Received: 5 February 2021 | Revised: 18 March 2021 | Accepted: 19 March 2021

DOI: 10.1002/wcms.1534

ADVANCED REVIEW

WIREs COMPUTATIONAL MOLECULAR SCIENCE WILEY

Computational investigations of polymerase enzymes: Structure, function, inhibition, and biotechnology

Inacrist Geronimo | Pietro Vidossich | Elisa Donati | Marco De Vivo

Laboratory of Molecular Modelling and Drug Discovery, Istituto Italiano di Tecnologia, Genoa, Italy

Correspondence
Marco De Vivo, Laboratory of Molecular Modelling & Drug Discovery, Istituto Italiano di Tecnologia, Via Morego 30, Genoa 16163, Italy.
Email: marco.devivo@iit.it

Funding information
Associazione Italiana per la Ricerca sul Cancro, Grant/Award Number: IG 23679

Abstract

DNA and RNA polymerases (Pols) are central to life, health, and biotechnology because they allow the flow of genetic information in biological systems. Importantly, Pol function and (de)regulation are linked to human diseases, notably cancer (DNA Pols) and viral infections (RNA Pols) such as COVID-19. In addition, Pols are used in various applications such as synthesis of artificial genetic polymers and DNA amplification in molecular biology, medicine, and forensic analysis. Because of all of this, the field of Pols is an intense research area, in which computational studies contribute to elucidating experimentally inaccessible atomistic details of Pol function. In detail, Pols catalyze the replication, transcription, and repair of nucleic acids through the addition, via a nucleotidyl transfer reaction, of a nucleotide to the 3′-end of the growing nucleic acid strand. Here, we analyze how computational methods, including force-field-based molecular dynamics, quantum mechanics/molecular mechanics, and free energy simulations, have advanced our understanding of Pols. We examine the complex interaction of chemical and physical events during Pol catalysis, like metal-aided enzymatic reactions for nucleotide addition and large conformational rearrangements for substrate selection and binding. We also discuss the role of computational approaches in understanding the origin of Pol fidelity—the ability of Pols to incorporate the correct nucleotide that forms a Watson–Crick base pair with the base of the template nucleic acid strand. Finally, we explore how computations can accelerate the discovery of Pol-targeting drugs and engineering of artificial Pols for synthetic and biotechnological applications.

This article is categorized under:

Structure and Mechanism > Reaction Mechanisms and Catalysis
Structure and Mechanism > Computational Biochemistry and Biophysics
Software > Molecular Modeling

KEYWORDS

enzymatic catalysis, molecular simulation, polymerase function, reaction mechanism

1 | INTRODUCTION

Nucleic acid polymerases (Pols) are essential to the survival and propagation of life.[1,2] These enzymes are responsible for DNA and RNA replication, DNA damage repair, transcription (DNA → RNA), and reverse transcription (RNA → DNA). The discovery of these vital enzymes and their critical functional implications have indeed been recognized by several Nobel prizes. Starting from the 1959 Nobel Prize in Physiology/Medicine awarded to Severo Ochoa and Arthur Kornberg "*for their discovery of the mechanisms in the biological synthesis of RNA and DNA*",[3] to the one in 2006 in Chemistry awarded to Roger D. Kornberg "*for his studies of the molecular basis of eukaryotic transcription*",[4] until the most recent one in 2015, again in Chemistry, received by Tomas Lindahl, Paul Modrich, and Aziz Sancar "*for having mapped and explained how the cell repairs its DNA and safeguards the genetic information*."[5]

Because of the central role of Pols in the transfer of genetic information, gene mutations or changes in the expression level have been implicated in various human diseases.[6,7] For example, POLH (gene encoding DNA polymerase [DNA Pol] η) mutations are linked to a variant type of xeroderma pigmentosum (XP-V),[8] POLB (DNA Pol β) mutations are linked to adenocarcinoma of the colon,[9] and POLG (DNA Pol γ) mutations to Alpers–Huttenlocher syndrome, childhood myocerebrohepatopathy spectrum, myoclonic epilepsy myopathy sensory ataxia, progressive external ophthalmoplegia, ataxia neuropathy spectrum, and other mitochondrial diseases.[10,11] On the other hand, overexpression of POLK (DNA Pol κ) has been observed in lung cancer[12] and POLQ (DNA Pol θ) in colorectal,[13] breast,[14] and nonsmall cell lung cancers.[15] Thus, targeting Pols with small-molecule inhibitors is an attractive therapeutic approach.[16–18] This also applies to the case of viral diseases, such as the common cold,[19] hepatitis C virus infection,[20] and severe acute respiratory syndrome (SARS),[21] because RNA polymerases (RNA Pols) are needed for the replication of viral RNA and thus involved in the genetic variability of RNA viruses.[22,23] At present, there is significant focus on the RNA-dependent RNA Pol of SARS coronavirus 2 (SARS-CoV-2)[24] in the search for effective treatments for coronavirus disease 2019 (COVID-19), which has caused the current pandemic and global health, social, and economic crisis.

