DNA Polymerases

Structure, Function, and Modeling

3

Tamar Schlick

---

1. Introduction

1.1 *Venerable task*

The transmission of our instruction book for life — the DNA genome — from one generation to the next describes both similarities to our parents and differences that make us unique. How this duplication occurs and how this process leads to differences via mutations is a fundamental problem in both basic and applied research. As Watson and Crick observed in their landmark 1953 *Nature* work reporting the structure of the DNA, the double-helical DNA structure lends itself naturally to template-directed replication. Many DNA polymerases are responsible for this venerable task of linking nucleotides in a rapid process that uses each DNA strand as a template for another duplex DNA nearly identical to the original.

Deviations from identity, however, occur when errors are introduced in this copying process — nucleotide deletion, insertion, inversion, or translocation — as well as damage from external sources like radiation or hazard chemicals, including cigarette smoke and environmental pollutants. Such mutations are handled both by proofreading activities of the replicative DNA polymerase themselves as well as specialized cellular repair processes involving a variety of DNA polymerases with expertise in handling specific error types: base excision repair, mismatch repair, chemical adduct removal, double-strand break repair, and many others. Because any error that is not promptly repaired can lead to many human diseases such as skin abnormalities and various cancers, significant research effort has been expended into understanding polymerase replication and repair mechanisms at the atomic level.

Computer modeling plays an important role in these endeavors by helping link static molecular views from X-ray crystallography and NMR with macroscopic kinetic data that measure polymerase efficiency and fidelity. See Refs. 1 and 2 for exceptional works linking crystal forms via a free-energy pathway of dynamic configurational changes. A detailed atomic and molecular-level understanding of how DNA polymerases operate can help suggest possible therapeutic approaches to enhance polymerase fidelity, identify factors that degrade polymerase performance,

and assist in the development of pharmacological approaches to diseases such as cancer, heart disease, diabetes, and a variety of neurological conditions that are associated with DNA damage and incorrect DNA repair.

1.2 *Replication and repair fidelity*

In replication, DNA polymerases orchestrate the addition of new nucleotides to a growing chain of DNA by catalyzing a nucleotidyl transfer reaction which increases the primer strand by one base pair. The relative ability of a polymerase to incorporate a *correct* nucleotide rather than an *incorrect* unit from a pool of structurally similar molecules is a measure of its *fidelity*. DNA polymerases catalyze the same nucleotidyl transfer reaction in which an incoming nucleotide unit termed dNTP (2′-deoxyribonucleoside 5′-triphosphate) is inserted at the 3′ end of a growing DNA primer strand (Figure 1). They must select the correct dNTP unit to pair with the templating base (e.g. dGTP opposite template C and other Watson-Crick base pair combinations) from a pool of structurally similar nucleotides (e.g. dATP, dCTP, and dTTP). Intriguingly, the accuracies exhibited by proofreading-deficient DNA polymerases from different families span a wide range of error rates (see Table 1), from 10^{-6} to near unity, namely one error per 10^6 nucleotides incorporated (for high-fidelity pol α) to nearly one error per nucleotide for low-fidelity pol X. The fidelity of a polymerase is defined as the inverse of the misinsertion frequency (see Table 1 footnote).

For recent reviews, see Refs. 3 and 4.

1.3 *Polymerase structure and function*

DNA polymerases are shaped like a hand with fingers, palm and thumb subdomains[10] (Figure 2). (Spartan pol X lacks the fingers subdomain). The fingers and thumb subdomains are associated
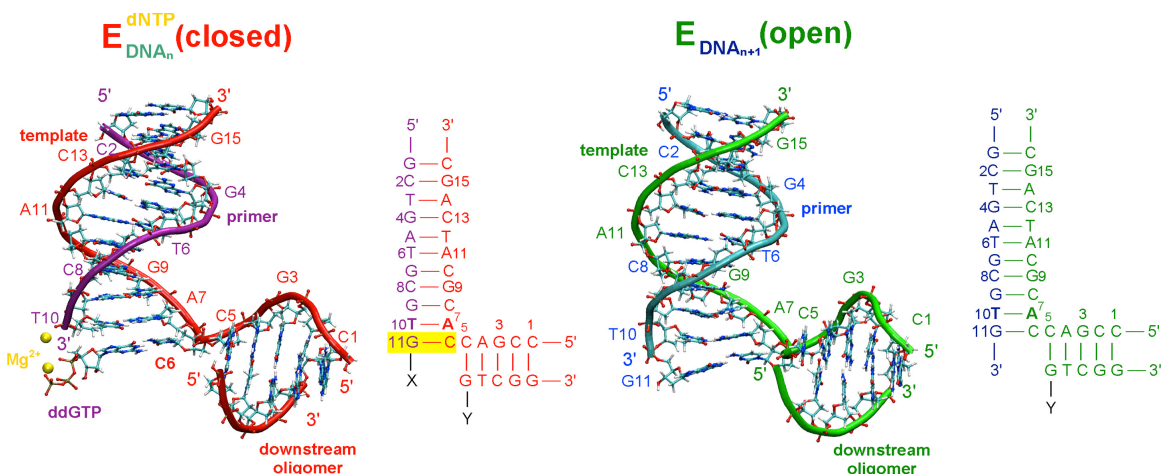

Figure 1. Extension of a growing DNA primer by one nucleotide. Left, correct incoming G opposite template C6. Right, incoming nucleotide covalently bound to primer. See Figure 3 for pathway context of the two DNA shown here within the polymerase complex.

Table 1.   Kinetic data for DNA Polymerases.

Enzymes (family)	Fidelity[a]	$k_{pol}(s^{-1})$[b]
Klentaq1 (A)	10^4	20
HIV-1 RT (RT)	10^4	70
Pol β (X)	10^3	10
Pol λ (X)	10^3	4
Pol μ (X)	10^3	0.08
Pol X (X)	~1	0.8
Dpo4 (Y)	10^2	0.3

[a] Fidelity = reciprocal of misinsertion error frequency = $[(k_{pol}/K_d)_c + (k_{pol}/K_d)_i] / (k_{pol}/K_d)_i$, where c and i denote correct and incorrect nucleotide incorporation, and K_d is the apparent equilibrium dissociation constant of dNTP.

