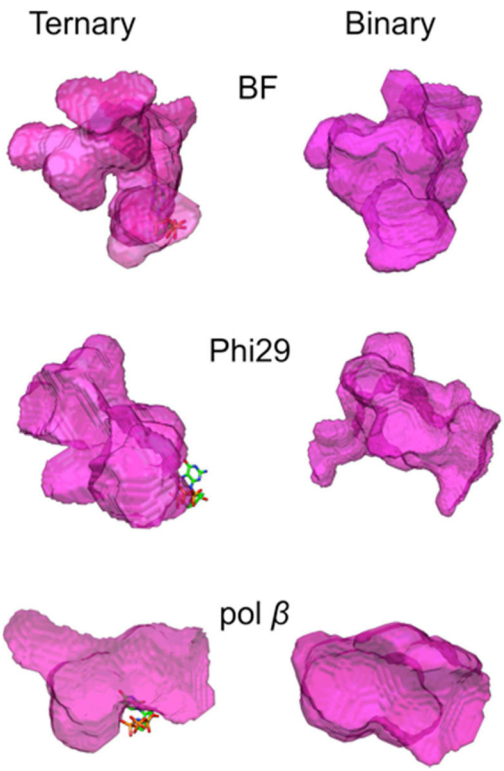

Figure 16.

Altered channel topology when transitioning from the binary DNA complex to a ternary complex. The channel volume (semi-transparent pink surface) of each complex, as viewed in Figure 16, is shown without the polymerase. The incoming nucleotide of the ternary complexes is shown in a stick representation.

TABLE 1

Crystallographic Ternary Complex DNA Polymerase Structures Analyzed in this Study

family	PDB ID	ref	biological function	polymerase	accessory domain[a]	abbreviation
A	3KTQ	6	Repair Replication	*Thermus aquaticus* DNA polymerase I, large fragment	N-terminal	Klentaq
	3EZ5	7	Repair Replication	*Bacillus stearothermophilus* DNA polymerase I, large fragment	3′–5′ Exonuclease	BF
	1T7P	8	Replication	T7 DNA polymerase[b]	3′–5′ Exonuclease	T7
B	2PYJ	9	Replication	Phi29 DNA polymerase	3′–5′ Exonuclease	Phi29
	3IAY	10	Replication	DNA polymerase δ	3′–5′ Exonuclease	pol δ
	3NCI	11	Replication	RB69 gp43 DNA polymerase	3′–5′ Exonuclease	RB69
C	3F2B	12	Replication	*Geobacillius kaustophilus* PolC	3′–5′ Exonuclease	PolC
	3E0D	13	Replication	*Thermus aquaticus* DNA Polymerase III α-subunit	CTD, β-binding, PHP	pol III
X	2FMS	14	Repair	Human DNA polymerase β	Lyase	pol β
	2PFO	15	Repair	Human DNA polymerase λ	Lyase	pol λ
	2IHM	16	Repair	Mouse DNA polymerase μ	8-kDa	pol μ
Y	2R8H	17	Translesion synthesis	*Sulfolobus solfataricus* DNA polymerase IV (Dpo4)	Little finger	Dpo4
	3H4D	18	Translesion synthesis	Human DNA polymerase ι	PAD	pol ι
	3IN5	19	Translesion synthesis	Human DNA polymerase κ	PAD, N-Clasp	pol κ
	3MR2	20	Translesion synthesis	Human DNA polymerase η	PAD	pol η

[a]These accessory domains are associated with the polymerase domain in the indicated structures. Since some of the structures represent truncated proteins, the list is not complete.

[b]The T7 DNA polymerase structure includes its processivity factor thioredoxin. Since it is ~50 Å from the polymerase active site, it is not included in the channel calculations.