Importantly, Pols are also utilized in biotechnological applications,[25–28] including the Nobel-Prize-winning (1993, Chemistry) polymerase chain reaction (PCR) method invented by Kary B. Mullis.[29,30] The PCR method is used for applications such as DNA amplification,[31] site-directed mutagenesis,[32] diagnosis of infectious diseases,[33] and analyses of clinical,[34] forensic,[35] and environmental[36] samples. Pols are also applied in synthetic biology to build synthetic genetic polymers for the storage and propagation of genetic information.[25–28] Examples of synthetic nucleotides (i.e., xeno-nucleic acids) include threose nucleic acids,[37] 1,5-anhydrohexitol nucleic acids,[38] and locked nucleic acids.[39] Additionally, Pols can be used to produce mirror-image polymers (i.e., L-nucleic acid polymers), which would not be recognized by the immune system and could evade nucleases,[40,41] and artificially expanded genetic information systems, which expand the genetic alphabet through the inclusion of unnatural base pairs (UBPs).[42,43] These applications in biotechnology and synthetic biology require Pols that can tolerate extreme conditions, resist inhibitors, or incorporate synthetic or modified nucleotides.[26] Strategies to produce such engineered Pols include random and targeted mutagenesis, directed evolution (e.g., compartmentalized self-replication and droplet-based optical polymerase sorting), and domain replacement.[25,26]

With the significant and continuous advances in the algorithms and performance of molecular modeling techniques,[44–46] computational approaches have become an increasingly used tool in Pol research. In this review, we highlight the role of computational studies in providing a molecular-level understanding of the Pol catalytic mechanism and properties of Pols targeted for therapeutics, biotechnology, and synthetic biology. In Section 2, we define the properties that are critical to the specific functional roles of Pols, namely, fidelity and processivity, and provide an overview of the structure and catalytic mechanism of Pols. In Section 3, we discuss key findings from structural studies that have significantly advanced our understanding of the Pol catalytic mechanism and functional properties. In Section 4, we highlight the contribution of computational studies in resolving the open questions on the catalytic mechanism and origin of Pol fidelity. In Section 5, we give a few examples of the contribution of computational studies to the development of Pol technology. Finally, in the conclusion, we provide an outlook on important and challenging Pol research areas that can benefit from computational approaches.

2 | FUNDAMENTALS OF POLYMERASES

2.1 | Polymerase properties and classification

DNA and RNA synthesis by Pols involves the template-guided addition of deoxyribonucleoside triphosphate (dNTP) or ribonucleoside triphosphate (NTP), respectively, to the 3′-end of the growing primer via a nucleotidyl transfer reaction.

Accurate and rapid D(R)NA synthesis depends on two important properties of Pols: fidelity and processivity. Fidelity is the ability of a Pol to select the correct (d)NTP from solution that forms a complementary Watson–Crick (WC) base pair with the templating base (G–C and A–T [DNA] or A–U [RNA] base pair). The fidelity or error rate can be quantified as the ratio of the catalytic efficiencies for incorrect and correct nucleotide incorporation, that is, $(k_{pol}/K_d)_{incorrect}/(k_{pol}/K_d)_{correct}$ in presteady-state kinetics or $(k_{cat}/K_M)_{incorrect}/(k_{cat}/K_M)_{correct}$ in steady-state kinetics.[47] This intrinsic fidelity can be further improved by the proofreading or exonuclease activity of the Pol. On the other hand, processivity is the average number of nucleotides (nt) incorporated during a single binding event and is calculated as the ratio of k_{pol} to the rate of dissociation of the Pol from the primer-template (k_{off}).[48]

