CHAPTER 4

Intrinsic Motions of DNA Polymerases Underlie Their Remarkable Specificity and Selectivity and Suggest a Hybrid Substrate Binding Mechanism

MEREDITH C. FOLEY,[a] KARUNESH ARORA[b] AND TAMAR SCHLICK*[a]

[a] Department of Chemistry and Courant Institute of Mathematical Sciences, New York University, 251 Mercer Street, New York, New York 10012, USA; [b] Department of Chemistry & Biophysics Program, University of Michigan, Ann Arbor, MI 48109, USA
*Email: schlick@nyu.edu

4.1 Introduction

DNA replication is a fundamental biological process involving numerous enzymes that help to ensure the maintenance of genomic integrity.[1] DNA polymerase enzymes are key components of this cellular replication machinery. Their primary function is to selectively incorporate correct nucleotide [*i.e.*, a dNTP (where N = A, T, C or G), 2′-deoxyribonucleoside 5′-triphosphate] to pair with the template base in accordance with Watson-Crick DNA base pairing rules (*i.e.*, A opposite T or G opposite C).[2] The ability of polymerases to

RSC Biomolecular Sciences No. 24
Innovations in Biomolecular Modeling and Simulations: Volume 2
Edited by Tamar Schlick

Published by the Royal Society of Chemistry, www.rsc.org

incorporate the correct nucleotide opposite template DNA bases determines their synthesis fidelity.[3] Depending on the polymerase and the DNA context, this ability can range from one incorrect insertion per almost every attempt (*e.g.*, pol X)[4] to ~1 error in every 10^6 to 10^7 insertion events (*e.g.*, pol ε).[5] DNA polymerase malfunction is implicated in several cancers and neurological disorders.[6–8] Therefore, understanding the fundamental principles that underlie selective nucleotide incorporation by DNA polymerases is medically important. The mechanistic insights into DNA polymerase mechanisms may open avenues for the rational drug design against several human cancers and other diseases associated with the genesis of DNA mutations and uncontrolled replication.[9,10]

During the process of nucleotide selection/incorporation, DNA polymerases undergo large-scale and long-time conformational changes that are thought to provide a mechanism for checking for correct base pairing. The dynamical transitions of DNA polymerases between functionally important conformations may play a crucial role in determining their fidelity.[11] Therefore, understanding polymerase dynamics can yield important insights into the process of nucleotide incorporation. Solving this challenging, medically-relevant problem requires a combination of theory, simulation, and experiment. Theory and simulation can provide an atomically detailed picture of both the thermodynamics and kinetics of conformational changes, and experiment can provide detailed structural information and macroscopic kinetics data. However, despite several experimental and theoretical efforts, the underlying principles that govern the nucleotide selection mechanism of DNA polymerases are not well understood. Static crystal structures of DNA polymerases have provided excellent starting points to begin understanding polymerase architecture in a detailed manner, but lack dynamical information crucial for understanding the polymerase mechanisms.[12,13] Similarly, experimental kinetic studies that probe mechanistic questions of polymerase mechanisms provide reaction rate data but no information on the corresponding structural evolution.[14–16] Computer simulations can bridge this gap and help relate kinetic data from experiments to functionally important structural changes that occur during the molecular recognition event.[17,18] Rapid increases in the availability of computer power and algorithmic advances have made possible increasingly longer simulations of large biomolecules in atomic details.[17,19,20] Thus, despite uncertainties and limitations, bimolecular modeling and simulation are becoming a full partner with experiment, as recently surveyed.[17] However, elucidating atomically detailed conformational changes taking place during the binding of substrates at increasing length and time scales as well as the corresponding underlying energy landscape remains a considerable challenge.[21,22]

We have applied standard and enhanced sampling molecular dynamics simulations, as well as advanced trajectory analysis techniques to describe aspects of DNA polymerase mechanisms. The studies have elucidated common dynamical characteristics of DNA polymerases, which are important for understanding polymerase function. In this chapter, we focus on presenting mechanistic insights gained from simulations of X-family DNA polymerases

from our group rather than methodological details, which can be found in the individual papers.[22–31] We begin by summarizing the knowledge of DNA polymerase structure and function available from experiment. We then briefly discuss polymerase simulation and trajectory analysis methods. We follow with a description of the key results from simulations of pol β, pol λ, pol X, and pol μ complexes containing correct or incorrect incoming nucleotides, which provide fresh insights into how the structural reorganization of polymerase complexes may relate to function and fidelity. We then present the main findings from correlated motions and principal component analysis (PCA) of our simulations: Correlated motions analysis reveals important differences between the behavior of dynamic networks within correct and incorrect nucleotide systems and suggest a key role of the dNTP in adjusting the motions of various substrate complexes. Furthermore, PCA of dynamics trajectories indicates that the dominant collective motions that occur during incoming nucleotide-dependent conformational transitions, and captured in simulations, are described by largest eigenvector-eigenvalue pairs and are motions intrinsic to the polymerase structure. Finally, we relate these intrinsic motions with our observation that the incoming nucleotide affects both active-site organization and dynamic networks: we propose a hybrid substrate binding mechanism for polymerases that incorporates features of both the induced-fit[32] and conformational selection[33,34] models for substrate binding. We conclude by proposing several possible ways this emerging view of DNA polymerase dynamics can be exploited in the structure-based drug design for the treatment of diseases resulting from DNA polymerase errors.

4.2 DNA Polymerase Structure and Function

As shown in Figure 4.1, all DNA polymerases are shaped like a hand; however, despite this homology, the motions of polymerases can vary significantly. Indeed, many higher fidelity polymerases including T7 DNA pol,[35] the large fragment of *Thermus aquaticus* DNA pol I (Klentaq1),[36] *Bacillus stearothermophilus* DNA pol I fragment (BF),[37] and the Klenow Fragment (KF) of *Escherichia coli* pol I[38] exhibit large-scale rearrangements of the fingers subdomain. Moderate-fidelity pol β and low-fidelity pol X also have large-scale protein rearrangements. Yet several lower-fidelity polymerases like pol λ,[24,39,40] pol μ,[30] terminal deoxynucleotidyltransferase,[41] and Y-family polymerases such as Dpo4[42] bind dNTP without large-scale conformational transitions. Intriguingly, the magnitude of DNA motion can also differ significantly among polymerase complexes. Lower fidelity pol λ utilizes large DNA shifting,[24,40] and Dpo4 exhibits DNA sliding.[43] In this chapter, we focus on the X family of DNA polymerases for which there are abundant experimental data.[9,44–46] Representative motions of pol β, pol X, and pol λ are shown in Figure 4.2.

Extensive kinetic, structural, and computational evidence[47–49] suggests that DNA polymerases follow a common catalytic mechanism (Figure 4.3). First, the polymerase binds DNA that will be extended. Second, a conformational

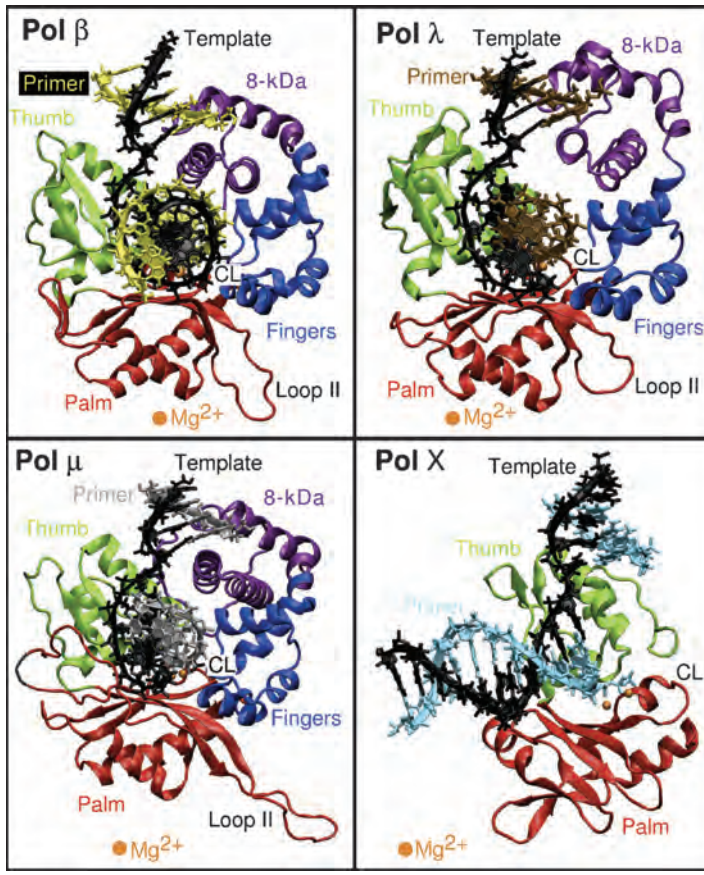

Figure 4.1 Structures of repair DNA polymerases from the X-family (pol β, pol λ, pol X, and pol μ). CL, conserved palm loop connecting two β-strands that contains two of the three catalytic aspartate residues.

change occurs that is associated with dNTP binding and transitions the complex from an inactive to active state. Third, a chemical reaction occurs, which requires two divalent ions, a catalytic and nucleotide-binding ion. This reaction is typically associative, *i.e.*, the new O–P bond begins to form before the P–O bond to the leaving group completely breaks, producing a phosphorane transition state,[50] and has been investigated quantum mechanically in many polymerase systems.[48,51–65] Fourth, following chemistry, the pyrophosphate moiety is released and the enzyme returns to its initial substrate-free conformation.

To determine how conformational changes within the polymerase complexes relate to function and fidelity we analyze the motions of pol β, pol λ, pol X, and pol μ complexes containing correct or incorrect incoming nucleotides. Results of this analysis are discussed in the following sections.