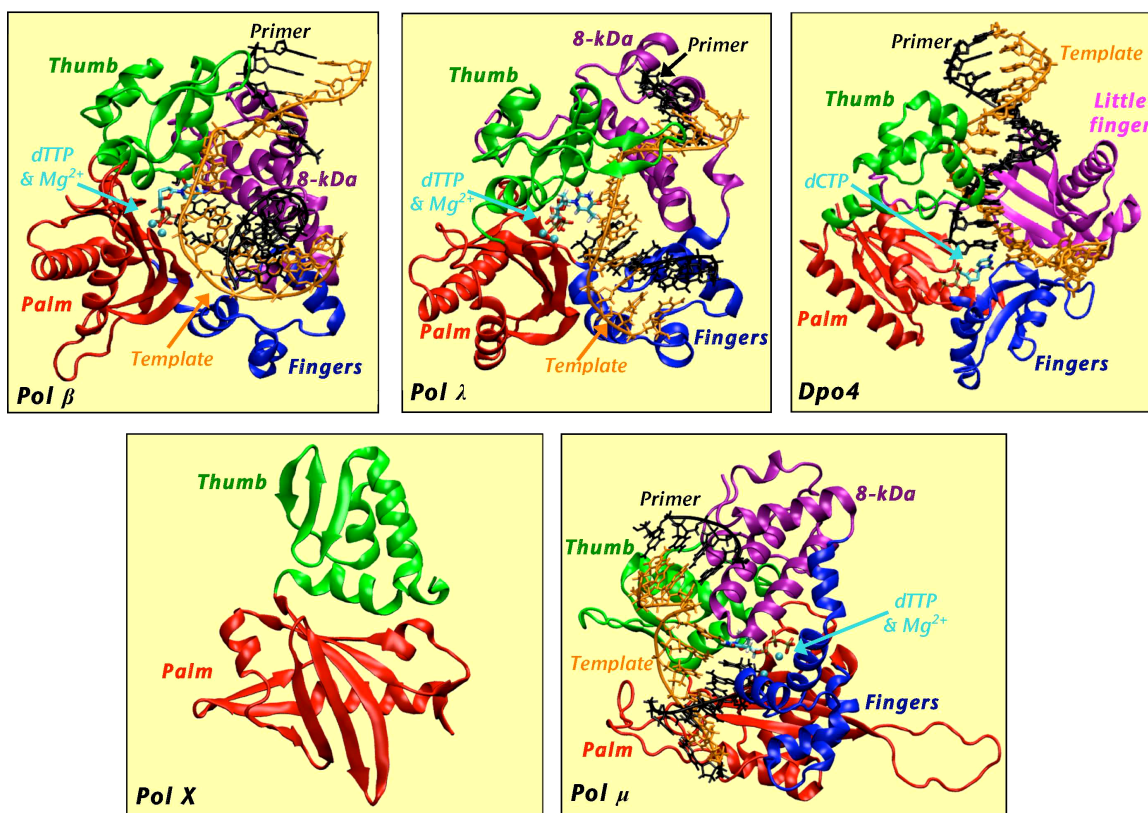[b] k_{pol}: rate of nucleotide incorporation for first-enzyme turnover.

Figure 2.   DNA polymerases structures: pol β,[5] pol λ,[6] Dpo4,[7] pol X,[8] and pol μ.[9]

with grasping the DNA and the incoming nucleotide unit, while the palm subdomain is the site of the chemical reaction which increases the DNA chain by one base pair.

Many families of DNA polymerases are recognized (A, B, C, D, X, Y, and reverse transcriptases), including eukaryotic DNA polymerases α, δ, ε in the B family, bacterial

polymerase III in the C family, Y-family members η, ι, and κ, and X-family members β, λ, μ, X, and Tdt.

In this article, we focus on several pol X family members (pol β, λ, μ, and X) and one pol Y member (Dpo4) (Figure 2) which we have studied by computer modeling. Each of these family-X enzymes has intriguing characteristics. Pol β is a repair enzyme with moderate fidelity that functions primarily in base excision repair (BER); it is thought to have a tight active site guided by the *induced-fit* mechanism for base-pair recognition (see below). Pol X is a highly error prone polymerase that considers five base pairs intrinsically correct (the four Watson-Crick base pairs plus G:G). Pol λ is a low to moderate fidelity enzyme that plays a back-up role to pol β in BER and possibly has a role in non-homologous end joining (NHEJ), a pathway by which double-stranded DNA breaks are repaired; it has a unique propensity to generate frame-shift errors through single-base deletions rather than base substitutions. Intriguingly, pol λ also has an inherent ability to use substrates with minimal base-pairing homology and can easily tolerate extra-helical nucleotides. Pol μ is also mainly associated with the NHEJ repair process; like pol λ, it can handle a gapped template, but unlike this relative it can also handle a primer terminus unpaired with the template. Dpo4 is a low-fidelity enzyme that is associated with lesion bypass. Both pol λ and pol μ also have an N-terminal BRCT domain (named after the C-terminal domain of the breast cancer suppressor protein BRCA1) for mediating interactions with other proteins involved in NHEJ. All five enzymes lack a proofreading domain for editing DNA replication mistakes.

Many high-resolution X-ray crystallographic structures of enzyme/DNA complexes are available for these DNA polymerase. A common induced-fit mechanism has been proposed to operate in their kinetic cycles (Figure 3) based on key crystallographic structures.[11] As illustrated by the superimposed open and closed forms of pol β in Figure 4, the binding of the dNTP substrate induces a conformational change from open binary to closed ternary complex; the chemical reaction of adding another base pair to the DNA then follows (Figure 1), after-which the enzyme returns to the open state and is ready for another cycle of polymerization. In particular, an induced-fit mechanism of the enzyme is thought to direct the opening and closing motion of the *thumb* subdomain in pol β (*fingers* in right-handed polymerases) so as to guide the correct selection of the template residue. Details of this process are, however, unknown at atomic level.

While similar structures are available for pol λ,[6] more limited structural information about the reaction pathways of pol μ and pol X is available. For pol μ, only a ternary closed pol μ/DNA/ddTTP complex has been resolved,[9] and for pol X, NMR complexes without the substrate are available.[12,13]

However, it has become apparent in recent years that not all polymerases have well defined open and closed states. Moreover, even when it exists, such a conformational rearrangement may involve any combination of protein and DNA motion. The reasons for such preferences remain unclear, but motions are thought to be related to polymerase function and error profile. Alternative frameworks to interpret fidelity, such as conformational sampling in which a larger range of conformational can help select correct substrates, are also actively being explored.

Besides structural information as mentioned above, a wealth of biochemical and kinetic data for these enzymes is also available, including kinetic and error rate data from many sources, such as shown in Table 1.

(a)

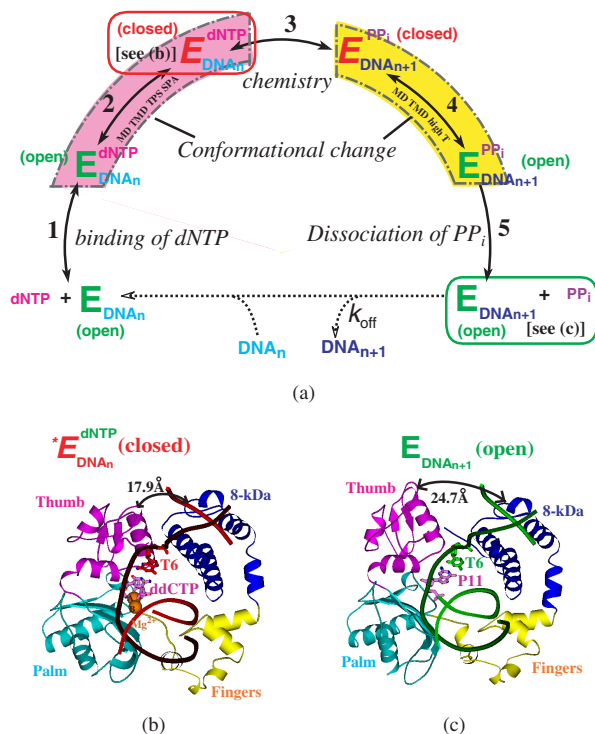(b)   (c)