On the basis of sequence homology, DNA Pols from the three domains of life can be classified into seven families: A, B, C, D, X, Y, and reverse transcriptase (RT).[47] These different DNA Pol families vary significantly in fidelity and processivity depending on their functional roles. For instance, A- and B-family DNA Pols, which are involved in DNA lesion bypass and mitochondrial DNA replication and repair,[49] have the highest fidelity (10^{-5} to 10^{-6} without exonuclease activity)[50,51] and highest processivity (>100 nt/s).[25] Thermostable DNA Pols from these families, such as DNA Pol I (A-family) from *Thermus aquaticus* (Taq), are also employed in PCR applications.[25] Y-family DNA Pols, which specialize in the bypass of lesions such as cyclobutane thymine-thymine dimer and 7,8-dihydro-8-oxo-2′-deoxyguanosine (8-oxoG) during DNA synthesis, have the lowest fidelity (10^{-1} to 10^{-3})[50,51] and lowest processivity (<20 nt/s).[48] A notable example is DNA Pol η, which has also been implicated in somatic hypermutation[52] and chemoresistance to anticancer agents like cisplatin.[53]

Archaea and Bacteria have only one class of RNA Pols to transcribe cellular genomes, while Eukarya has three classes of RNA Pols: (1) RNA Pol I, which transcribes ribosomal RNA genes, (2) RNA Pol II, which synthesizes messenger RNA and a subset of small noncoding RNAs, and (3) RNA Pol III, which synthesizes transfer RNAs, 5S RNA, and the majority of small noncoding RNAs.[54] Because RNA Pol II transcribes protein-encoding genes, its fidelity is very critical to the eukaryotic life cycle. This RNA Pol reportedly discriminates against the incorrect NTP, as well as dNTP, with a transcription error rate of 10^{-5}.[55,56] On the other hand, viral RNA Pols are medium-fidelity Pols with mutation frequencies of 10^{-3} to 10^{-5}, which allow RNA viruses to quickly adapt to different host cell environments.[22,57] However, these mutation rates can be reduced to as low as 10^{-7} in coronaviruses because of the proofreading activity of $3' \rightarrow 5'$ exoribonucleases.[57]

Taken together, all these results and experimental evidence suggest that Pol fidelity and processivity can be modulated by numerous chemical variables through different mechanisms.[48] As we will be discussing here, general hypotheses to explain Pol fidelity and processivity can be derived and tested on the basis of computations integrated with experimental data.

2.2 | Structural domains of polymerases

DNA Pols are structurally diverse in accordance with the differences in their functional roles and properties.[49] For example, A- and B-family DNA Pols have an exonuclease $3' \rightarrow 5'$ proofreading domain that enhances their fidelity by as much as 100-fold (10^{-7} to 10^{-8}).[50,51] X-family DNA Pols have an additional lyase domain involved in the excision of apyrimidinic sites.[58] Y-family DNA Pols have a smaller structure with fewer contacts with the incoming dNTP and DNA, which contributes to their low processivity.[59] Because they have a more open and solvent-exposed active site that can accommodate modified template bases, nucleotide adducts, and noncomplementary nucleotides, they also exhibit low fidelity. Nevertheless, most DNA Pols typically consist of a catalytic domain resembling a right hand (Figure 1a).[49,50] On the basis of this architecture, the catalytic domain can be further subdivided into palm, fingers, and thumb subdomains, although only the first two are present in DNA Pol X from African swine fever virus. The palm subdomain contains the catalytic residues, while the fingers and thumb subdomains interact with the incoming dNTP and DNA substrates, respectively. The conventional right-hand-based nomenclature, however, is the opposite of the one originally proposed for the X-family DNA Pol, DNA Pol β (i.e., the fingers and thumb subdomains were swapped).[58] To avoid confusion, a function-based nomenclature is adopted for this family: the catalytic domain is referred to as the polymerase domain and the subdomains as the D (DNA binding), C (catalytic), and N (nascent base-pair binding) subdomains.

In comparison, RNA Pols are more structurally conserved. Viral single-subunit DNA- and RNA-dependent RNA Pols also adopt the right-hand architecture with seven conserved structural motifs: motifs A–E are located in the palm subdomain, while motifs F and G are located in the fingers subdomain.[64] On the other hand, DNA-dependent RNA