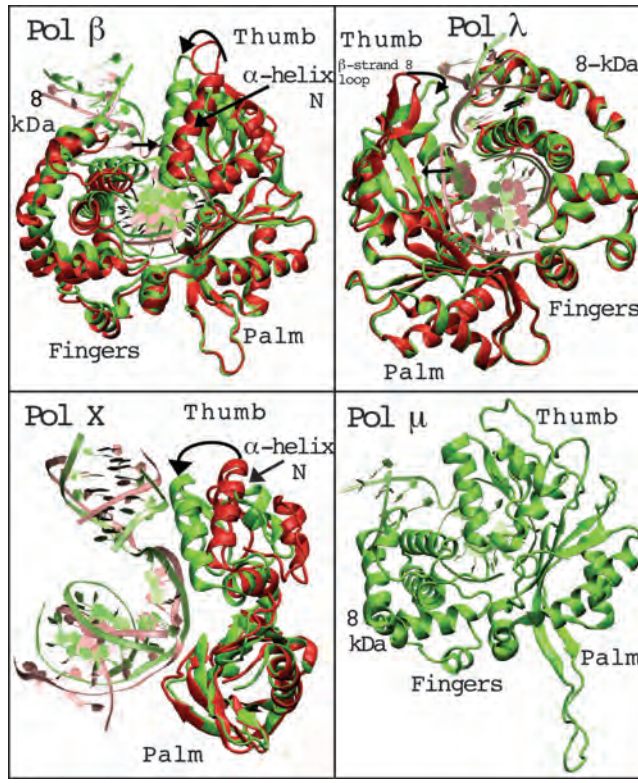

Figure 4.2 Motions of pol β, pol λ, pol X, and pol μ upon binding the correct incoming nucleotide. Structures colored red or pink represent the inactive or open conformation. Structures colored green or light green represent the active or closed conformation. Pol β images are derived from open (PDB entry 1BPX) and closed (PDB entry 1BPY) X-ray crystal data. Pol β is depicted using X-ray crystal data for the protein/DNA binary complex (PDB entry 1XSL) and protein/DNA/dTTP ternary complex (PDB entry 1XSN). Pol X coordinates are taken from the initial and final snapshots of the pol X/C:G trajectory. The pol μ form is derived from X-ray crystal data for the before chemistry ternary complex (PDB entry 2IHM); missing loops are modeled.

4.3 Methods

Atomic models of polymerases were prepared from either high-resolution crystal structures or NMR structures deposited in RCSB Protein Data Bank (PDB).

4.3.1 Molecular Dynamics

The standard molecular dynamics (MD) protocols we use for simulations of pol β, pol X, pol λ, and pol μ analyzed in this work are fully described in our

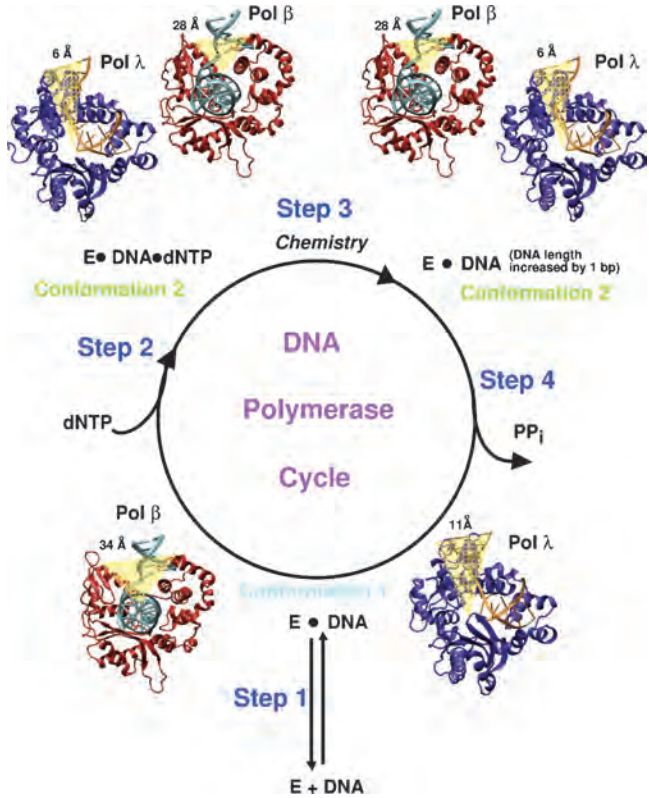

Figure 4.3 General DNA polymerase catalytic cycle for dNTP insertion. Conformational changes in pol β and pol λ are shown for reference.

prior published works.[23,24,29–31,66] Essentially, all simulations are performed at 300 K with the CHARMM force field.[67] For pol β and pol X, the CHARMM program[68] is used with the Langevin multiple time-step LN integrator;[69] we use an inner time step $\Delta\tau = 1$ fs for updating local bonded interactions, a medium time step $\Delta\tau_m = 2$ fs for updating non-bonded interactions within 7 Å, and an outer timestep $\Delta\tau = 150$ fs for calculating the remaining terms. For pol λ and pol μ, simulations are performed using the NAMD program[70] with a 2 fs time step; full electrostatics are computed using the PME method.[71] All simulations utilize the SHAKE algorithm to constrain bonds involving hydrogen atoms.

4.3.2 Principal Component Analysis

To identify the most significant fluctuation modes of pol β, pol X, and the I492A pol λ mutant, dynamics trajectories of correct dNTP systems are analyzed using PCA.[72] PCA describes the overall dynamics of systems with a few collective, "essential" degrees of freedom in which anharmonic motion takes

place. Many subdomain motions have been examined by PCA.[73–77] In brief, the positional covariance matrix, C, is constructed from equilibrated MD trajectories after removing the rotational and translational motions of systems by alignment to the initial protein configuration by means of a least-square-fit procedure using all protein heavy atoms. For M snapshots of N atoms, C is a $3N \times 3N$ covariance matrix:

$$C = M^{-1} \sum_{k=1}^{M} \left[(r_i(k) - \langle r_i \rangle) \right] \cdot \left[(r_j(k) - \langle r_j \rangle) \right],$$

where r_i and r_j are position vectors of two atoms i and j in the fitted structure and the angular brackets ($\langle r_i \rangle$) represent averages over the trajectories. The eigenvectors of the covariance matrix, V, together with their corresponding eigenvalues, λ, are obtained by diagonalizing the covariance matrix C, *i.e.*,

$$V^T C V = \wedge, \quad \text{or} \quad C V_n = \lambda_n V_n, \quad n = 1, 2, \ldots 3N,$$

where $\wedge$ is the diagonal matrix with eigenvalues λ_i. $\wedge = diag(\lambda_1, \lambda_2, \ldots \lambda_{3N})$. Each eigenvector V_n defines the direction of motion of N atoms as an oscillation about the average structure $\langle X \rangle$. The normalized magnitude of the corresponding eigenvalue is a measure of the amplitudes of motion along the eigenvector V_n as calculated by $\lambda_i / \sum_i \lambda_i$. When eigenvalues are arranged in decreasing order, the first few describe the largest positional fluctuations.

Note that although principal component analysis is not reliable for predicting long-time dynamics of proteins[78] due to sampling errors, it is reasonable when large-scale motions are captured during the simulation length.[78] In our straightforward simulations of pol β, pol λ, and pol X, transitions between functionally important states were captured and thus PCA can be informative.

4.3.3 Correlated Motion Analysis

The extent of correlated motions (positive or negative) for two molecular components can be quantified by calculating the covariance between the two units (*e.g.*, atoms, residues, and subdomains).[79] To investigate correlated motions among the DNA polymerase, DNA, and the incoming nucleotide as well as compare correlated motions among DNA polymerases, we calculated the normalized covariance, C_{ij}, for the displacement of all C_α or heavy atom pairs, i and j, as given by:

$$C_{ij} = \frac{\langle \Delta \mathrm{r}_i \cdot \Delta r_j \rangle}{\left(\langle \Delta r_i^2 \rangle \langle \Delta r_j^2 \rangle \right)^{1/2}}$$

where Δr_i is the displacement from the mean position of the i $t\square$ atom determined from all configurations in the trajectory segment being analyzed. C_{ij} is normalized to be between -1 and 1. For "in phase" motions (*i.e.*, when two

atoms are moving in the same directions), $C_{ij} = 1$, and, for completely anti-correlated motions (*i.e.*, when two atoms are moving in opposite directions), $C_{ij} = -1$.

Covariance analysis is performed on 10–40 ns MD trajectories of pol β, pol λ, pol X, and pol μ bound to both correct and incorrect nucleotides.[23–25,27,30,31,66,70] Pol β simulations are initiated from an intermediate conformation between the open and closed crystal forms whereas pol X simulations are initiated from a fully open conformation. For pol λ and pol μ, simulations are started from comparatively closed conformations found in their X-ray crystal structures with the correct dNTP bound.[40,80] We also analyzed simulations of the I492A pol λ mutant bound to the correct incoming nucleotide,[24] which began from the X-ray crystal structure of the wild-type polymerase conformation before dNTP binding[40] that also resembles the closed forms of pol β and pol X; this simulation captures a large rotation in the thumb loop containing β-strand 8 and a shift in the DNA to the active position as suggested for wild-type pol λ from X-ray crystal data.[40] For these covariance analyses, rotations and translations of the proteins occurring during the trajectories are removed by alignment to the initial protein configuration using a least-square-fit procedure for all protein heavy atoms.

4.4 Results and Discussions

4.4.1 Similarities in the Dynamics of X-Family DNA Polymerases and Their Effects on Function and Fidelity

4.4.1.1 Similar "Gate-keeping" Residues Identified for Different Large-Scale Polymerase/DNA Motions

Pol β, pol λ, pol μ and pol X exhibit varying amounts of protein and DNA motion: pol β and pol X show large-scale motions of the thumb, pol λ displays small loop motion and large DNA shifting, and pol μ reveals only local dNTP-binding pocket motions (Figure 4.2). Yet, collectively, our analyses suggest underlying mechanical similarities in their use of analogous palm residues as gate-keepers for key rearrangements. These residues are Arg258 for pol β, Ile492 for pol λ, and Phe102 for pol X.

Specifically, our enhanced sampling simulations of pol β by transition path sampling (TPS) have suggested that rotation of Arg258 in the palm is rate-limiting in the conformational closing pathway.[22] X-ray crystal structures initially suggested that the movement of Arg258 has an important role in active-site assembly because Arg258 forms a salt bridge to Asp192 in the open form, which keeps the aspartate away from the metal ion binding area, and a rotation in Arg258 releases the aspartate so that it can coordinate the ions in the closed form.[12] Additional evidence for the gate-keeping role of Arg258 in pol β's large-scale thumb conformational change comes from TPS simulations of the R258A mutant that suggest that thumb closing occurs more quickly in

this mutant relative to wild-type pol β.[48] This thumb/residue relationship is further supported by stopped-flow fluorescence experiments, which demonstrated an increased rate of opening following the chemical reaction in the R258A mutant.[14]

Pol λ's Arg258 analogue, Ile492, cannot form the same type of interactions as Arg258 in pol β due to its neutral charge. However, the I492A mutant displays increased thumb loop motion compared to wild-type pol λ and shifts the DNA toward the active DNA position.[24] This suggests that Ile492 may control a similar gate for pol λ's conformational transitions. In pol X, the motion of the analogous residue, Phe102, may similarly be related to large-scale thumb motion together with fluctuations in several adjacent thumb residues (*i.e.*, His115, Phe116, and Val120).[25]

These motions of analogous active-site residues in different polymerases with varying fidelities are significant because they indicate how pico- to nano-second protein side-chain fluctuations can affect the large-scale subdomain motions. Mutations in these side chains can alter the amplitude of the subdomain motion[81] and hence function of the enzyme.