Figure 3. Proposed DNA polymerase pathway for nucleotide insertion (a), with anchoring crystallographic pol β conformations in (b) closed (red) and (c) open (green) states. E: DNA polymerase; dNTP: 2′-deoxyribonucleoside 5′-triphosphate; PP$_i$: pyrophosphate; DNA$_n$ and DNA$_{n+1}$: DNA before and after the nucleotide incorporation to the DNA primer strand. Adapted with permission from "L. Yang, K. Arora, W. A. Beard, S. H. Wilson and T. Schlick. 2004. Critical role of magnesium ions in DNA polymerase β's closing and active site assembly. *J Am Chem Soc* **126**, 8441–8453." Copyright (2004) American Chemical Society.
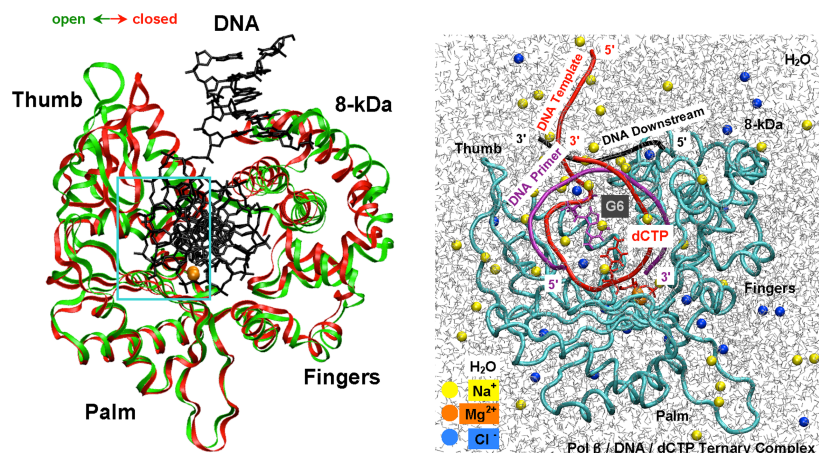

Figure 4. Pol β Structure and Modeling. Overview of the pol β ternary (closed, red) and binary (open, green) states as resolved by crystallography (left) and the solvated, modeled ternary complex (right). Adapted from Journal of Molecular Biology, 317/5, L. Yang, W. A. Beard, S. H. Wilson, S. Broyde and T. Schlick, Polymerase β simulations suggest that Arg258 rotation is a slow step rather than large subdomain motion per se., 651–671, Copyright (2002), with permission from Elsevier.

2. Modeling Studies

2.1 *Need for modeling*

Molecular modeling and simulation are critically needed to interpret and extend experimental findings to provide geometric, thermodynamic, and dynamic explanations that link observed structures to pathways and mechanisms in the goal of interpreting biological function. In particular, it remains a challenge to establish the sequences of motions involved in polymerase cycles and define the sequence of subdomain and residue rearrangements as well as define the rate-limiting step.

Modeling and simulation, while still subject to many approximations and inherent sampling limitations, have nonetheless demonstrated success in many important applications[14–16] from protein folding to RNA structure prediction to membrane dynamics; see Ref. 16 for a recent field perspective and specific examples. With increasing improvements in force fields and algorithms and the growing availability of high-speed computing platforms, modeling and simulation are well on their way to becoming full partner with experiment, if not a field on its own right.[16] Below, a sampling of results from our group's studies on DNA polymerases mechanisms is described.

2.2 *Conformational pathways in pol β, pol λ, pol X, and pol μ*

Dynamics simulations have lended support to the induced-fit mechanism for pol β[17]and pol X[8], in which the correct incoming base triggers the requisite conformational change while an incorrect incoming nucleotide hampers the process (Figure 5). Moreover, studies revealed a cascade of subtle side-chain conformational events following thumb closing but prior to chemistry (Figure 6), including a slow and possibly "gate-keeping" rearrangement of Arg258[1,18–21] that likely directs the catalytic cycle of pol β. This directed sequence enhances the polymerase's selectivity for the correct incoming partner to the template base by allowing an incorrect incoming nucleoside residue to dissociate.[1,22] Significantly, the slow step involving Arg258 has been verified experimentally.[23]

For pol X, which lacks the fingers domain of pol β, we have also shown that the two available NMR structures interconvert to one another at physiological salt.[8] For pol λ, simulations suggest that, besides motions in the thumb subdomain and active-site residues (Ile492, Tyr505, Phe506, Arg517), the DNA undergoes significant correlated motions upon binding the correct nucleotide, when the ions bind too to the primer/template terminus.[24] Dpo4 also utilizes more subtle movements of its little finger and fingers subdomains to bypass small oxidative lesions like 8-oxoguanine.[25]

Recent modeling work on pol μ suggests that the enzyme may not have a well defined open state.[26] The enzyme tends to remain in the closed state even when a substrate is removed from the active site or when the complex is forced to an open-like state. Like its relatives, pol μ exhibits subtle residue rearrangements, with His329, Asp330, Gln440, and Glu443 playing key roles in nucleotide accommodation. The DNA/protein contact environment also appears fragile in this enzyme.

For pol λ, the motions identified above of the primer DNA may be relevant to the inherent tendency for base pair deletions because subtle protein/DNA interactions regulate pol λ stability and hence its error propensity (see below). Thus, though pol λ does not demonstrate large-scale
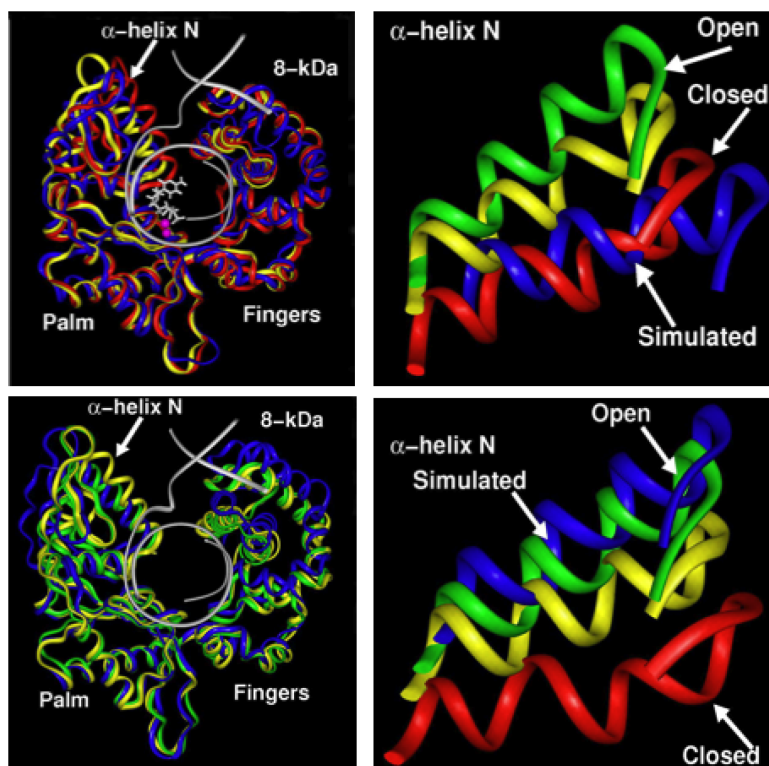
Figure 5.   *In silico* evidence for pol β's induced fit mechanism C$_\alpha$ traces of superimposed pol β/DNA complex with dCTP (top left) and without dCTP (bottom left) for the intermediate starting structure (yellow), crystal closed (red), and crystal open (green) and the trajectory final structures (blue).[17] Notable are the residue motions in the thumb subdmain and the 8-kDa domain. The positions of α-helix N in the simulated systems are comparied to the crystal structures in panels on the right (top, with dCTP, and bottom, without dCTP). Adapted from Biophysical Journal, 87/5, K. Arora and T. Schlick, *In silico* evidence for DNA polymerase β's substrate-induced conformational change, 3088–3099, Copyright (2004), with permission from Elsevier.

subdomain movements as pol β, the above correlated motions may also serve as gate-keepers by controlling the evolution of the reaction pathway. For pol μ, the lack of well defined transtions upon binding the substrate may be related to its ability to handle unpaired primer termini. The significant DNA motions in polymerase cycles for both pol λ[24] and Dpo4[25] may suggest a common feature for moderate-to-low fidelity polymerases. See Table 2 for a summary.