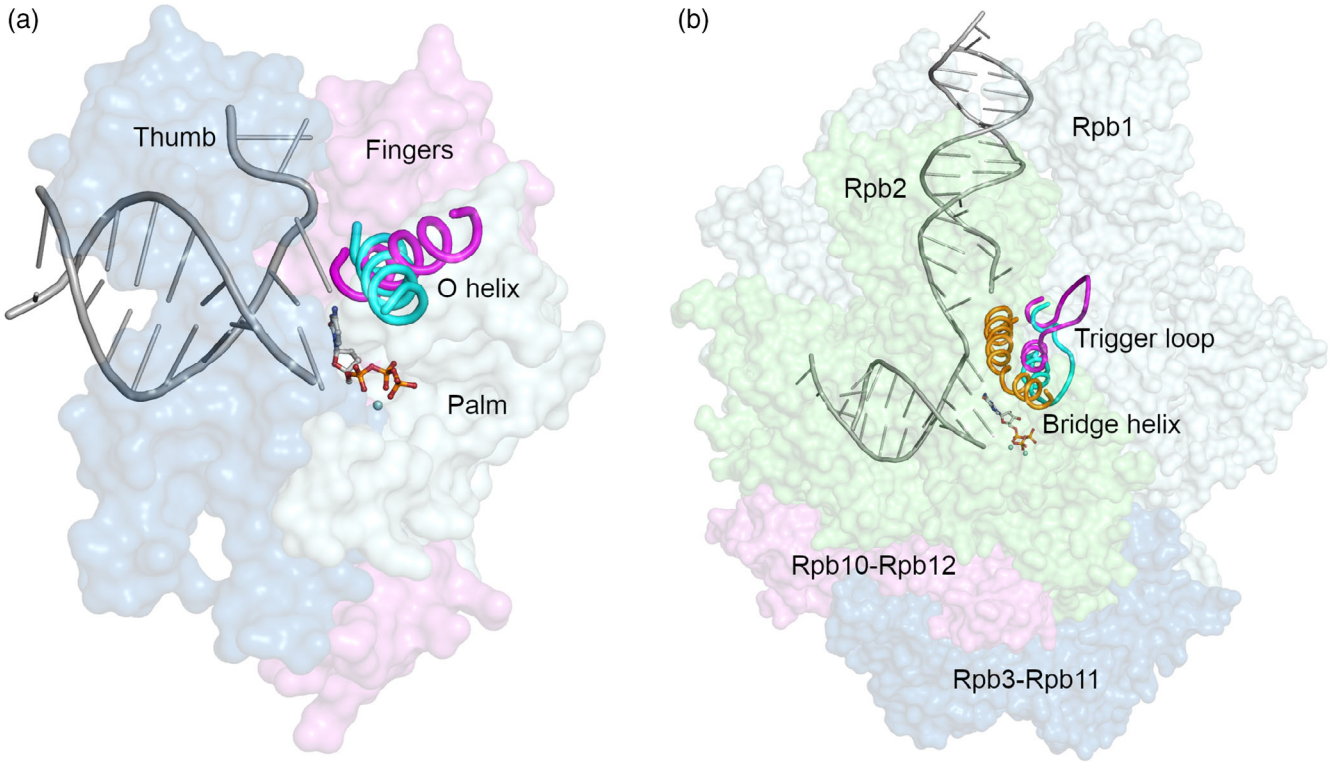

FIGURE 1 Crystal structures of (a) DNA polymerase (Pol) I (PDB ID 2HVI[60]) and (b) RNA Pol II (PDB ID 2E2H[61]). DNA Pols, as well as viral RNA Pols, adopt a right-hand architecture consisting of palm, fingers, and thumb subdomains. Archaeal, bacterial, and eukaryotic RNA Pols have multiple subunits with the catalytic subunits (Rpb1 and Rpb2) and assembly platform (Rpb3-Rpb11 and Rpb10-Rpb12) forming the minimal configuration capable of RNA polymerization. Pols may undergo an open → closed conformational transition upon nucleotide binding. The O helix (DNA Pol, A) and trigger loop (RNA Pol, B) are shown in both their open (magenta, PDB IDs 1L3U[62] and 1Y1V,[63] respectively) and closed (cyan, PDB IDs 2HVI and 2E2H, respectively) conformations. The bridge helix (orange) of RNA Pol does not significantly change in conformation upon nucleotide binding

Pols from the three domains of life have multiple subunits (Figure 1b).[65] The subunits of bacterial RNA Pols are named using Greek letters, those of archaeal RNA Pols are named Rpo followed by a number, and those of eukaryotic RNA Pols are named Rpb, again followed by a number. Using the Rpb nomenclature, the conserved core of multi-subunit RNA Pols consists of Rpb1 and Rpb2, which carry out NTP incorporation and translocation, and Rpb3, Rpb11, and Rpb6, which are involved in enzyme assembly and transcription regulation. The catalytic subunits Rpb1 and Rpb2 contain the catalytic residues, bridge and trigger helices, binding sites for downstream DNA and RNA–DNA hybrid, secondary NTP entry channel, and loop and switch regions responsible for handling the nucleic acid scaffold. The assembly platform is formed by the complex of Rpb10–Rpb12 and Rpb3–Rpb11. Together, the catalytic subunits and assembly platform constitute the minimal configuration of multi-subunit RNA Pols capable of RNA polymerization.