TABLE 2

Conserved Position of Critical Constituents of the Active Site of X- family DNA Polymerases[a]

PDB ID (pol)	catalytic metal (A)			nucleotide metal (B)			pro-R_P oxygen (dNTP)		
	X	y	z	x	y	z	x	y	z
2FMS (β)	0.00	0.20	0.19	−0.02	−0.07	0.31	0.04	−0.11	0.31
2PFO (λ)	0.56	−0.16	−0.02	0.38	−0.31	0.09	0.64	−0.39	0.07
2IHM (μ)	−0.55	−0.05	−0.18	−0.39	0.39	−0.40	−0.69	0.51	−0.37
Ave.[b]	3.40	7.42	17.16	2.44	3.94	16.43	3.96	5.33	16.10
S.D.	0.56	0.18	0.19	0.41	0.36	0.36	0.67	0.47	0.34

[a]The values represent the absolute distance (Å) in coordinate space (x,y,z) of the respective atoms from the average position of all three X-family polymerases after MSA (Ave. – atom(x,y, or z)).

[b]The absolute coordinate position (x,y,z) of the respective atom average of the three X-family polymerases.

TABLE 3

Conserved Position of Critical Constituents of the Active Site of A-family DNA Polymerases[a]

PDB ID (pol)	catalytic metal (A)			nucleotide metal (B)			pro-R_P oxygen (dNTP)		
	x	y	z	x	y	z	x	y	z
1T7P (T7)	0.27	0.39	−0.43	0.07	0.51	−0.05	0.25	0.48	−0.15
3KTQ (Klentaq)	−0.28	−0.02	−0.15	−0.17	−0.05	−0.14	−0.49	−0.10	−0.10
3EZ5 (BF)	0.02	−0.38	0.59	0.11	−0.48	0.20	0.23	−0.38	0.24
Ave.[b]	47.41	26.19	−5.65	44.84	23.50	−5.76	45.24	25.49	−4.72
S.D.	0.39	0.39	0.52	0.15	0.50	0.18	0.42	0.44	0.21

[a]The values represent the absolute distance (Å) in coordinate space (x,y,z) of the respective atoms from the average position of all three A-family polymerases after MSA (Ave. – atom(x,y, or z)).

[b]The absolute coordinate position (x,y,z) of the respective atom average of the three A-family polymerases.

TABLE 4

General DNA Polymerase Characteristics

family	PDB ID	mol. mass (kDa)	dNTP	active site ions (A/B)	channel volume[a] (Å^3)	conserved residues[b]	organism
I. Ternary Complex Structures							
A	1T7P	90.8	ddGTP	Mg/Mg	2651	R429, R452, D475, E480, R518, K522, H506, D654, E655	bacteriophage T7
	3KTQ	60.9	ddCTP	Mg/Mg	2651	R573, R595, D610, E615, R659, K663, H639, D785, E786	*Thermus aquaticus*
	3EZ5	66.2	ddATP	Zn/Zn	2675[c]	R615, R637, D653, E658, H682, R702, K706, D830, E831	*Bacillus stearothermophilus*
B	2PYJ	66.7	dGTP	Mg/Mn	7980	D249, K383, D456, D458, E486, K498	bacteriophage Phi29
	3IAY	104.5	dCTP	Ca/Ca[d]	5937	D608, K701, D762, D764, E802, K814	yeast (pol δ)
	3NCI	104.7	dCTP	Ca/Ca	6668	D411, K560, D621, D623, E686, K706	bacteriophage RB69
C	3F2B	117.2	dGTP	–/Mg	NO[e]	R893, D973, D975, K1096, D1098	*Geobacillus kasustophilus*
	3E0D	137.6	dATP	–/Ca	NO	R424, D463, D465, K616, D618	Eubacteria
X	2FMS	38.2	dUMPNPP	Mg/Mg	2976	R183, D190, D192, R254, D256	human (pol β)
	2PFO	37.4	dUMPNPP	Mn/Mg	2688	R420, D427, D429, R488, D490	human (pol λ)
	2IHM	40.6	ddTTP	Na/Mg	2430	R323, D330, D332, R418, D420	human (pol μ)
Y	2R8H	40.3	dGTP	Ca/Ca	NO	D7, R51, D105, E106, K152, K159	*Sulfolobus solfataticus*
	3H4D	44.0	dGTP	Mg/Mg	NO	D34, R71, D126, E127, K207, K214	human (pol ι)
	3IN5	58.0	dATP	–/Mg	NO	D107, R144, D198, E199, K321, K328	human (pol κ)
	3MR2	48.6	dAMPNPP	Mg/Mg	NO	D13, R55, D115, E116, K224, K231	human (pol η)