4.4.1.2 *Efficiency of Incorrect Nucleotide Insertion Suggested by Comparison to Correct Systems*

The wealth of kinetics and error rate data for both high and low fidelity DNA polymerases show that these enzymes vary widely in their ability to discriminate against incorrect nucleotides. X-ray crystal structures for select DNA polymerases from the A, X, and Y-families bound to non-lesioned DNA and mismatches have resolved some important questions raised by these biochemical data. These include identifying pol ι's unusual active-site stabilization of T/U:G that facilitates its insertion,[82,83] and revealing how sensitive pol β/DNA interactions and conformational states are to the location of the mismatch within the DNA sequence and to the incorporation state of the mismatch.[84–86] Complementary computational studies of pol β mismatch insertion by several groups have also provided insights into conformational transitions before chemistry,[27,87] chemical reaction pathways,[56,61,65] transition states,[88–90] and the potential for mismatch extension.[91]

Our simulations of incorrect dNTP complexes of pol β, pol X, pol μ, and pol λ suggest specific dynamics features of mismatch distortions that can predict how easily an incorrect dNTP would be inserted. Namely, three major areas of motion appear useful for characterizing mismatch system distortions (see Figure 4.4): altered large-scale motions of protein subdomains or DNA; unusual local active-site motions in protein side chains and DNA base pairs; and rearrangements involving the reactive atoms of the primer terminus, incoming nucleotide, and metal ions and their ligands. Mismatches that are poorly inserted are associated with thumb opening motion or DNA shifting toward an inactive position as well as disordered active sites. Mismatches that are inserted more easily have polymerase/DNA/dNTP complex geometries similar to correct base pair systems.

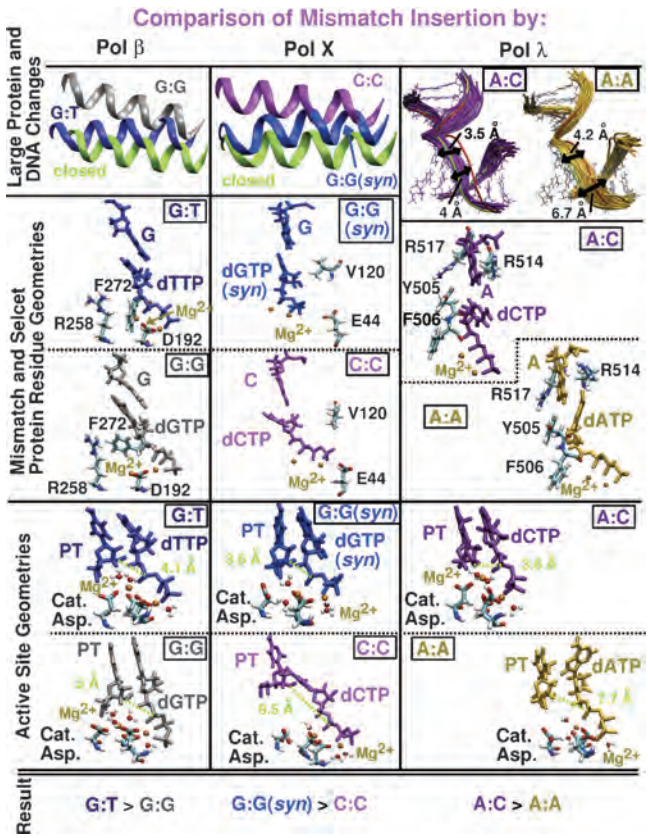

Figure 4.4 Comparison of mismatch insertion by pol β, pol X, and pol λ. Mismatches that are relatively easily inserted by these polymerases (*e.g.*, G:T, G:G (*syn*), and A:C for pol β, pol X, and pol λ, respectively) are compared to mismatches that more difficultly inserted (*e.g.*, G:G, C:C, and A:A for pol β, pol X, and pol λ, respectively). PT, primer terminus.

For example, pol X's binding of G:G (*syn*), which resembles WC base pairs, triggers thumb *closing* motion like the four WC pairs; this mispair geometry preference has since been validated by solution structures capturing pol X bound to G:G(*syn*).[92] Likewise, the pol β/G:T complex exhibits thumb closing and the pol λ/A:C complex has less DNA motion toward the inactive position. Conversely, mismatches that are weakly inserted appear to change the basic closing dynamics, reverting instead to thumb opening (*e.g.*, pol β/G:G and pol X/C:C complexes) or significant DNA motions toward an inactive position (*e.g.*, pol λ/A:A complex); these protein and DNA motions distort specific protein side chain/DNA interactions in critical ways.

For example, in pol β/G:G, Phe272 and Arg258 do not transition to their active or closed complex positions and, in pol X/C:C, greater variability in the side chains of Val120 and Glu44 occurs than in correct base pair systems.

In pol λ/A:A, rearrangements in Tyr505, Phe506, Arg514, and Arg517 occur that alter interactions with the DNA and lead to additional distortions such as pairing between the templating base and the primer terminus, and a rotation in the dATP that considerably lengthens the O3′–P_α distance. Mismatches in pol β and pol X that are inserted with greater difficultly also have longer O3′–P_α distances as well as less direct interactions between the Mg^{2+} ions and both the primer terminus' O3′ and the catalytic aspartates due to the presence of additional water molecules in the active site. The disordered nature of mismatch complexes may also explain why certain DNA polymerase mismatch complexes are less amenable to crystallization.

Aside from these motions, pol λ mismatch studies reveal the relatively poorer electrostatic interactions that incorrect incoming nucleotides have with the polymerase active-site pocket as compared to the correct nucleotide.[31] This feature could aid discrimination of incorrect nucleotides by triggering some of the rearrangements described above that promote nucleotide release. In pol λ, the 2-amino group of incorrect dGTPs has especially poor active-site energetic interactions and these may cause the large dGTP motions in the T:G and A(*syn*):G simulations.[31]

Interestingly, changes in the insertion tendencies of purine dNTPs have been experimentally observed following the addition or removal of functional groups like the 2-amino group for several different DNA polymerases.[93–95] This suggests that DNA polymerase fidelity may be based, at least in part, on the formation of unfavorable energetic interactions of the substrate with the polymerase. These interactions assist in discriminating between incoming nucleotides based on nucleotide functional groups residing in non-WC positions within the active site. Incorrectly bound nucleotides require rearrangements within the protein/DNA/dNTP complex to improve interactions and may also result in active-site disorder and a reduced efficiency of incorporation.

Together, these mismatch data suggest that DNA polymerase systems utilize variations in motions, structures and energetics to discriminate between correct and incorrect dNTPs. Thus, natural variations in the structure and sequences of polymerases contribute to differences in nucleotide discrimination.

4.4.2 PCA Reinforces the Functional Importance of Thumb and DNA Motions

Our PCA of pol β, pol X, and the I492A pol λ mutant simulations, which capture the characteristic thumb and/or DNA motions of each polymerase when bound to the correct dNTP, reveals the dominant motion in these systems: the largest ten principal components contribute respectively 79.6%, 84.5%, and 74.3% to the amplitude of fluctuations defined by the covariance matrices. Projection of each MD trajectory along the largest eigenvalue-eigenvector pair captures large subdomain and DNA motions in these systems, suggesting that these motions are functionally relevant (Figure 4.5[a]). Previously reported normal mode analysis using a simple elastic network model

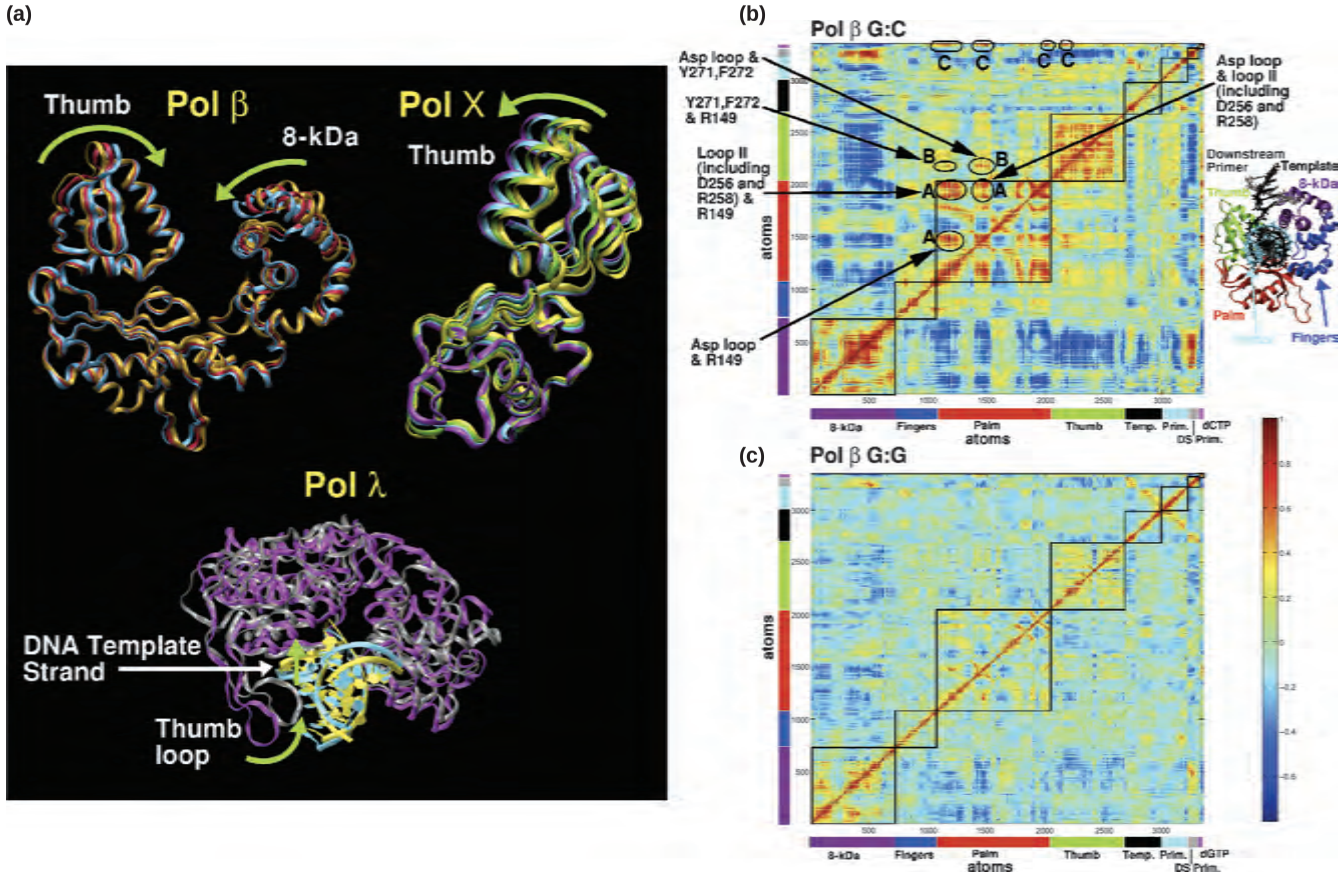

Figure 4.5 PCA data for X-family polymerases and pol β covariance matrices for correct and incorrect dNTP systems. (a) Projection of pol β, pol X, and I492A pol λ mutant MD trajectories on the first principal component captures large subdomain and DNA motions in these systems. Green arrows show direction of motion in labeled protein and DNA regions. (b) Covariance matrix for protein/DNA/dNTP heavy atoms in pol β from simulations with correct and incorrect nucleotides bound. Colors: purple (8-kDa domain), blue (fingers), red (palm), green (thumb), black (DNA template strand), cyan (DNA primer strand), silver (DNA downstream primer), and magenta (dNTP).