2.3 | Catalytic cycle of polymerases

Despite their structural diversity, Pols follow a common catalytic mechanism for D(R)NA synthesis, which consists of not only the chemical reaction itself, but also physical steps including substrate binding, conformational change, and translocation along the primer-template (Figure 2). The catalytic cycle of RNA Pols consists of initiation, elongation (i.e., NTP addition and translocation), and termination phases.[65] In this regard, this review examines the elongation phase of RNA synthesis, which is analogous to the DNA synthetic process in DNA Pol.

2.3.1 | Nucleotide addition cycle

The catalytic cycle of DNA Pols begins with binding to the DNA substrate.[48] The dNTP then binds to the resulting DNA/Pol binary complex to form the initial DNA/Pol/dNTP ternary complex (Figure 2). This may be followed by an

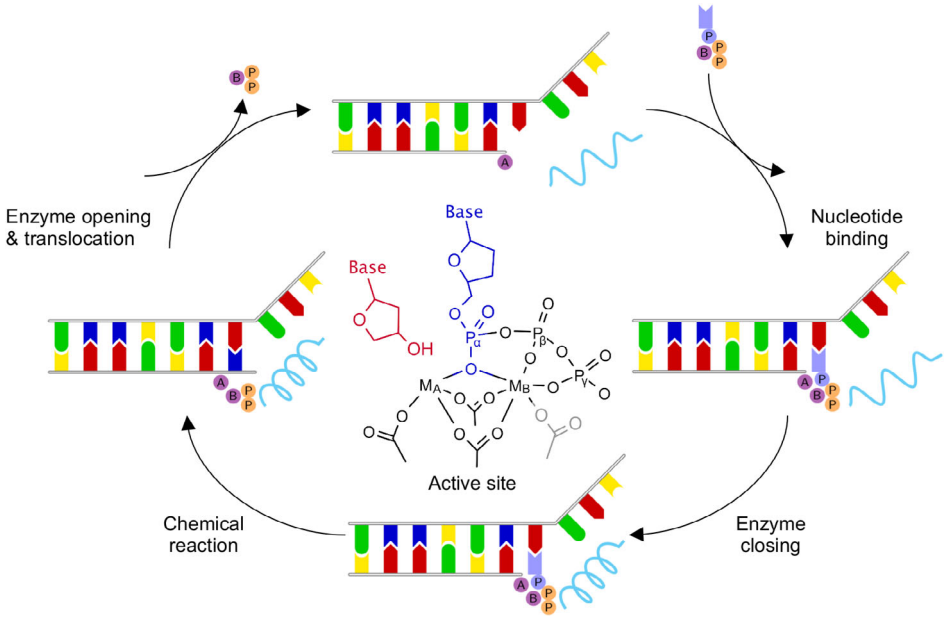

FIGURE 2 Nucleotide addition cycle of polymerases (Pols). The active site is shown in the center. Multi-subunit RNA Pols often have an additional conserved acidic residue (gray) coordinated exclusively to the B-site metal (M_B). In the chemical step, the 3′-OH group of the primer terminus (red) is deprotonated and attacks the P_α atom of the incoming nucleotide (blue)

open → closed conformational change (mainly of the fingers or N-subdomain, Figure 1a) and local active site rearrangement. These structural changes render a reactive ternary complex wherein the dNTP is aligned with the 3′-end of the primer DNA for nucleophilic attack (Figure 2). Y-family DNA Pols, however, do not undergo any global conformational change because the protein is already prealigned for dNTP binding and catalysis.[66] After nucleotide incorporation, a pyrophosphate (PPi) group is released and a reverse conformational change occurs (Figure 2).[48] However, the order of these two events has not been definitely established.[67,68] Subsequently, the DNA Pol translocates by one base pair along the DNA, leaving the active site free for the binding and incorporation of the next dNTP.[48] Alternatively, the DNA Pol can dissociate from the extended DNA and bind to a new DNA substrate.