2.3 *Pol β's closing pathway*

Our extensive studies on mammalian pol β based on pioneering crystal complexes solved in various forms by the Wilson group[11] have revealed key aspects of the conformational (Figures 6–7) and chemical pathways (Figure 8) of the enzyme and suggested several factors responsible for fidelity discrimination.[27–29] These biological pathways for both correct and incorrect base-pair systems were resolved by developing the enhanced sampling approach *transition path sampling*[30] for biomolecules[1] and an efficient free-energy approach termed *BOLAS*.[31]
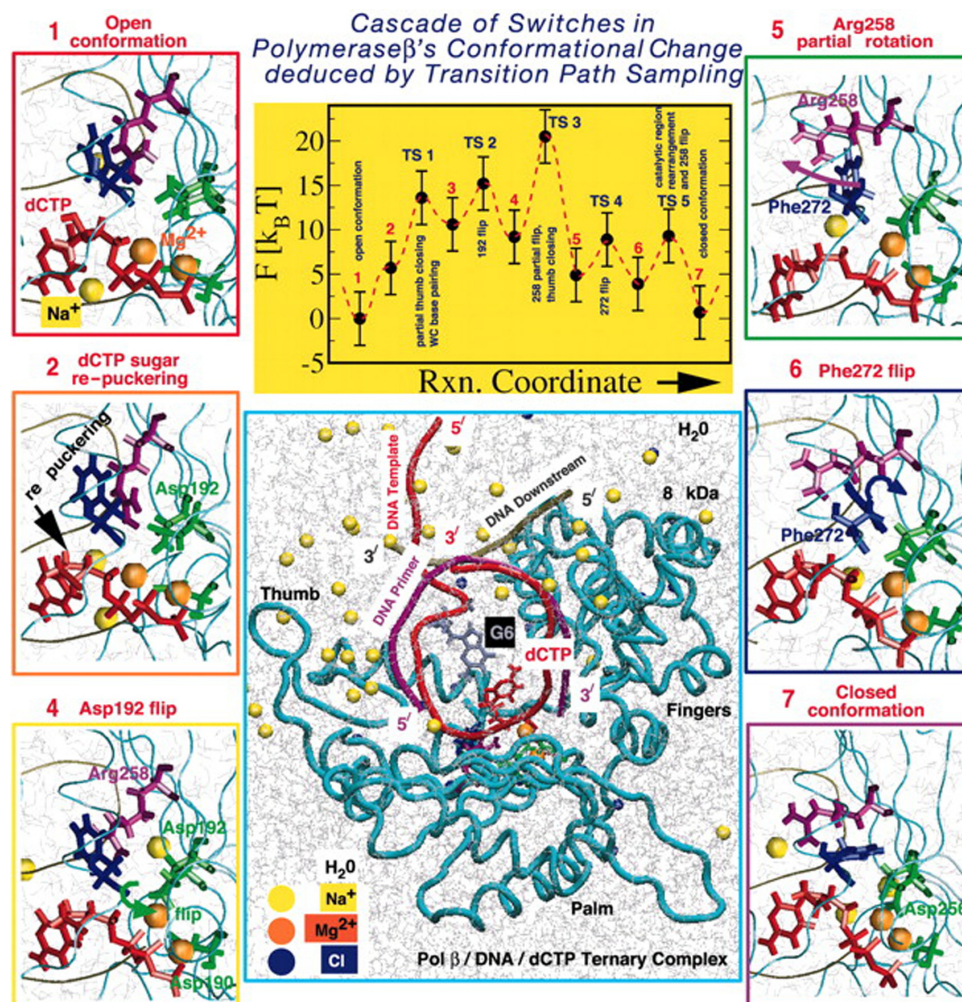
Figure 6.   Pol β's closing pathway.[1] Overall captured reaction kinetics profile (from TPS) for the conformational transition of pol β (for G:C) from open (state 1) to closed (state 7) forms showing free energies (in k_BT) associated with the different transition state regions.[1] The meta-stable basins (in red) along the reaction coordinate are numbered 1–7. Reprinted from R. Radhakrishnan and T. Schlick. 2004. Orchestration of cooperative events in DNA synthesis and repair mechanism unraveled by transition path sampling of DNA polymerase β's closing, **101**, 5970–5975, Copyright (2004) National Academy of Sciences, U.S.A.

Specifically, transition path sampling simulations for pol β for the correct G:C system[1,22] delineated the specific transition states and energies involved in a complex geometric/energetic landscape (Figure 6). The critical roles of key residues Arg258, Phe272, Asp192, and Tyr271 emerged: a delicate system of checks and balances directs the system to the chemical reaction and likely facilitates enzyme discrimination of the correct from the incorrect unit. Analogous simulations on the mispair G:A (Figure 7) suggest that while free-energy barriers for the matched and mismatched systems are comparable to one another for the conformational pathway, the closed state of the mismatch is much less stable than its open state, unlike the G:C complex, in which the closed and open states have equal stabilities. Thus, different sequences of transition states in the
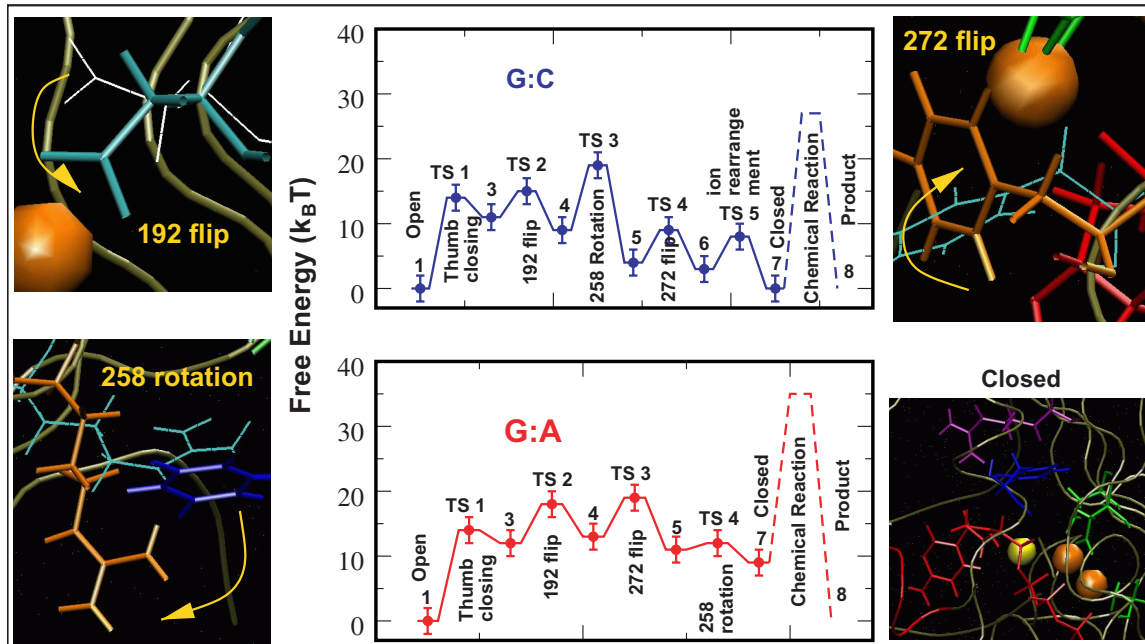
Figure 7. Pol β kinetic profile. Overall captured reaction kinetics profile for pol β's closing transition followed by chemical incorporation of dNTP for G:C and G:A systems.[22] The barriers to chemistry (dashed peaks) are derived from experimentally measured k_{pol} values. The profiles were constructed by employing reaction coordinate characterizing order parameters in conjunction with transition path sampling. The potential of mean force along each reaction coordinate is computed for each conformational event. Adapted with permission from "R. Radhakrishnan, K. Arora, Y. Wang, W. A. Beard, S. H. Wilson and T. Schlick. 2006. Regulation of DNA repair fidelity by molecular checkpoints: "Gates" in DNA polymerase β's substrate selection. *Biochem* **45**, 15142–15156." Copyright (2006) American Chemical Society.

correct versus incorrect basepair complexes dictate different conformational pathways toward an ideal two-metal ion catalysis geometry.