of several DNA polymerase crystal structures[96] also supports the functional relevance of these large-scale motions. Pol λ's thumb loop motion, in particular, has also been found to be functionally important through experimental deletion studies.[97]

4.4.3 Coupled Polymerase/DNA/dNTP Motions Bridge Spatial Gaps and Impact DNA Polymerase Function and Fidelity

To provide an overview of the dynamics of each polymerase system that can help in identifying important features for function and fidelity, we perform covariance analysis of the simulation data for correct, incorrect, mutant, and misaligned DNA complexes (Figures 4.5 and 4.6 and 4.7 and 4.8). For comparative purposes, we also include covariance analysis of the pol μ/correct dNTP complex (Figure 4.8[a]), which does not exhibit large-scale protein or DNA motions. Table 4.1 summarizes the correlated motions data for all four polymerases. This type of analysis provides an overview of the dynamics of each polymerase system that is useful for identifying their importance for function and fidelity.

4.4.3.1 Coupled Motions Create Synchronized Dynamic Networks with Correct dNTPs

Underlying similarities in the dynamics of pol β, pol X, pol λ, and pol μ when bound to the correct dNTP are summarized in Table 4.1. For example, within pol β's correct G:C system, a network within the palm exists that joins different regions of the subdomain with the active site. This palm network includes: a conserved X-family loop joining two β-strands and containing two of the three catalytic aspartates;[98] a region adjacent to this loop containing the third catalytic aspartate (Asp256) and nearby gate-keeping residue Arg258; a segment of the palm near the junction with the fingers that includes Arg149, which binds to the dNTP's triphosphate moiety; and loop II, which is more distant from the active site (see highlighted regions in Figure 4.5[b] labeled A). These palm regions are correlated with areas in the thumb (*e.g.*, residues near Tyr271 and Phe272 in-helix M) as well as the dCTP (see highlighted regions in Figure 4.5[b] labeled B and C). Correlated motions between the palm and thumb, and fingers and palm, highlight the coordination of the subdomains during thumb closing.

Correct dNTP systems of pol X, pol λ, and pol μ show similar areas of correlated motions, but these coupled motions are less extensive than in pol β (see Table 4.1, and Figures 4.6–4.8). Interestingly, analysis of the pol λ I492A mutant simulation, which captures thumb loop and DNA conformational changes, reveals more extensive coupled motions as in pol β Figure 4.7). Comparison of pol λ aligned and misaligned DNA complexes with the correct dNTP show similar correlated motions (Figure 4.7), which agrees with their similar structures and pol λ's efficient handling of both of these types of

Table 4.1 Summary of correlated motions occurring in X-family DNA polymerases.

DNA polymerase	*Correlated motions in correct dNTP systems*	*Correlated motions in incorrect dNTP systems*	*Correlated motions in misaligned DNA systems*
Pol β.	1. Within all subdomains 2. In palm: conserved Asp loop, Asp256–Arg258, loop II, and Argl49 region 3. Between palm/thumb: Argl49 region, conserved Asp loop, and α-helix M 4. With dNTP: all listed palm and thumb regions 5. With DNA: palm and thumb are similar	In G:G system: 1. Fewer correlated motions within and between subdomains, DNA, and dNTP 2. In palm: fewer correlations among conserved Asp loop, loop II, and Argl49 region	NA*
Pol X	1. Within all subdomains, but less than pol β 2. In palm: conserved Asp loop and N-terminus (similar to pol β Argl49 region) 3. Between palm/thumb: conserved Asp loop, C-terminus, and α-helices D & E (like pol β α-helices M & N) 4. With dNTP: α-helices D & E and conserved Asp loop 5. With DNA: more with palm than thumb	1. In C:C system, more widespread correlated motions in palm and palm/thumb; fewer between palm and upstream template/downstream primer paired segment 2. In G:G (*syn*) system, more similar to correct dNTP system, but greater thumb/DNA correlated motions 3. In both, fewer correlated motions between thumb/dNTP	NA

Pol λ	1. Within all subdomains, but less than pol β and pol X 2. In palm: conserved Asp loop, Asp490, loop II, and Arg386 region (similar to pol β Argl49 region) 3. Between palm/thumb: conserved Asp loop, Tyr505–Arg517 in α-helices M & N, and C-terminus 4. With dNTP: all listed palm and thumbregions 5. With DNA: more with thumb than palm	1. A:C system is similar to correct dNTP system 2. A:A system has fewer correlated motions among the conserved Asp loop, Asp490, and loop II as well as fewer protein/dNTP correlated motions than correct dNTP system	Same as pol λ correct dNTP system
I492A Pol λ	Similar areas of correlated motions to wild-type pol λ, but with intensities more like correlated regions in pol β and pol X	NA	NA
Pol μ	1. In palm: conserved Asp loop, loop II, and Pro289 region (similar to pol P Argl49 region and pol λ Arg386 region) 2. Between palm/thumb: conserved Asp loop, Pro289 region, and Gly435–Glu443 in α-helix M & N 3. With dNTP: all listed palm and thumb regions 4. With DNA: more with thumb than palm	NA	NA

*NA, not analyzed

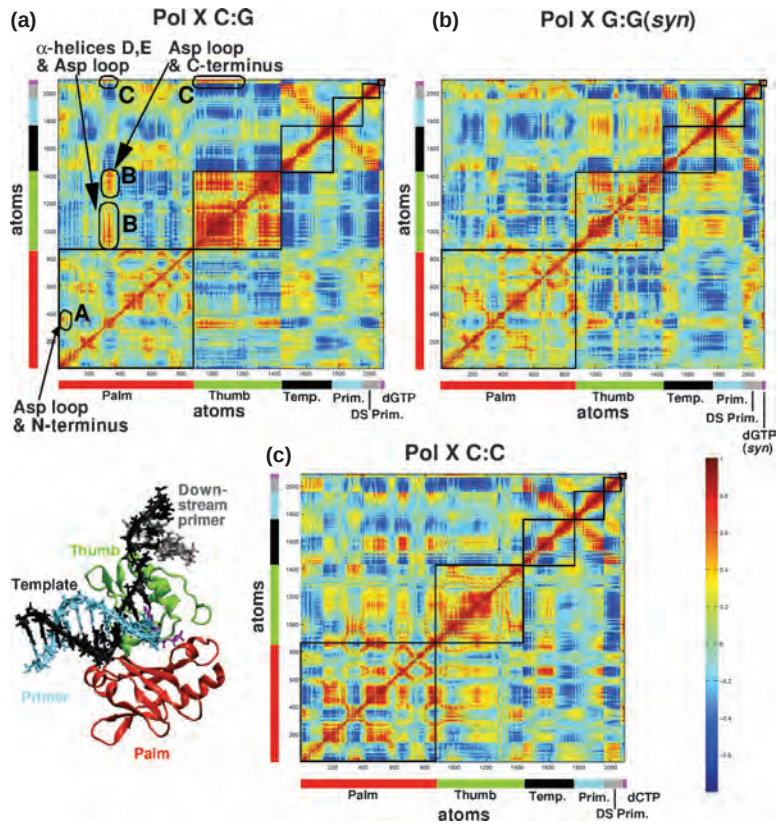

Figure 4.6 Covariance matrix for protein/DNA/dNTP heavy atoms in pol X from simulations with correct and incorrect nucleotides. Colors: purple (8-kDa domain), blue (fingers), red (palm), green (thumb), black (DNA template strand), cyan (DNA primer strand), silver (DNA downstream primer), and magenta (dNTP).

substrates.[97] In these correct pol X/C:G, pol λ and I492A pol λ/A:T, and pol μ/A:T complexes, correlated motions within the palm involve the conserved palm loop containing two catalytic aspartates and areas analogous to pol β's Arg149 region (*i.e.*, pol X's N-terminus, pol λ's Arg386 region, and pol μ's Pro289 region); see A regions in Figures 4.6–4.8. In pol λ and pol μ, palm correlated motions involve loop II as in pol β; pol λ also has correlations with its third catalytic aspartate (Asp490) like pol β. Between the palm and thumb, correlated motions in these polymerases include the conserved aspartate loop with the thumb α-helices analogous to pol β's α-helices M and N; see B' regions in Figures 4.6–4.8. In pol X and pol λ, coupled motions between the palm and thumb also involve the C-terminus whereas, in pol μ, coupled motions also involve the Pro289 region. The I492A pol λ mutant differs from wild-type pol λ in that additional correlated motions occur among the Arg386 palm region, the thumb including α-helices M and N, and