The elongation phase of RNA synthesis is similar to the catalytic cycle of DNA Pols.[69,70] The NTP binds to the transcription elongation complex, which consists of the open-conformation RNA Pol and the RNA duplex or RNA–DNA hybrid formed in the initiation phase (Figure 2). The mechanism for this step slightly differs between the RNA Pol classes. The single-subunit RNA Pol from bacteriophage T7 (T7 RNA Pol) has separate preinsertion and insertion sites for the substrate,[69] while in multi-subunit RNA Pols, the positions of the substrate in the preinsertion and insertion states nearly overlap.[71] However, a template-independent entry site, where the NTP initially binds before moving to the insertion site, has been proposed for multi-subunit RNA Pols.[72] Unlike these DNA-dependent RNA Pols, viral RNA-dependent RNA Pols do not have a distinct entry or preinsertion site.[73] Once the NTP moves to the insertion site and establishes WC H-bond interactions (Figure 2), the active site closes through the folding of the trigger loop (or rotation of the O helix in single-subunit RNA Pols, Figure 1), leading to the formation of a reactive configuration.[69,70] Subsequently, the nucleotide is incorporated, PPi is released, the active site re-opens, and the RNA Pol translocates along the RNA duplex or RNA–DNA hybrid (Figure 2). RNA Pols continue to transcribe until it receives a termination signal (termination phase), and unlike DNA Pols, cannot re-associate with the DNA template or RNA once it dissociates.[74]

2.3.2 | Active site and chemical reaction mechanism

Pols have been shown to follow the two-metal-ion mechanism for nucleotidyl transfer (Figure 2, center).[75] The two-metal-ion mechanism is also employed by other nucleic-acid-processing enzymes, such as type II topoisomerases, ribonuclease H enzymes, and endonucleases, for DNA and RNA cleavage.[76–79] The metal ion at the A-site (M_A) is coordinated to two or three conserved acidic residues (Asp or Glu), two of which are also coordinated to the metal ion at the B-site (M_B).[50,64] In the case of multi-subunit RNA Pols, another conserved acidic residue is coordinated exclusively

to M_B.[54] Additionally, M_B is coordinated to the triphosphate group of the incoming (d)NTP and in some cases, a backbone carbonyl oxygen.[50,54,64] The reaction begins with the deprotonation of 3′-OH to form the nucleophile, which then attacks the P_α atom of (d)NTP in an S_N2-like reaction to form a pentacovalent phosphate transition state.[75] A new bond between the O3′ and P_α atoms is formed, thereby extending the primer by one nucleotide. Concurrently, the phosphodiester bond between the α- and β-phosphates of (d)NTP is broken, leading to the release of PPi. PPi has been suggested to undergo protonation by a general acid during the reaction on the basis of solvent deuterium isotope effect data for various DNA Pols and RNA Pols.[80]

3 | KEY STRUCTURAL STUDIES OF POLYMERASES

The catalytic cycle and chemical reaction mechanism of Pols discussed in Section 2 have been established primarily through X-ray crystallography and cryogenic electron microscopy (cryo-EM) studies. In this section, we briefly discuss landmark structural studies of Pols that not only provided a detailed picture of the catalytic cycle, but also served as a framework for computational studies aimed at addressing unresolved questions regarding the catalytic cycle and determining the contributing factors to Pol catalytic efficiency, fidelity, and processivity.

3.1 | Conformational changes for catalysis

DNA Pol β, of the X-family DNA Pols, has served as the model enzyme for structural studies of the global and local conformational changes in DNA Pols during the catalytic cycle. For this DNA Pol, at least one crystal structure for each catalytic step has been solved. The apoenzyme structure shows an extended protein conformation with the lyase domain positioned away from the polymerase domain (PDB ID 1BPD).[81] In complex with single-nucleotide-gapped DNA, the enzyme adopts a doughnut-like conformation as the lyase domain interacts with the N-subdomain (PDB ID 3ISB).[82] The dNTP/DNA/Pol ternary complex structure indicates a transition from an open to a closed conformation upon dNTP binding through the rotation of the N-subdomain (PDB ID 2FMS).[83] This conformational change is accompanied by local rearrangements that bring the active site to a reactant state, namely, the movements of R283 to H-bond with the templating base, conserved acidic residues (D190, D192, and D256) to coordinate to M_A, and N279 and R183 to H-bond with the dNTP. The structure of the ternary product complex with PPi suggests that the enzyme remains in the closed conformation after the reaction (PDB ID 4KLL),[84] while the structure of the binary product complex with only the nicked product DNA (PDB ID 1BPZ) shows a return to the open conformation once PPi is released.[85]