2.4 *Pol β's chemical mechanism for G:C vs. G:A systems*

A novel hybrid molecular/quantum mechanics approach (MM/QM) combining quasi-harmonic free energies with MM/QM dynamics[27] helped delineate a proton-hop mechanism for the chemical reaction of pol β.[27–29] Specifically, a series of transient intermediates is linked via a Grotthuss hopping mechanism of proton transfer between water molecules and the three conserved aspartate residues in the enzyme's active site (Figure 8). In the G:C system, the rate-limiting step is the initial proton hop with a free-energy of activation of at least 17 kcal/mol, which corresponds reasonably to measured k_{pol} values. Fidelity discrimination in pol β can be explained by a significant loss of stability of the closed ternary complex of the enzyme in the G:A system and a much higher activation energy of the initial step of nucleophilic attack, namely de-protonation of the terminal DNA primer O3′H group. Thus, subtle differences in the enzyme active site between matched and mismatched base pairs generate significant differences in catalytic performances.
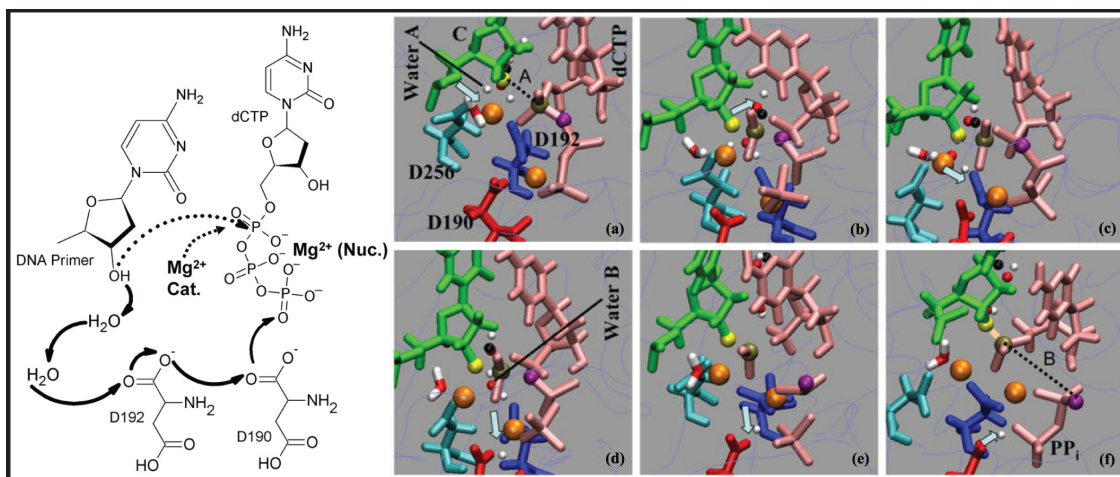
Figure 8.   Pol β's chemical synthesis reaction. Left: Schematic drawing of the mechanism of concerted proton hops during phosphoryl transfer in pol β. Solid arrows indicate the migration path of the proton and the dotted arrows represent the nucleophilic attack.[27] Right: Captured reaction intermediates for pol β's phosphoryl transfer in the G:C system:[27] (a) reaction state of the closed nucleotide-bound enzyme state; (b) deprotonation of the O3′H to water; (c,d) proton transfers to Asp192; (e) proton transfers to Asp190; (f) proton reaches the pyrophosphate unit to obtain the final product. The colors represent: cyan (D256), red (D190), blue (D192), pink (dCTP), green (CYT: terminal DNA primer), black (the O3′H-proton), yellow (the O3′ oxygen, attacking nucleophile), tan (central phosphorus), purple (leaving O3A oxygen), and orange (Magnesium). The oxygens and hydrogens of water molecules are in red and white, respectively. The arrows denote the location and direction of proton hop. Adapted from Biochemical and Biophysical Research Communications, 350/3, R. Radhakrishnan and T. Schlick, Correct and incorrect nucleotide incorporation pathways in DNA polymerase β's, 521–529, Copyright (2006), with permission from Elsevier.

2.5 *Pol λ's pathway and slippage tendency*

For pol λ, conformational transitions occur via flips of side-chain residues (Ile492, Tyr505, Phe506) and other motions associated with Arg514 and Arg517 in the active site en route to the binary (inactive), though no large-scale subdomain motion was noted.[24] However, significant DNA motion was observed, and this was proposed to be related to the tendency of this polymerase to generate deletion errors.

Dynamics simulations also identified residue Arg517 as crucial to protein/DNA stabilization through mutant simulations (Figure 9). The discrete orientations of the 517 residue can impact protein-coupled DNA stability by forming unfavorable electrostatic interactions, which lower the stability of the ternary complex and move the system toward the binary conformation. Residue 517's critical interactions with the DNA also help interpret pol λ's slippage tendency. Fragile protein/DNA interactions in pol λ might lead to deletion mutations because the DNA "slips". Through side-by-side experiments and additional simulations of other Arg517 mutants in pol λ, we validated the results from simulation and developed the hypothesis that Arg517 plays an important role in modulating deletion error generation.[32,33]

Modeling of pol λ bound to incorrect incoming nucleotides[34] revealed a wide range of DNA motion and protein residue side-chain motions, as well as distinct differences compared to the reference (correct base pair) system (see Figure 10). This led us to suggest key base-checking roles in pol λ of active site residues Arg517, Tyr505 and Phe506. On the basis of
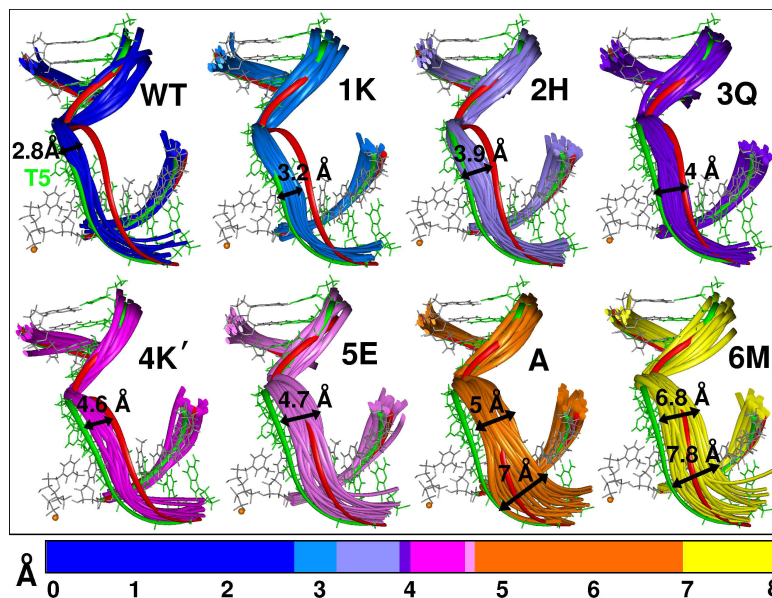
Figure 9. Residue 517 as key regulator of DNA/protein interactions in pol λ. Ranges of DNA motion occurring in each 517 mutant simulation as well as in the wild-type (WT) simulation with the nucleotide-binding ion.[32] Positions of the DNA are shown with reference to the crystal binary (red, PDB entry 1XSL) and ternary (green, PDB entry 1XSN) DNA positions. DNA bases of the template and primer strands are green and grey, respectively. A: Arg517Ala pol λ mutant. Reprinted with permission from "M. C. Foley and T. Schlick. 2008. Simulations of DNA pol λ R517 mutants indicate 517's crucial role in ternary complex stability and suggest DNA slippage origin. *J Amer Chem Soc* **130**, 3967–3977." Copyright (2008) American Chemical Society.
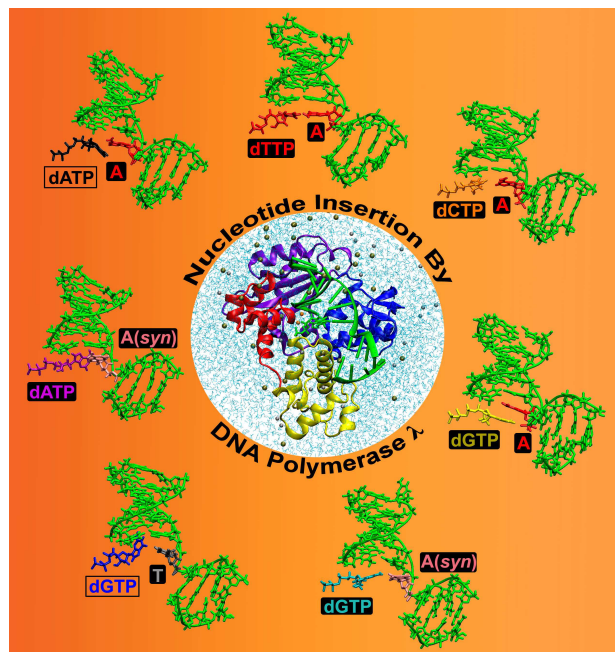