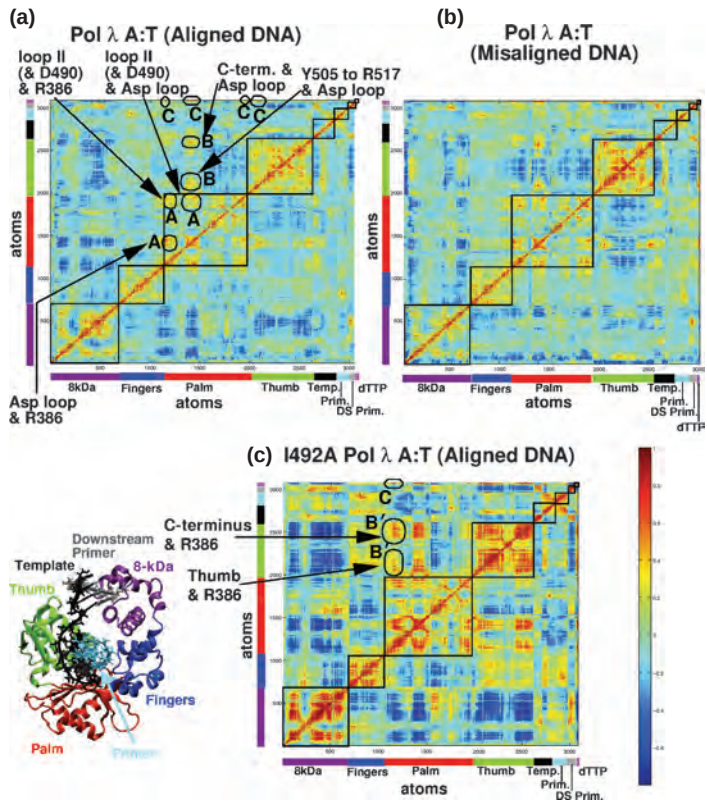

Figure 4.7 Covariance matrix for protein/DNA/dNTP heavy atoms in pol λ from simulations with correct nucleotides in both aligned and misaligned DNA contexts as well as bound to the I492A mutant. Colors: purple (8-kDa domain), blue (fingers), red (palm), green (thumb), black (DNA template strand), cyan (DNA primer strand), silver (DNA downstream primer), and magenta dNTP).

the C-terminus; see B regions in Figure 4.7. As in pol β, all these palm and thumb regions are coupled to the incoming nucleotide; see C regions in Figures 4.6–4.8.

Several of these correlated motion regions have been shown experimentally to affect pol β's function and fidelity. For instance, altering the length of loop II, but not its composition, decreases pol β's fidelity.[99,100] In pol β, Phe272 in α-helix M has also been shown to play a role in fidelity.[101] In addition to changing rates of pol β's opening motion,[14] mutation of Arg258 to alanine can lead to a reduced dNTP binding affinity.[2] Similarly, mutation of nearby Ile260 affects function and can lead to greater misinsertion rates due to higher dNTP binding affinities.[102,103] Mutants of Ile174, which is part of the β-strand attached to the conserved aspartate palm loop, also have higher error rates that are hypothesized to result from changes in protein dynamics.[45] Indeed, all these

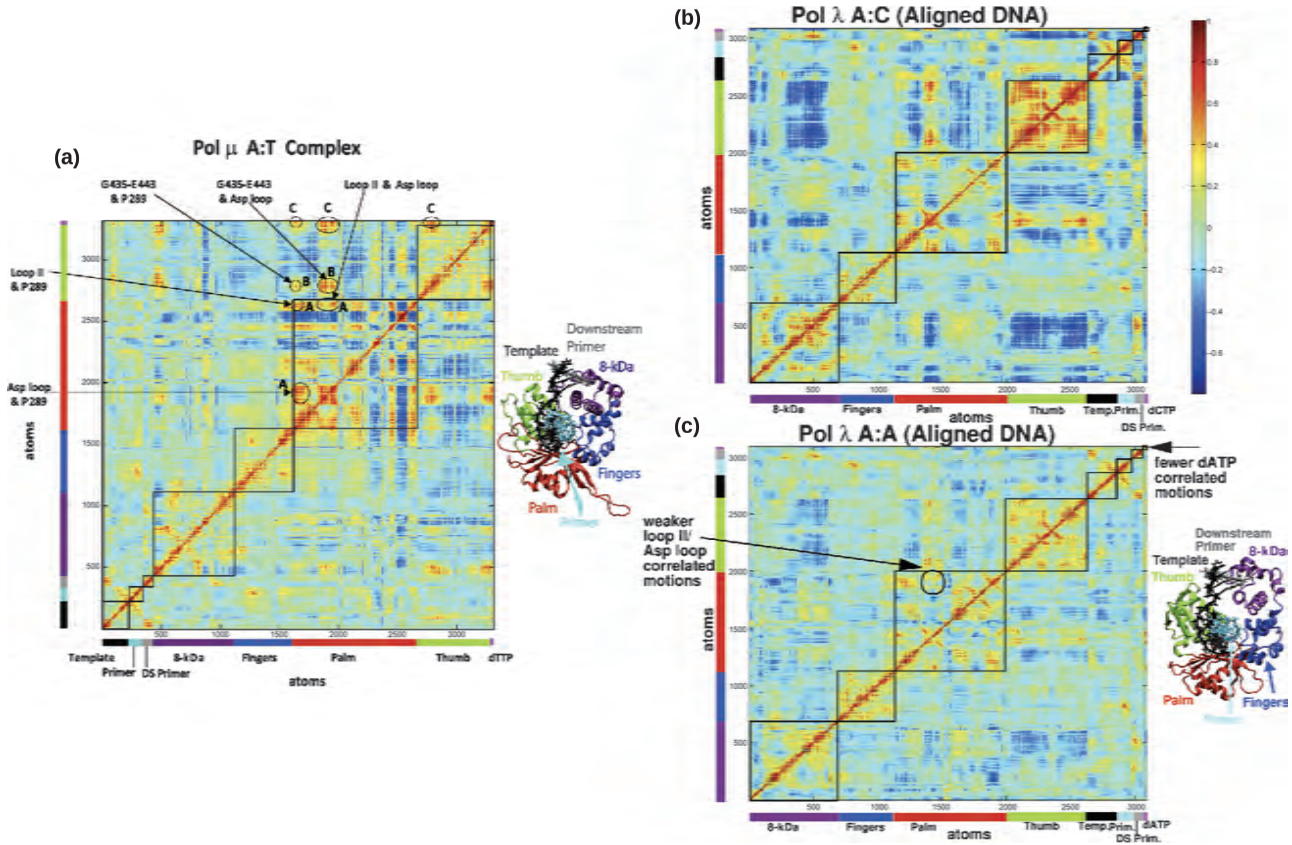

Figure 4.8 Covariance matrices for correct pol μ/A:T system and pol λ mismatch systems. (a) Covariance matrix for protein/DNA/dNTP heavy atoms in pol m simulation with the correct nucleotide. Colors: purple (8-kDa domain), blue (fingers), red (palm), green (thumb), black (DNA template strand), cyan (DNA primer strand), silver (DNA downstream primer), and magenta (dNTP). (b) Covariance matrix for protein/DNA/dNTP heavy atoms in pol λ for incorrect dNTP systems. Colors: purple (8-kDa domain), blue (fingers), red (palm), green (thumb), black (DNA template strand), cyan (DNA primer strand), silver (DNA downstream primer), and magenta (dNTP).

mutations may disrupt dynamic coupling with the active site that is important for function and fidelity.

Based on these connections between pol β residues participating in coupled motions and function/fidelity, we hypothesize that similar relationships may occur in the other X-family polymerases. For example, the shorter length of loop II in pol λ may contribute to its having a higher deletion error rate than pol β since changes to the length of loop II in pol β altered its rate of frameshift mutations.[99] It is also likely that mutations within the conserved aspartate loop, aside from the catalytic aspartates, impair the function and/or fidelity of all these polymerases. Similarly, mutations to the C-terminus of pol X and pol λ may alter the behavior of these polymerases.

From our comparison of correlated motions, differences also emerge in protein/DNA correlated motions of all the X-family polymerases. For example, the I492A pol λ mutant has more thumb/DNA correlated motions than palm/DNA motions; pol X has the opposite; pol β has about the same between the palm or thumb and DNA; and pol μ has greater DNA correlated motions with the thumb than palm. Subtle differences in DNA binding may thus be important for DNA stabilization within the active position. The palm and thumb may better share interactions in pol β, while pol X may depend on its palm and pol λ and pol μ rely on their thumbs to secure the DNA. Pol β's combined palm and thumb interactions with the DNA may hamper large-scale DNA motions. Pol β, I492A pol λ, and pol μ show similar correlated motions between the DNA and the 8-kDa domain, which agrees with the similar DNA binding role proposed for the 8-kDa domain in these enzymes.

4.4.3.2 Dynamic Networks are Perturbed within Incorrect dNTP Complexes

Our comparison of X-family DNA polymerase handling of correct and incorrect dNTPs uncovers trends in the dynamics of DNA polymerase mismatch complexes that may suggest how easily an incorrect dNTP is inserted by a polymerase (see Figure 4.4). Generally, the more closely the thumb, DNA, and active-site conformational changes resemble correct dNTP systems, the more likely the incorrect nucleotide will be inserted. Our covariance analysis of these systems reveals further differences between the dynamics of correct and incorrect dNTP complexes. Pol β bound to the G:G mismatch shows far fewer correlated motions (Figure 4.5[c]). In pol X and pol λ incorrect dNTP complexes, correlated motions also differ from correct dNTP systems (Figures 4.6 and 4.8). In particular, all mismatch complexes show fewer correlated motions between the protein and dNTP. Significantly, the extent of the changes in correlated motion corresponds to the enzyme's efficiency in handling the incorrect dNTP. That is, pol X/G:G(syn) and pol λ/A:C complexes show greater similarity to correct dNTP complexes than pol X/C:C and pol λ/A:A complexes. In pol X, some correlated motions not involving the dNTP are upregulated in mismatch complexes unlike in pol β and pol λ where they are generally reduced.

4.4.3.3 *Potential Roles of Coupled Motions in DNA Polymerase Complexes*

The similarities in correlated motions within pol β, pol X, pol λ, and pol μ despite their different large-scale protein subdomain and DNA motions can likely be explained in part by conserved X-family structural elements. For example, the palm loop joining two β-strands with two catalytic aspartates, in particular, is common to wider-range of nucleotidyl transferases in the X family.[98] The highly efficient propagation of signals within proteins by aspartate residues that bind metal ions[104] agrees with the high degree of correlated motions involving the conserved aspartate loop in these polymerases. We suggest that these coupled motions not only are important for ligand binding, but also play a role in active-site preorganization, and likely affect fidelity. Studies of high-fidelity BF suggest that coupled motions promote enzyme catalysis by driving fluctuations in the distance between the reactive O3′ and P_α atoms.[105] Though the direct role of large-scale enzyme motions in catalysis is an area of debate,[106] with arguments both in favor[107,108] and against it,[109] in DNA polymerases where conformational changes are rate-limiting, a direct relationship to catalysis can be envisioned. However, chemistry rather than conformational changes is hypothesized to be rate-limiting in some DNA polymerases; for many DNA polymerases, the rate-limiting step has not yet been determined.