3.2 | Phosphodiester bond formation for nucleotide insertion

Noteworthy, the time-resolved X-ray crystallography study of DNA Pol η allowed the first real-time monitoring of phosphodiester bond formation.[86] Nucleotidyl transfer was initiated by immersing a nonreactive ground-state crystal of the dNTP/DNA/Pol ternary complex (with only an inert Ca^{2+} ion in the B-site, PDB ID 4ECQ) in 1 mM aqueous Mg^{2+}. The proper alignment of the primer 3′-OH and dNTP was observed once both metal sites were occupied by Mg^{2+} (PDB ID 4ECR). The α-phosphate was refined in a pentacovalent transition state with inverted configuration at the peak of bond formation (PDB ID 4ECV), thereby confirming the proposed S_N2-like reaction. The same experimental technique was employed for DNA Pol β to compare phosphodiester bond formation in the matched and mismatched complexes.[84] In the mismatched complex (PDB ID 4LVS), the template strand was observed to shift upstream (relative to its position in the matched complex, PDB ID 4KLD) with concomitant rotation of the primer 3′-OH away from M_A, resulting in a nonreactive ground state. Although the reactant state with M_A-O3′ coordination was not captured, the product structure (PDB ID 4KLT) clearly showed the occurrence of phosphodiester bond formation. However, unlike the case of the matched complex, the base of the incorrect dNTP broke its interaction with the templating base during the reaction (PDB ID 4KLQ). Moreover, the α-phosphate of the incorrect dNTP flipped away from the metal ions (PDB ID 4KLS), which was attributed to the strain on the DNA caused by the template shift.

3.3 | Sugar pucker conformations during catalysis

The ribofuranose sugar ring in nucleic acids is not flat but puckered with at least one of the five atoms out of plane.[87,88] In solution, dNTP and NTP molecules exist in a dynamic equilibrium between the C3′- and C2′-endo conformations.[89,90] The populations, energetics, and timescale of interconversion of these two conformations are highly dependent on the sugar ring substituents. Notably, in the active site of Pols, the sugar moieties of both the primer terminus and incoming (d)NTP typically adopt the C3′-endo conformation, as evinced by crystallographic data.[86,91] In nature, the most common helical topologies of DNA and RNA are the B-form (C2′-endo) and A-form (C3′-endo), respectively. The barrier for interconversion of these topologies in DNA duplex structures is small, as supported by several NMR and computational studies.[88,92–94]

Despite the preferred C2′-endo conformation of the B-form of DNA duplexes, several crystallographic structures of A-, B-, X-, and Y-family DNA Pols show the 3′-end of the primer adopting a C3′-endo conformation either in pre- or postreactant state.[62,83,86,95–100] The impact of the sugar pucker conformation on catalytic efficiency (k_{cat}/K_M) was tested for both Pol β[101] and Pol η[86] using a primer with a ribonucleotide (C3′-endo) at the 3′ end. The catalytic efficiencies were comparable whether the 3′-end nucleotide was a ribonucleotide or a deoxyribonucleotide. Additionally, the sugar pucker conformation was demonstrated to have an effect on both nucleotide incorporation and extension by Pols. For example, nucleotides constrained to a C3′-endo conformation (such as 2′-fluororibonucleotides) are preferentially incorporated by RNA Pols. Equally, 2′-fluoroarabino nucleotides, which prefer the C2′-endo conformation, are preferentially incorporated by DNA Pols. Interestingly, the consequent product of such incorporation is very difficult to extend further.[102] On these bases, it can be inferred that sugar puckering plays a nontrivial role for correct and efficient nucleotidyl transfer reaction catalyzed by Pols.

3.4 | Complete transcription cycle of an RNA polymerase

RNA-dependent RNA Pols allow the copying of RNA from RNA and are essential enzymes of RNA viruses, enabling the synthesis of viral proteins by host cells and their assembly with viral genome into new progeny.[103] In the influenza virus, the enzyme is composed of three intertwined polypeptide chains (known as PA, PB1, and PB2) of about 700 amino acids each.[104] The influenza RNA Pol is capable of both replication and transcription of viral RNA. Transcription takes place by a unique process, known as cap-snatching, in which different domains act in concert. The cap-binding domain (from PB2) binds a messenger RNA fragment of the host cell, which is cleaved by the endonuclease domain (from PA) and used as primer at the Pol site (from PB1). Replication is an unprimed process, which is guided by the structural reorganization of a priming loop that folds into the Pol core and aligns the first nucleotides of the nascent viral RNA. Recent studies have provided structural insight into both processes,[105] attesting the tremendous advances achieved by cryo-EM structure determination. In particular, the authors, through carefully designed experiments, captured the influenza RNA Pol in different "instants" of the transcription cycle,[106,107] demonstrating the complex structural reorganization that the enzyme undergoes to perform catalysis. Likely, this structural knowledge will greatly help the discovery of new and effective antiviral drugs against influenza.