Figure 10. Pol λ's handling of different mismatched base pairs revealed from dynamics simulations.[34]

trends in the electrostatic interactions, we proposed an order of increasing mismatch insertion by pol λ: A:C > A:G > A (*syn*):G > T:G > A (*syn*):A > A:A. This sequence agrees with available kinetic data for incorrect nucleotide insertion opposite template adenine, with the exception of T:G, which may be more sensitive to the insertion context.

Recent modeling of aligned versus misaligned DNA[35] has also suggested that pol λ stabilizes the latter more tightly than aligned DNA, with active-site electrostatics playing a major role in this stabilization. This astonishing fact may be related to this enzyme's ability to handle substrates with minimal base pairing at the cost of generating a high rate of base deletions.

2.6 *Dpo4's handling of 8-oxoG*

Since Dpo4 has a prominent role in lesion bypass, it is natural to study Dpo4 in the context of complexes to small lesions like 8-oxoguanine (8-oxoG) (Figure 11), one of the most common lesions due to oxidative damage such as cigarette smoke and air polutants. These lesions are handled by base excision repair (BER) and nucleotide excision repair (NER) pathways, but the structural details are not well understood. The 8-oxoG lesion has been studied extensively in a variety of contexts both experimentally and by molecular modeling. These lesions can introduce errors because adenines opposite 8-oxoG can lead to mismatched *syn* 8-oxoG:A mispairs (Figure 11) which in turn can lead to C→T transversions.
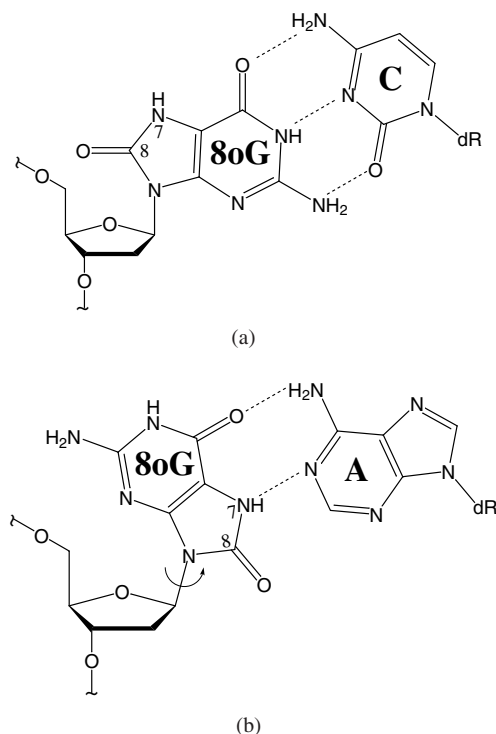

(a)

(b)

Figure 11.   8-oxoG (8oG) structure and base pairing possibilities. (a) 8-OxoG in an *anti* orientation forms a Watson-Crick base pair with dCTP, (b) 8-oxoG in a *syn* orientation forms a Hoogsteen base pair with dATP.

Table 2.   Inactive to active complex transitioning.

Enzymes	Motions Involved
Pol β	large-scale thumb motion, subtle protein side-chain motion
Pol λ	large DNA shift, small thumb loop motion, subtle protein side-chain motion
Pol X	large-scale thumb motion, subtle protein side-chain motion
Pol μ	subtle protein side-chain motion
Dpo4	DNA sliding, little finger rotation, subtle protein side-chain motion

Experiments have shown that Dpo4 has a 70-fold preference to insert dCTP rather than dATP opposite 8-oxoG.[36] Since the ternary structure is highly stable before chemistry, both with and without the incoming nucleotide at the active site, we hypothesized that the "induced-fit" mechanism does not operate in Dpo4. A rapid ~12 degree little-finger subdomain rotation occurring in our simulation after chemistry with ions in the active site suggested a low energy barrier for the transition between ternary and binary conformations. In addition, our studies of Dpo4 pathways for the correct insertion of dCTP opposite 8-oxoG using a hybrid quantum-classical mechanics approach revealed that the most favorable reaction path involves initial deprotonation of O3′H via two bridging water molecules to dCTP, overcoming an overall energy barrier of approximately 20 kcal/mol. The proton then migrates to the γ-phosphate oxygen of dCTP as the nucleotide is joined to the primer terminus and PP_i is formed. In contrast, initiating the chemical reaction from the less ideal state of the crystal structure requires a much higher activation energy barrier (29 kcal/mol) due to longer distances for O3′H proton transfer and distorted conformations of the proton acceptors. Compared to the higher fidelity pol β, Dpo4 has a higher chemical reaction barrier, which may result from its more solvent-exposed active site.