4.4.4 Hybrid Conformational Selection/Induced-fit Mechanism May Better Account for Intrinsic Polymerase Motions

DNA polymerase ligand interactions have been described most frequently by Koshland's induced-fit concept.[32] As illustrated in Figure 4.9, this model purports that correct substrate binding triggers one conformational change whereas incorrect substrate binding triggers a different conformational change that may be less compatible with catalysis. Conformational selection is an alternative model that provides wider scope for the intrinsic dynamics of biomolecules in ligand recognition/binding events;[33,34,110,111] it has been applied to various systems.[112–115] In this dynamical view (Figure 4.9), the apo (*i.e.*, unliganded) protein traverses through an ensemble of pre-existing conformations on an evolutionarily determined energy landscape, which includes conformations structurally similar to the ligand-bound form. The ligand then selects the favored conformation for binding from this ensemble, causing a population-shift in the ensemble toward the ligand-bound form.

Our correlated motion analysis provides further support for conformational selection playing a role in dNTP binding events since it emphasizes that functionally important motions are encoded by the structures of these X-family polymerases. Our PCA work also underscores the importance of intrinsic large-scale thumb and DNA motions of polymerase/DNA/dNTP complexes in the assembly of the active site before chemistry. When incorrect nucleotides are bound by the polymerase, intrinsic large-scale and correlated motions are hampered, preventing the proper assembly of the active site for catalysis. As the

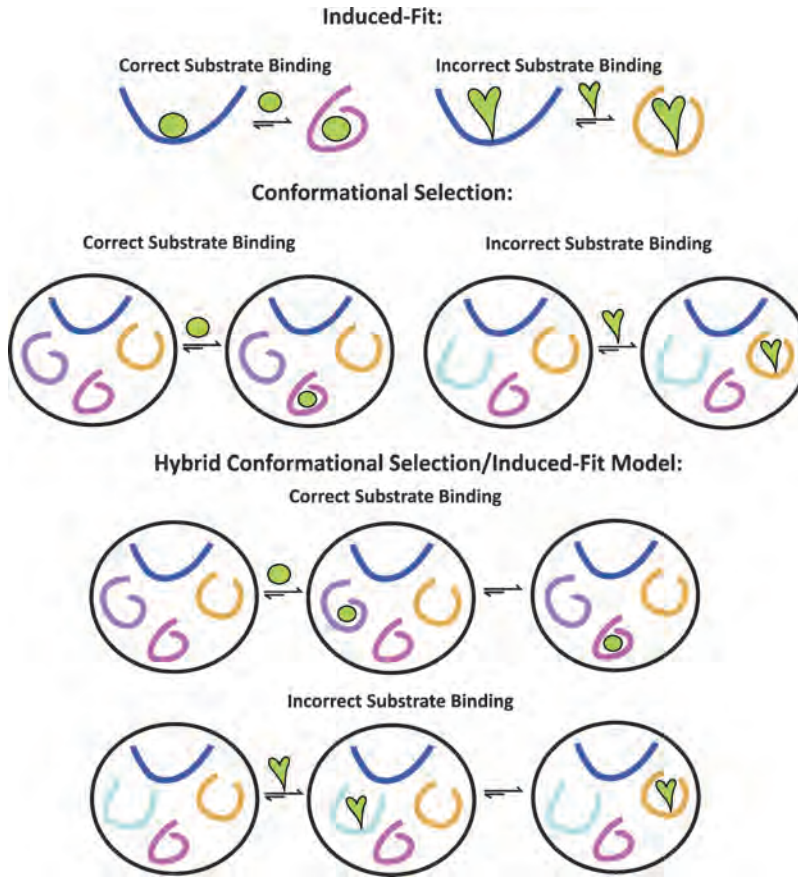

Figure 4.9 Models for protein/ligand binding with correct and incorrect substrates. Protein conformations are indicated by colored curves and ligands are closed green shapes. In induced fit, correct substrate binding triggers a protein conformational change while incorrect substrate binding triggers a different conformational change. Poor fit and interactions of the incorrect substrate require the protein to form a more accommodating conformation. In conformational selection, the correct substrate selectively binds to one conformation from an ensemble of conformations, and shifts the ensemble toward this form. Incorrect substrates selectively bind to a different conformation. In a hybrid conformational selection/induced-fit model, the correct substrate selectively binds to a protein conformation close to the final protein/substrate conformation. The binding of the correct substrate induces the transition to the final conformation. For incorrect substrates, binding selectively occurs to a different conformation and triggers rearrangements that produce a different conformation from that of correct substrate complexes because of poorer interactions between the binding partners. Note, ensembles do not include all possible conformations for simplicity.

likelihood of incorrect nucleotide insertion increases, the motions and active-site assembly more closely resemble the correct case (*e.g.*, G:G (*syn*) by pol X and A:C by pol λ). Pol λ, which handles both aligned and misaligned DNA substrates almost equally well, has nearly identical correlated motions and active-site assemblies.

These intrinsic motions, combined with the dNTP's ability to "fine-tune" these motions and affect active-site assemblies, support the use of a hybrid conformational selection/induced-fit model for DNA polymerase/substrate binding (Figure 4.9) to reconcile the cumulative data. In this model, intrinsic motions of the polymerase/DNA complex such as pol β's and pol X's open-to-closed thumb motion, pol λ's thumb loop and large-scale DNA shifts, and some correlated motions would exist prior to dNTP binding. From this ensemble of possible polymerase/DNA conformations, the correct dNTP selectively binds to a near closed or other active conformation (*e.g.*, pol λ bound to DNA in the active position) that causes a population shift within the ensemble toward this form. The bound dNTP then induces small adjustments in protein side chain (*e.g.*, catalytic aspartate and other active-site residues), nucleotide (*e.g.*, closer positioning of O3′ and P_α atoms), and ion positions within the active site as well as stronger correlated motions within the complex, connecting the movement within the active site to the rest of the polymerase/DNA complex. These changes together result in a closed or active complex ready for catalysis.

Incorrect dNTPs that are relatively efficiently handled by the polymerase (*e.g.*, G:G [*syn*] by pol X and A:C by pol λ) would also selectively bind to a near-closed or active conformation; however, the suboptimal fit of an incorrect dNTP within the active site would induce active-site changes that differ from correct dNTP binding. For example, the pol λ /A:C active site has an additional water molecule and catalytic aspartate residues in different orientations from correct dNTP systems.[31] For incorrect dNTPs that are relatively poorly inserted by a polymerase (*e.g.*, G:G by pol β, C:C by pol X, and A:A by pol λ), dNTP binding may occur to a variable state that better accommodates its poor interactions and fit within the active site. The resulting incomplete organization of the active site and fewer coupled motions between the active site and the rest of the complex would reduce the efficiency for inserting an incorrect dNTP (see disordered active sites in Figure 4.4).

Although our present findings only suggest intrinsic correlated motions within DNA polymerase complexes, the applicability of a hybrid mechanism as described here could be probed through enhanced sampling simulations to provide evidence of minor populations of closed conformations without a dNTP bound or open conformations with the dNTP bound.[112] An induced-fit component to this mechanism would be apparent if active sites only fully assemble in the presence of the correct dNTP. Calculation of the relative free energies of complexes with and without the dNTP in both active and inactive forms would also suggest the relative population of each conformation within ensembles occurring before and after dNTP binding. Similar hybrid mechanisms have been applied to other biomolecules.[115–119] Experimental support for

this model comes from recent single-molecule FRET data showing open and closed forms present with and without the correct dNTP, which suggests that DNA polymerase motions are encoded by their structures rather than triggered upon substrate binding.[120] Other, similar experimental studies show that DNA motions may not be triggered by substrate binding.[121]

4.4.5 Utilizing the New Hybrid DNA Polymerase Mechanism for Therapeutic Purposes

An improved understanding of the interrelation between polymerase dynamics and function could be useful in targeting DNA polymerases for therapeutic purposes. The induced-fit mechanism implies that DNA polymerases are mostly static structures since conformational changes are only stimulated by ligand binding. Thus, a virtual screening for small molecules that only bind to the open and closed crystal states may exclude some functional small molecules. A consideration of the intrinsic motions of DNA polymerases supported by experimental and computational data increases the range of conformations to which small molecules could bind. In addition, utilizing an ensemble of protein conformations in virtual screening may assist in identifying small molecule candidates with improved binding to specific DNA polymerases.

Currently, DNA polymerases are used as drug targets in the treatment of various cancers. DNA polymerases can subvert DNA repair by inserting similar but incorrect nucleoside analogues that cripple or kill cancerous cells.[10] An important class of existing anticancer agents, including azacitidine, gemcitabine, fludarabine, cladribine, and cytarabine, are analogues of correct DNA bases. Interestingly, some of the effectiveness of drugs like azacitidine may lie in their ability to also change DNA methylation patterns through inhibition of DNA methylating enzymes; thus reversing epigenetic changes made by cancers to promote their growth.[122] The effectiveness of such agents in treating cancers is limited by their nonspecific modifications to DNA. The work by Goodman and coworkers focusing on pol β's interactions with nonhydrolyzable dNTPs opens new directions for the development of novel polymerase-specific inhibitors with increased deliverability.[9,123] The inclusion of DNA polymerase-specific dynamics may help design more selective anticancer agents.

DNA polymerases can also be roadblocks to effective cancer therapy[6,124] since they can quickly repair the DNA of cancer cells damaged by chemotherapy and radiation. Because the survival of even a few cancer cells after therapy can lead to a recurrence of the cancer, inhibiting certain polymerases temporarily might be an important step in achieving better long-term results.[125] For example, it has been found that inhibiting pol β can be beneficial when treating colon cancer with the DNA alkylating drug temozolomide.[126]

It has recently been demonstrated in adenylate kinase that dynamics fluctuations can be selectively modified without altering enzyme structure to modulate the binding affinity of small molecules.[127] Thus, the differences in the dynamics of pol β, pol λ, and pol X as revealed by our correlated motions

analysis could be used to selectively target specific polymerases to modulate function. For example, in pol X, decoupling protein interactions from the upstream template and downstream primer may hamper thumb closing as was observed in the C:C mismatch system. This approach could be used to treat African Swine Fever Virus infections by inhibiting the virus' ability to repair its DNA. Similarly, in pol λ, deletion errors could be decreased by increasing DNA coupling to the palm to provide extra stabilization for aligned DNA as in pol β. In contrast, it might be useful to increase the deletion error frequency of an enzyme such as pol β, which has been associated with trinucleotide-repeat disorders like Huntington's disease and fragile X syndrome.[128] Our studies[66] suggest that increasing favorable electrostatic interactions between pol β's thumb and misaligned DNA, possibly through stronger dynamical coupling between the thumb and DNA, may result in more frequent deletion errors. Because of the widespread involvement of defective DNA repair in human diseases, the potential applications of a better understanding of DNA polymerase activity are numerous.