II. Binary Complex Structures

family	DNA polymerase	PDB ID	ref	$F_{vol.}$[f]
A	BF	3EYZ	7	0.11
	Klentaq	4KTQ	6	0.18
B	Phi29	2PZS	9	0.19
X	β	3ISB	46	0.08
	β	3ISB	46	–0.01[g]
	λ	1XSL	47	0.04

[a] The volumes of the channel were estimated employing the two-rolling sphere method (8 and 3 Å radii).

[b] Active site aspartates that coordinate both active site metals are structurally conserved across all families (underlined).

[c] The volume of the channel was estimated employing the two-rolling sphere method (9 and 3 Å radii).

[d] An additional Ca^{2+} ion is observed 2.7 Å from an oxygen on the γ-phosphate of the incoming nucleotide.

[e] Not observed.

[f] $F_{vol.} = Volume_{binary} - Volume_{ternary}/Volume_{binary}$.

[g] A ternary complex structure with an active site mismatch (dAMPCPP–dG, incoming nucleotide–templating nucleotide; PDB ID 3C2M)[48] used in calculation.

TABLE 5

Residues in Crystal Structures of Ternary Substrate Complexes (C) Relative to that of Full-length Enzyme (T)

polymerase[a] (family)	C	T	C/T	domains[b]
Klentaq (A)	539	540	0.998	Pol, Exo
Phi29 (B)	567	575	0.986	Pol, Exo
PolC (C)[c]	1006	1455	0.691	Pol
pol β (X)	326	335	0.973	Pol, Lyase
pol ι (Y)	373	390	0.956	Pol

[a] The PDB ID codes are given in Table 1.

[b] Polymerase domain, Pol; Exonuclease Domain, Exo.

[c] Among 1455 residues, amino-terminal residues (1–232) and the 3′–5′ proofreading exonuclease domain (412–617) were removed from the engineered construct.

TABLE 6

Distance (Å) from DNA Primer Terminus or Incoming Nucleotide to the Entrance of the Channel

DNA polymerase (family)[a]	C3′–entrance[b]	dNTP–entrance[c]
Klentaq (A)	12.5	10.8
Phi29 (B)	22.4	24.9
pol β (X)	13.3	13.1

[a] The PDB ID codes for the ternary complexes are given in Table 1.

[b] Distances calculated from C3′ of the primer terminus to the entrance of the channel defined by an imaginary plane (see text).

[c] Distances calculated from the center of mass of the incoming dNTP to the entrance of the channel.

Table 7

Calculated Channel Volumes for Crystallographic Structures from Various DNA Polymerase Families

polymerase (family), PDB ID	total channel (A)	DNA + dNTP[b] (B)	dNTP (C)	channel (A)−(B)+(C)
ternary complexes				
T7 (A), 1T7P	3633	1406	424	2651
Klentaq (A), 3KTQ	3546	1347	452	2651
BF (A), 3EZ5	4085	1839	429	2675
Phi29 (B), 2PYJ	11246	3719	453	7980
δ (B), 3IAY	8028	2484	393	5937
RB69 (B), 3NCI	7595	1374	447	6668
β (X), 2FMS	3338	751	389	2976
λ (X), 2PFO	3131	835	392	2688
μ (X), 2IHM	2737	712	405	2430
λ (X), 1XSN[c]	3331	715	401	3017
β (X), 3C2M[d]	3647	812	433	3268
binary DNA complexes				
BF (A), 3EYZ	4991	1979	–	3012
Klentaq (A), 4KTQ	3734	514	–	3220
Phi29 (B), 2PZS	12820	2985	–	9835
β (X), 3ISB	3232	–	–	3232
λ (X), 1XSL	3470	341	–	3129

[a] Units are $Å^3$.

[b] Volume that penetrates the channel (A).

[c] Ternary complex structure with a single bound Mg^{2+} in site B.

[d] Ternary complex structure with an active site mismatch (dAMPCPP–dG; incoming nucleotide-templating nucleotide).