2.7 *Pol X mispair incorporation*

MD simulations of pol X bound to different mismatched nascent base pairs (i.e. C:C, A:G, G:G (*anti*), and G:G (*syn*)) were designed to explore a range of incorporation difficulty by pol X: G:G is very frequently inserted by pol X, A:G is moderately inserted, and C:C is relatively infrequently inserted compared to pol X's incorporation of the correct G:C base pair. Intriguingly, our simulations provide an explanation for this G:G preference: in the G:G (*syn*) mismatch simulation, the thumb exhibits a large-scale conformational change from an open to closed state that is similar to what occurs with the correct G:C base pair; in contrast, the A:G and G:G (*anti*) systems only display partial thumb closing and the C:C system maintains the thumb open (Figure 12). Thus, only the G:G (*syn*) base pair fits well within the helix via a Hoogsteen base pair arrangement that is geometrically similar to standard Watson-Crick base pairs. The more open and pliant active site of pol X, compared to pol β, allows pol X to accommodate bulkier mismatches such as G:G, while the more structured and organized pol β active site imposes higher discrimination that results in higher fidelity.
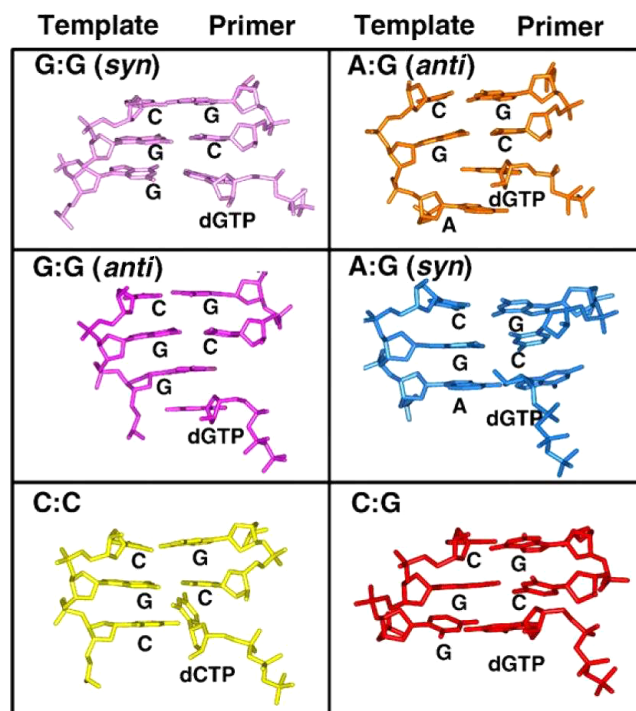
Figure 12.   Geometry of template–primer DNA base pairs bound to pol X.[37] Reprinted from Journal of Molecular Biology, 384/5, B. A. S. Benítez, K. Arora, L. Balistreri and T. Schlick, Mismatched base pair simulations for ASFV Pol X/DNA complexes help interpret frequent G:G misincorporation, 1086–1097, Copyright (2008), with permission from Elsevier.

3. Conclusion

Our genomic imprint is both venerable and vulnerable. The rapid duplication of our DNA requires a team of DNA polymerase molecules working in concert, each copying a sliver of a chromosome, with a rescue team of proofreaders and repair agents. The fascinating atomic and molecular aspects of these replication and repair processes are avidly being pursued by both experimentalists and modelers alike. Modeling in this field is well anchored in many available crystal endpoints and kinetic data, and can help shed insights into the operation of these fascinating workhorses. With rapid improvements in modeling and simulation,[16] theoretical techniques are becoming more reliable and hence attractive, but they require expertise in simulation methodology and data analysis to sort through a voluminous amount of information.

In this chapter, a taste of some studies in our group focusing on pol β, γ, μ, X, and Dpo4 have been presented. The modeling tools involve standard dynamics simulations, enhanced sampling protocols, hybrid classical/quantum methods, free energy methods, principal component analysis, and others.

The various simulations have shed insights into the conformational pathways involved in these enzymes — the transition from binary polymerase/DNA to ternary polymerase/DNA/dNTP complexes — and highlighted subtle protein-residue transitions, DNA motions, and protein

subdomain motions that are polymerase specific. Studies of these enzymes with mismatches help describe how well or how poorly the active site tolerates mismatched base pair systems, a fact that can be correlated to both the conformational transitions and the fidelity profiles of these enzymes. Indeed, pol λ's architecture is vulnerable to deletion errors, Dpo4's flexible active site handles lesions, pol X's open active site tolerates mispairs easily, and pol μ's lack of specific active-site interactions may allow it to perform repair with unpaired templates.

Mechanistic studies of the chemical reactions helped suggest atomic-level conformational steps in the process as well as associated free energy barriers. Together, such results can be combined to describe the kinetic cycle of polymerases as a monopoly board with conformational, pre-chemistry, and chemistry avenues, as shown in Figure 13. We suggest that while the conformational avenue is dictated by a defined sequence of conformational changes, the "pre-chemistry" avenue[38] involves stochastic sampling of active-site rearrangements to reach the chemical-reaction competent state and can help interpret different fidelity profiles for different polymerases. The experimental confirmation of some modeling predictions, such as the importance of Arg517 for pol λ,[33] the gate-keeping role of Arg258 in pol β,[23] or the G:G *syn* orientation preferred for pol X,[39] lend confidence in the modeling. Undoubtedly, ongoing studies via simulation and experiment will help dissect these separate avenues and the roles of key molecular components in polymerase mechanisms, including alternatives to the induced-fit mechanism. All such findings have important ramifications to understanding DNA synthesis and repair fidelity and, ultimately, the design of drugs to treat human diseases arising from polymerase errors.
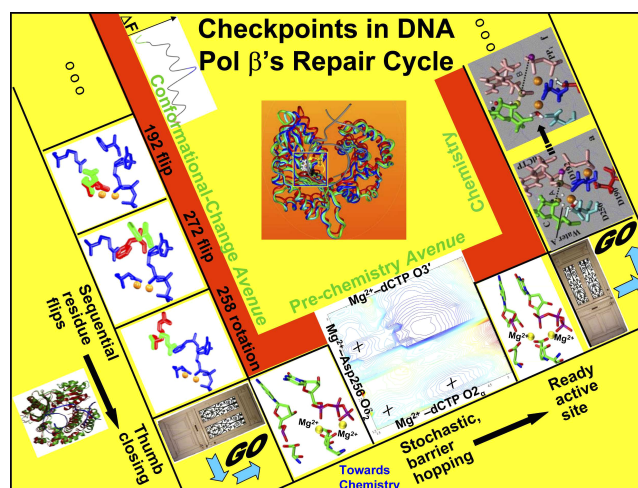

Figure 13. Pol β kinetic pathway. Sequential events and corresponding "gates" controlling pol β's fidelity:[38] the conformational change avenue, which comprises of Arg192 flip, Phe272 flip, and Arg258 rotation accompanying thumb subdomain closing motions upon incoming nucleotide binding; the pre-chemistry avenue, which involves the stochastic reorganization of the protein catalytic region, particularly the coordinating ligands of the two binding metal ions; and the chemistry avenue, where the incoming nucleotide is finally connected onto the primer terminus and the primer is extended by one residues.

Acknowledgment

Suggested Additional Reading Materials

A. A. Golosov, J. J. Warren, L. S. Beese and M. Karplus. 2010. The mechanism of the translocation step in DNA replication by DNA polymerase I: A computer simulation. *Structure* **18**, 83–93.

J. Yamtich and J. B. Sweasy. 2010. DNA polymerase family X: Function, structure, and cellular roles. *Biochim Biophys Acta* **1804**, 1136–1150.

C. E. McKenna, B. A. Kashemirov, L. W. Peterson and M. F. Goodman. 2010. Modifications to the dNTP triphosphate moiety: From mechanistic probes for DNA polymerases to antiviral and anti-cancer drug design. *Biochim Biophys Acta* **1804**, 1223–1230.

R. Radhakrishnan and T. Schlick. 2004. Orchestration of cooperative events in DNA synthesis and repair mechanism unravelled by transition path sampling of DNA polymerase β's closing. *Proc Natl Acad Sci U S A* **101**, 5970–5975.

References

1. R. Radhakrishnan and T. Schlick. 2004. Orchestration of cooperative events in DNA synthesis and repair mechanism unraveled by transition path sampling of DNA polymerase β's closing. *Proc Natl Acad Sci U S A* **101**, 5970–5975.

2. A. A. Golosov, J. J. Warren, L. S. Beese and M. Karplus. 2010. The mechanism of the translocation step in DNA replication by DNA polymerase I: A computer simulation. *Structure* **18**, 83–93.

3. J. Yamtich and J. B. Sweasy. 2010. DNA polymerase family X: Function, structure, and cellular roles. *Biochim Biophys Acta* **1804**, 1136–1150.

4. C. E. McKenna, B. A. Kashemirov, L. W. Peterson and M. F. Goodman. 2010. Modifications to the dNTP triphosphate moiety: From mechanistic probes for DNA polymerases to antiviral and anti-cancer drug design. *Biochim Biophys Acta* **1804**, 1223–1230.