4.5 Conclusion

DNA polymerases are critical components of living cells because of their essential roles in replicating and repairing DNA. Their many evolutionary conserved characteristics from viruses to humans underscores this point. As we decipher the subtle details of how they perform essential tasks like replication and repair, striking similarities in system dynamics can help us understand significant common elements in polymerase mechanisms. In this work, we have examined the motions of various X-family DNA polymerase complexes to better understand their role in function and fidelity. We demonstrated the existence of certain characteristic motions that recur within various dNTP contexts. Specifically, correlated protein and dNTP motions occur within correct dNTP complexes and are altered within incorrect dNTP complexes. The incorrect dNTP complexes suggest that their different characteristic correlated motions depend on the dNTP. When we consider the differences in polymerase subdomains and the variation in functionally important subdomain and DNA motions among these enzymes, the degree of homology in their correlated motions is notable. We propose that these correlated motions derive in part from shared structural motifs common to X-family polymerases. From the effects of known experimental pol β mutations, we suggest that mutations within correlated motion regions have the potential to affect polymerase function and fidelity. Similar intrinsic motions may also occur in other DNA polymerases.

Such intrinsic polymerase motions suggest a broader view to interpret DNA polymerase mechanisms, namely a hybrid conformational selection/induced-fit model for DNA polymerases that better reflects both the intrinsic motions of polymerases and the highly specific nature of polymerase/ligand interactions. This proposed broader view may open new avenues for structure-based drug

design by targeting inherent motions related to polymerase malfunction. Both these ideas can be tested by further computations and experiments.

Acknowledgements

We thank Dr. Benedetta Sampoli Benitez and Mr. Yunlang Li for providing some of the simulation data used in this study. Research described in this article is supported by NSF grant MCB-0316771, NIH grant R01 ES012692, Philip Morris USA Inc. and Philip Morris International, and the American Chemical Society's Petroleum Research Fund award (PRF #39115-AC4) to T. Schlick.

References

1. J. H. J. Hoeijmakers, *Nature*, 2001, **411**, 366–374.
2. K. L. Menge, Z. Hostomsky, B. R. Nodes, G. O. Hudson, S. Rahmati, E. W. Moomaw, R. J. Almassy and Z. Hostomska, *Biochemistry*, 1995, **34**, 15934.
3. W. A. Beard, D. D. Shock, B. J. Vande Berg and S. H. Wilson, *J. Biol. Chem.*, 2002, **277**, 47393–47398.
4. A. K. Showalter and M.-D. Tsai, *J. Am. Chem. Soc.*, 2001, **123**, 1776.
5. K. Shimizu, K. Hashimoto, J. M. Kirchner, W. Nakai, H. Nishikawa, M. A. Resnick and A. Sugino, *J. Biol. Chem.*, 2002, **277**, 37422.
6. L. A. Loeb and J. R. J. Monnat, *Nat. Rev. Genet.*, 2008, **9**, 594.
7. D. Starcevic, S. Dalal and J. B. Sweasy, *Cell Cycle*, 2004, **3**, 998.
8. A. Copani, J. J. Hoozemans, F. Caraci, M. Calafiore, E. S. Van Haastert, R. Veerhuis, A. J. Rozemuller, E. Aronica, M. A. Sortino and F. Nicoletti, *J. Neurosci.*, 2006, **26**, 10949.
9. C. E. McKenna, B. A. Kashemirov, L. W. Peterson and M. F. Goodman, *Biochim. Biophys. Acta*, 2010, **1804**, 1223.
10. A. J. Berdis, *Biochemistry*, 2008, **47**, 8253.
11. K. A. Jhonson, *Annu. Rev. Biochem.*, 1993, **62**, 685.
12. M. R. Sawaya, R. Prasad, S. H. Wilson, J. Kraut and H. Pelletier, *Biochemistry*, 1997, **36**, 11205.
13. W. A. Beard, W. P. Osheroff, R. Prasad, M. R. Sawaya, M. Jaju, T. G. Wood, J. Kraut, T. A. Kunkel and S. H. Wilson, *J. Biol. Chem.*, 1996, **271**, 12141.
14. M. Bakhtina, M. P. Roettger, S. Kumar and M.-D. Tsai, *Biochemistry*, 2007, **46**, 5463.
15. C. M. Joyce and S. J. Benkovic, *Biochemistry*, 2004, **43**, 14317–14324.
16. S. S. Patel, I. Wong and K. A. Johnson, *Biochemistry*, 1991, **30**, 511–525.
17. T. Schlick, R. Collepardo-Guevara, L. A. Halvorsen, S. Jung and X. Xiao, *Quart. Rev. Biophys.*, 2011, **44**, 191–228.
18. M. Karplus and J. Kuriyan, *Proc. Natl. Acad. Sci. USA*, 2005, **102**, 6679–6685.

19. T. Schlick, *Molecular Modeling and Simulation: An Interdisciplinary Guide*, Springer-Verlag, New York, 2010.
20. E. H. Lee, J. Hsin, M. Sotomayor, G. Comellas and K. Schulten, *Structure*, 2009, **17**, 1295–1306.
21. T. Schlick, *F1000 Biology Reports*, 2009, **1**, 51.
22. R. Radhakrishnan and T. Schlick, *Proc. Natl. Acad. Sci. USA*, 2004, **101**, 5970.
23. B. A. Sampoli Benitez, K. Arora, L. Balistreri and T. Schlick, *J. Mol. Biol.*, 2008, **384**, 1086.
24. M. C. Foley, K. Arora and T. Schlick, *Biophys. J.*, 2006, **91**, 3182.
25. B. Sampoli Benitez, K. Arora and T. Schlick, *Biophys. J.*, 2006, **90**, 42.
26. K. Arora and T. Schlick, *J. Phys. Chem. B*, 2005, **109**, 5358–5367.
27. K. Arora, W. A. Beard, S. H. Wilson and T. Schlick, *Biochemistry*, 2005, **44**, 13328.
28. L. Yang, K. Arora, W. A. Beard, S. H. Wilson and T. Schlick, *J. Amer. Chem. Soc.*, 2004, **126**, 8441.
29. K. Arora and T. Schlick, *Biophys. J.*, 2004, **87**, 3088.
30. Y. Li and T. Schlick, *Biophys. J.*, 2010, **99**, 3463.
31. M. C. Foley and T. Schlick, *J. Phys. Chem. B*, 2009, **113**, 13035.
32. D. E. Koshland, *Proc. Natl. Acad. Sci. USA*, 1958, **44**, 98.
33. B. Ma, S. Kumar, C.-J. Tsai and R. Nussinov, *Prot. Engin.*, 1999, **12**, 713.
34. H. Frauenfelder, G. A. Petsko and D. Tsernoglou, *Nature*, 1979, **280**, 558.
35. S. Doublie, S. Tabor, A. M. Long, C. C. Richardson and T. Ellenberger, *Nature*, 1998, **391**, 251.
36. Y. Li, S. Korolev and G. Waksman, *EMBO J.*, 1998, **17**, 7514.
37. S. J. Johnson, J. S. Taylor and L. S. Beese, *Proc. Natl. Acad. Sci. USA*, 2003, **100**, 3895.
38. V. Purohit, N. D. F. Grindley and C. M. Joyce, *Biochemistry*, 2003, **42**, 10200.
39. J. D. Fowler, J. A. Brown, M. Kvaratskhelia and Z. Suo, *J. Mol. Biol.*, 2009, **390**, 368.
40. M. Garcia-Diaz, K. Bebenek, J. M. Krahn, T. A. Kunkel and L. C. Pedersen, *Nat. Struct. Mol. Biol.*, 2005, **12**, 97.
41. M. Delarue, J. B. Boul'e, J. Lescar, N. Expert-Bezancon, N. Jourdan, N. Sukumar, F. Rougeon and C. Papanicolaou, *EMBO J.*, 2002, **21**, 427.
42. J. D. Pata, *Biochim. Biophys. Acta*, 2010, **1804**, 1124.
43. Y. Wang, K. Arora and T. Schlick, *Protein Sci.*, 2006, **15**, 135.
44. A. F. Moon, M. Garcia-Diaz, V. K. Batra, W. A. Beard, K. Bebenek, T. A. Kunkel, S. H. Wilson and L. C. Pedersen, *DNA Repair*, 2007, **6**, 1709.
45. J. Yamtich, D. Starcevic, J. Lauper, E. Smith, I. Shi, S. Rangarajan, J. Jaeger and J. B. Sweasy, *Biochemistry*, 2010, **49**, 2326.
46. J. Yamtich and J. B. Sweasy, *Biochim. Biophys. Acta*, 2010, **1804**, 1136.

47. H. Zang, A. K. Goodenough, J.-Y. Choi, A. Irimia, L. V. Loukachevitch, I. D. Kozekov, K. C. Angel, C. J. Rizzo, M. Egli and F. P. Guengerich, *J. Biol. Chem.*, 2005, **280**, 29750.
48. R. Radhakrishnan, K. Arora, Y. Wang, W. A. Beard, S. H. Wilson and T. Schlick, *Biochemistry*, 2006, **45**, 15142.
49. T. A. Steitz, *J. Biol. Chem.*, 1999, **274**, 17395.
50. A. S. Mildvan, *Proteins: Struct. Funct. Gen.*, 1997, **29**, 401.
51. J. Florian, M. F. Goodman and A. Warshel, *J. Am. Chem. Soc.*, 2003, **125**, 8163.
52. J. Florian, M. F. Goodman and A. Warshel, *Biopolymers*, 2003, **68**, 286.
53. L. Wang, X. Yu, P. Hu, S. Broyde and Y. Zhang, *J. Am. Chem. Soc.*, 2007, **129**, 4731.
54. R. C. Rittenhouse, W. K. Apostoluk, J. H. Miller and T. P. Straatsma, *Proteins: Struct. Funct. Gen.*, 2003, **53**, 667.
55. G. Andres Cisneros, L. Perera, M. Garcia-Diaz, K. Bebenek, T. A. Kunkel and L. G. Pedersen, *DNA Repair*, 2008, **7**, 1824.
56. P. Lin, V. K. Batra, L. C. Pedersen, W. A. Beard, S. H. Wilson and L. G. Pedersen, *Proc. Natl. Acad. Sci. USA*, 2008, **105**, 5670.
57. P. Lin, L. C. Pedersen, V. K. Batra, W. A. Beard, S. H. Wilson and L. G. Pedersen, *Proc. Natl. Acad. Sci. U.S.A.*, 2006, **103**, 13294–13299.
58. Y. G. Abashkin, J. W. Erickson and S. K. Burt, *J. Phys. Chem. B*, 2001, **105**, 287.
59. L. Wang, S. Broyde and Y. Zhang, *J. Mol. Biol.*, 2009, **389**, 787.
60. M. D. Bojin and T. Schlick, *J. Phys. Chem. B*, 2007, **111**, 11244.
61. I. L. Alberts, Y. Wang and T. Schlick, *J. Am. Chem. Soc.*, 2007, **129**, 11100.
62. Y. Wang and T. Schlick, *J. Am. Chem. Soc.*, 2008, **130**, 13240.
63. R. Venkatramani and R. Radhakrishnan, *Prot. Sci.*, 2010, **19**, 815.
64. C. A. Sucato, T. G. Upton, B. A. Kashemirov, V. K. Batra, V. Martinek, Y. Xiang, W. A. Beard, L. C. Pedersen, S. H. Wilson, C. E. McKenna, J. Florian, A. Warshel and M. F. Goodman, *Biochem.*, 2007, **46**, 461.
65. R. Radhakrishnan and T. Schlick, *Biochem. Biophys. Res. Commun.*, 2006, **350**, 521.
66. M. C. Foley, V. A. Padow and T. Schlick, *J. Am. Chem. Soc.*, 2010, **132**, 13403.
67. J. MacKerell, A. D. and N. K. Banavali, *J. Comput. Chem.*, 2000, **21**, 105.
68. B. R. Brooks, C. L. Brooks, J. A. D. MacKerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York and M. Karplus, *J. Comput. Chem.*, 2009, **30**, 1545.
69. L. Yang, W. A. Beard, S. H. Wilson, S. Broyde and T. Schlick, *J. Mol. Biol.*, 2002, **317**, 651.