5. M. R. Sawaya, R. Prasad, S. H. Wilson, J. Kraut and H. Pelletier. 1997. Crystal structures of human DNA polymerase β complexed with gapped and nicked DNA: Evidence for an induced fit mechanism. *Biochemistry* **36**, 11205–11215.

6. M. Garcia-Diaz, K. Bebenek, J. M. Krahn, T. A. Kunkel and L. C. Pedersen. 2005. A closed conformation for the pol λ catalytic cycle. *Nat Struct Mol Biol* **12**, 97–98.

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

8. B. A. S. Benítez, K. Arora and T. Schlick. 2006. Induced-fit mechanism for the interaction of the African swine fever virus DNA polymerase X with its target DNA. *Biophys J* **90**, 42–56.

9. A. F. Moon, M. Garcia-Diaz, K. Bebenek, B. J. Davis, X. Zhong, D. A. Ramsden, T. A. Kunkel and L. C. Pedersen. 2007. Structural insight into the substrate specificity of DNA polymerase μ. *Nat Struct Mol Biol* **14**, 45–53.

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

11. W. A. Beard and S. H. Wilson. 2003. Structural insights into the origins of DNA polymerase fidelity. *Structure* **11**, 489–496.

12. M. W. Maciejewski, R. Shin, B. Pan, A. Marintchev, A. Denninger, M. A. Mullen, K. Chen, M. R. Gryk and G. P. Mullen. 2001. Solution structure of a viral DNA repair polymerase. *Nat Struct Biol* **8**, 936–941.

13. A. K. Showalter, I. J. Byeon, M. I. Su and M. D. Tsai. 2001. Solution structure of a viral DNA polymerase x and evidence for a mutagenic function. *Nat Struct Biol* **8**, 942–946.

14. T. Schlick. 2010. *Molecular Modeling: An Interdisciplinary Guide.* Springer-Verlag, New York, Second edition.

15. E. H. Lee, J. Hsin, M. Sotomayor, G. Comellas and K. Schulten. 2009. Discovery through the computational microscope. *Structure* **17**.

16. T. Schlick, R. Collepardo-Guevara, L. A. Halvorsen, S. Jung and X. Xiao. 2011. Biomolecular modeling and simulation: A field coming of age. *Quart Rev Biophys* **44**, 191–228.

17. K. Arora and T. Schlick. 2004. *In silico* evidence for DNA polymerase β's substrate-induced conformational change. *Biophys J* **87**, 3088–3099.

18. L. Yang, W. A. Beard, S. H. Wilson, S. Broyde and T. Schlick. 2002. Polymerase β simulations suggest that Arg258 rotation is a slow step rather than large subdomain motion *per se. J Mol Biol* **317**, 651–671.

19. L. Yang, W. A. Beard, S. H. Wilson, B. Roux, S. Broyde and T. Schlick. 2002. Local deformations revealed by dynamics simulations of DNA polymerase β with DNA mismatches at the primer terminus. *J Mol Biol* **321**, 459–478.

20. L. Yang, W. A. Beard, S. H. Wilson, S. Broyde and T. Schlick. 2004. The highly organized but pliant active-site of DNA polymerase β: Compensatory interaction mechanisms in mutant enzymes by dynamics simulations and energy analyses. *Biophys J* **86**, 3392–3408.

21. L. Yang, K. Arora, W. A. Beard, S. H. Wilson and T. Schlick. 2004. The critical role of magnesium ions in DNA polymerase β's closing and active site assembly. *J Am Chem Soc* **126**, 8441–8453.

22. R. Radhakrishnan and T. Schlick. 2005. Fidelity discrimination in DNA polymerase β: differing closing profiles for a mismatched G:A versus matched G:C base pair. *J Amer Chem Soc* **127**, 13245–13252.

23. M. Bakhtina, M. P. Roettger, S. Kumar and M.-D. Tsai. 2007. A unified kinetic mechanism applicable to multiple DNA polymerases. *Biochem* **46**, 5463–5472.

24. M. C. Foley, K. Arora and T. Schlick. 2006. Sequential side-chain residue motions transform the bionary into the ternery state of DNA polymerase A. *Biophys J* **91**, 3182–3195.

25. Y. Wang, K. Arora and T. Schlick. 2006. Subtle but variable conformational rearrangements in the replication cycle of *Sulfolobus solfataricus* P2 DNA polymerase IV may accommodate lesion bypass. *Protein Sci* **15**, 135–151.

26. Y. Li and T. Schlick. 2010. Modeling DNA polymerase μ motions: Subtle transitions before chemistry. *Biophys J* **99**, 3463–3472.

27. R. Radhakrishnan and T. Schlick. 2006. Correct and incorrect nucleotide incorporation pathways in dna polymerase β's. *Biochem Biophys Res Comm* **350**, 521–529.

28. I. L. Alberts, Y. Wang and T. Schlick. 2007. DNA polymerase β catalysis: Are different mechanisms possible? *J Amer Chem Soc* **129**, 11100–11110.

29. M. D. Bojin and T. Schlick. 2007. A quantum mechanical investigation of possible mechanisms for the nucleotidyl transfer reaction catalyzed by DNA polymerase β. *J Phys Chem B* **111**, 11244–11252.

30. C. Dellago and P. G. Bolhuis. 2007. Transition path sampling simulations of biological systems. *Top Curr Chem* **268**, 291–317.

31. R. Radhakrishnan and T. Schlick. 2004. Biomolecular free energy profiles by a shooting/umbrella sampling protocol, "BOLAS". *J Chem Phys* **121**, 2436–2444.

32. M. C. Foley and T. Schlick. 2008. Simulations of dna pol λ R517 mutants indicate 517's crucial role in ternary complex stability and suggest DNA slippage origin. *J Amer Chem Soc* **130**, 3967–3977.

33. K. Bebenek, M. Garcia-Diaz, M. C. Foley, L. C. Pedersen, T. Schlick and T. A. Kunkel. 2008. Substrate-induced DNA strand misalignment during catalytic cycling by DNA polymerase λ. *EMBO Reports* **9**, 459–464.

34. M. Foley and T. Schlick. 2009. The relationship between conformational changes in pol λ's active site upon binding incorrect nucleotides and mismatch incorporation rates. *J Phys Chem B* **113**, 13035–13047.

35. M. C. Foley, V. Padow and T. Schlick. 2010. The extraordinary ability of DNA pol λ to stabilize misaligned DNA. *J Amer Chem Soc* **132**, 13403–13416.

36. O. Rechkoblit, L. Malinina, Y. Cheng, V. Kuryavyi, S. Broyde, N. E. Geacintov and D. J. Patel. 2006. Stepwise translocation of Dpo4 polymerase during error-free bypass of an oxoG lesion. *PLoS Biol* **4**, e11.

37. B. A. S. Benítez, K. Arora, L. Balistreri and T. Schlick. 2008. Mismatched base pair simulations for ASFV Pol X/DNA complexes help interpret frequent G:G misincorporation. *J Mol Biol* **384**, 1086–1097.

38. R. Radhakrishnan, K. Arora, Y. Wang, W. A. Beard, S. H. Wilson and T. Schlick. 2006. Regulation of DNA repair fidelity by molecular checkpoints: "Gates" in DNA polymerase β's substrate selection. *Biochem* **45**, 15142–15156.

39. S. Kumar. 2008. *Biochemical, Mechanistic, and Structural Characterization of DNA Polymerase X from African Swine Fever Virus.* Ph.D. thesis, Department of Chemistry, Ohio State University. http://etd.ohiolink.edu/view.cgi?acc_num=osul211380265.