70. J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale and K. Schulten, *J. Comput. Chem.*, 2005, **26**, 1781.
71. T. A. Darden, D. M. York and L. G. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089.
72. A. Amadei, A. B. M. Linssen and H. J. C. Berendsen, *Proteins*, 1993, **17**, 314.
73. B. L. deGroot, X. Daura, A. E. Mark and H. Grubmüller, *J. Mol. Biol.*, 2001, **309**, 299.
74. D. M. F. van Aalten, B. L. deGroot, J. B. C. Findlay, H. J. C. Berendsen and A. Amadei, *J. Comput. Chem.*, 1997, **18**, 169.
75. W. Weber, H. Demirdjian, R. D. Lins, J. M. Briggs, R. Ferreira and J. A. McCammon, *J. Biomol. Struct. Dyn.*, 1998, **16**, 733.
76. L. Stella, E. E. Iorio, M. Nicotra and G. Ricci, *Proteins*, 1999, **37**, 10.
77. B. L. deGroot, D. M. F. van Aalten, A. Amadei and H. J. C. Berendsen, *Biophys. J.*, 1996, **71**, 1707.
78. M. A. Balsera, W. Wriggers, Y. Oono and K. Schulten, *J. Phys. Chem.*, 1996, **100**, 2567–2572.
79. T. Ichiye and M. Karplus, *Proteins*, 1991, **11**, 205.
80. A. F. Moon, M. Garcia-Diaz, K. Bebenek, B. J. Davis, X. Zhong, D. A. Ramsden, T. A. Kunkel and L. C. Pedersen, *Nat. Struct. Mol. Biol.*, 2007, **14**, 45.
81. K. A. Henzler-Wildman, M. Lei, V. Thai, S. J. Kerns, M. Karplus and D. Kern, *Nature*, 2007, **450**, 913.
82. K. N. Kirouac and H. Ling, *EMBO J.*, 2009, **28**, 1644.
83. R. Jain, D. T. Nair, R. E. Johnson, L. Prakash, S. Prakash and A. K. Aggarwal, *Structure*, 2009, **17**, 974.
84. J. M. Krahn, W. A. Beard and S. H. Wilson, *Structure*, 2004, **12**, 1823.
85. V. K. Batra, W. A. Beard, D. D. Shock, L. C. Pedersen and S. H. Wilson, *Mol. Cell*, 2008, **30**, 315.
86. V. K. Batra, W. A. Beard, D. D. Shock, L. C. Pedersen and S. H. Wilson, *Structure*, 2005, **13**, 1225.
87. R. Radhakrishnan and T. Schlick, *J. Am. Chem. Soc.*, 2005, **127**, 13245.
88. K. Henzler-Wildman and D. Kern, *Nature*, 2007, **450**, 964.
89. Y. Xiang, M. F. Goodman, W. A. Beard, S. H. Wilson and A. Warshel, *Proteins*, 2008, **70**, 231.
90. R. Rucker, P. Oelschlaeger and A. Warshel, *Proteins*, 2010, **78**, 671.
91. L. Yang, W. A. Beard, S. H. Wilson, B. Roux, S. Broyde and T. Schlick, *J. Mol. Biol.*, 2002, **321**, 459.
92. S. Kumar, The Ohio State University, 2008.
93. J. N. Patro, M. Urban and R. D. Kuchta, *Biochemistry*, 2009, **48**, 180.
94. M. Trostler, A. Delier, J. Beckman, M. Urban, J. N. Patro, T. E. Spratt, L. S. Beese and R. D. Kutcha, *Biochemistry*, 2009, **48**, 4633.
95. N. A. Cavanaugh, M. Urban, J. Beckman, T. E. Spratt and R. D. Kuchta, *Biochemistry*, 2009, **48**, 3554.

AQ1

96. M. Delarue and Y.-H. Sanejouand, *J. Mol. Biol.*, 2002, **320**, 1011.
97. M. Garcia-Diaz, K. Bebenek, J. M. Krahn, L. C. Pedersen and T. A. Kunkel, *Cell*, 2006, **124**, 331.
98. L. Aravind and E. V. Koonin, *Nucl. Acids Res.*, 1999, **27**, 1609.
99. G. C. Lin, J. Jaeger, K. A. Eckert and J. B. Sweasy, *DNA Repair*, 2009, **8**, 182.
100. G. C. Lin, J. Jaeger and J. B. Sweasy, *Nucl. Acids Res.*, 2007, **35**, 2924.
101. S.-X. Li, J. A. Vaccaro and J. B. Sweasy, *Biochemistry*, 1999, **38**, 4800.
102. S. Dalal, D. Starcevic, J. Jaeger and J. B. Sweasy, *Biochem.*, 2008, **47**, 12118.
103. D. Starcevic, S. Dalal and J. Sweasy, *Biochem.*, 2005, **44**, 3775.
104. A. Dutta and I. Bahar, *Structure*, 2010, **18**, 1140.
105. R. Venkatramani and R. Radhakrishnan, *Proteins*, 2008, **71**, 1360.
106. V. Nashine, S. Hammes-Schiffer and S. J. Benkovic, *Curr. Opin. Chem. Biol.*, 2010, **14**, 644.
107. E. Z. Eisenmesser, O. Millet, W. Labeikovsky, D. M. Korzhnev, M. Wolf-Watz, D. A. Bosco, J. J. Skalicky, L. E. Kay and D. Kern, *Nature*, 2005, **438**, 117.
108. Z. D. Nagel and J. P. Klinman, *Chem. Rev.*, 2010, **110**, PR41.
109. S. C. Kamerlin and A. Warshel, *Proteins*, 2010, **78**, 1339.
110. R. H. Austin, K. W. Beeson, L. Eisenstein, H. Frauenfelder and I. C. Gunsalus, *Biochemistry*, 1975, **14**, 5355.
111. B. F. Volkman, D. Lipson, D. E. Wemmer and D. Kern, *Science*, 2001, **291**, 2429.
112. K. Arora and C. L. Brooks, *Proc. Natl. Acad. Sci. USA*, 2007, **104**, 18496.
113. B. J. Grant, J. A. McCammon and A. A. Gorfe, *Biophys. J.*, 2010, **99**, L87.
114. C. G. Kalodimos, N. Biris, A. M. J. J. Bonvin, M. M. Levandoski, M. Guennuegues, R. Boelens and R. Kaptein, *Science*, 2004, **305**, 386.
115. O. F. Lange, N.-A. Lakomek, C. Fares, G. F. Schroder, K. F. A. Walter, S. Becker, J. Meiler, H. Grubmuller, C. Griesinger and B. L. de Groot, *Science*, 2008, **320**, 1471.
116. D. D. Boehr, R. Nussinov and P. E. Wright, *Nat. Chem. Biol.*, 2009, **5**, 789.
117. D. Tobi and I. Bahar, *Proc. Natl. Acad. Sci. USA*, 2005, **102**, 18908.
118. R. Grunberg, J. Leckner and M. Nilges, *Structure*, 2004, **12**, 2125.
119. T. Wlodarski and B. Zagrovic, *Proc. Natl. Acad. Sci. USA*, 2009, **106**, 19346.
120. Y. Santoso, C. M. Joyce, O. Potapova, L. Le Reste, J. Hohlbein, J. P. Torella, N. D. F. Grindley and A. N. Kapanidis, *Proc. Natl. Acad. Sci. USA*, 2010, **107**, 715.
121. P. J. Rothwell, S. Berger, O. Kensch, S. Felekyan, M. Antonik, B. M. Wohrl, T. Restle, R. S. Goody and C. A. M. Seidel, *Proc. Natl. Acad. Sci. USA*, 2003, **100**, 1655.
122. J. Kaiser, *Science*, 2010, **330**, 576.

123. G. K. Surya Prakash, M. Zibinsky, T. G. Upton, B. A. Kashemirov, C. E. McKenna, K. Oertell, M. F. Goodman, V. K. Batra, L. C. Pedersen, W. A. Beard, D. D. Schock, S. H. Wilson and G. A. Olah, *Proc. Natl. Acad. Sci. USA*, 2010.
124. J. H. J. Hoeijmakers, *N. Engl. J. Med.*, 2009, **361**, 1475.
125. A. Venkitaraman, *Cancer Cell*, 2009, **16**, 89.
126. A. S. Jaiswal, S. Banerjee, H. Panda, C. D. Bulkin, T. Izumi, F. H. Sarkar, D. A. Ostrov and S. Narayan, *Mol. Cancer Res.*, 2009, **7**, 1973.
127. T. P. Schrank, D. W. Bolen and V. J. Hilser, *Proc. Natl. Acad. Sci. USA*, 2009, **106**, 16984.
128. I. V. Kovtun, Y. Liu, M. Bjoras, A. Klungland, S. H. Wilson and C. T. McMurray, *Nature*, 2007, **447**, 447.