4 | COMPUTATIONAL STUDIES OF POLYMERASE CATALYTIC MECHANISM AND PROPERTIES

Although structural studies have shed light on the catalytic cycle of Pols and, to a certain extent, have addressed the main factors discriminating between the correct and incorrect (d)NTPs, they employ methods that cannot capture transition states or fast dynamic events (e.g., substrate binding, primer 3′-OH deprotonation, and PPi release). Moreover, obtaining crystal structures of mismatched complexes with the catalytic metal (M_A) is challenging because of the weak binding of the incorrect (d)NTP.[108] In this regard, computational approaches are a valuable means to fill the gap in our knowledge of the molecular factors influencing the catalytic efficiency, fidelity, and processivity of Pols. In this section, we focus on the application of computational approaches in elucidating key aspects of the chemical reaction that have eluded experimental probes, namely, the identity of the general base that deprotonates the primer 3′-OH, the role of the third metal ion observed in DNA Pol crystal structures, and the timing and mechanism of Pol translocation. We then examine diverse mechanistic hypotheses on the origin of Pol fidelity that have arisen from comparative computational studies of correct and incorrect (d)NTP incorporation.

4.1 | Proposed mechanisms for 3′-OH deprotonation and nucleotide incorporation

4.1.1 | Classic mechanisms

Coordination to M_A is believed to lower the pK_a of the primer 3′-OH[83] (ribose 3′-OH: pK_a in water = 13,[109] effective pK_a in protein = 8.1[110]), and the conventional hypothesis is that either a conserved first-coordination-shell ligand or active site water molecule accepts the dissociated proton from 3′-OH (Figure 3). For the model DNA Pol, DNA Pol β, Wilson et al. proposed a conserved aspartate residue bound to M_A (D256) as the general base for the deprotonation step on the basis of ONIOM quantum mechanics/molecular mechanics (QM/MM) calculations.[110,111] Florian, Goodman, and Warshel used free energy perturbation/empirical valence bond (FEP/EVB) calculations to investigate alternative pathways for T7 DNA Pol, including proton transfer to bulk solvent and to one of the nonbridging oxygens of the α-phosphate of the incoming dNTP.[112] However, proton transfer to the catalytic residue, D654, remained the most energetically favorable pathway. Similarly, the corresponding catalytic residues of DNA Pol η (E116),[113] DNA Pol λ (D490),[114] and human immunodeficiency virus (HIV) RT (D185)[115] were shown to be the most plausible base by QM/MM studies. However, none of these studies addressed how the pK_a of aspartate or glutamate (3.90 and 4.35, respectively, in water[74]) is perturbed by the protein environment such that either residue is able to act as a proton acceptor.[113,116] On the other hand, a molecular dynamics (MD) study of DNA Pol I showed that a histidine residue (H829, pK_a in water = 6.46[74]), located about 10 Å from 3′-OH in the crystal structure of the open-conformation complex (PDB ID 4YFU), moves toward the catalytic site during enzyme closing. Then, it occasionally interacts with 3′-OH upon departure of the Na^+ ion at the A-site.[117] On the basis of this observation, H829 was proposed as a potential proton acceptor, although the energetic feasibility of this mechanism was not assessed.[117]

A few of the studies discussed above[112–114] demonstrated that proton transfer to bulk solvent via H_2O is a high-energy process. However, subsequent computational studies of DNA Pol β[118] and DNA Pol η[116,119] demonstrated the feasibility of such proton transfer to a deprotonated Mg_A-bound H_2O (Mg_A-bound OH^- p$K_a \approx 11.2$[116]). In the case of DNA Pol β, the resulting Mg_A-bound OH^- acts as a conduit for the eventual proton transfer to bulk solvent according to EVB and pK_a calculations (using the semi-macroscopic version of the protein dipole Langevin dipole in its linear response approximation [PDLD/S]).[118] On the other hand, for DNA Pol η, Stevens and Hammes-Schiffer[116] demonstrated the feasibility of proton transfer to the Mg_A-bound OH^- using QM/MM finite temperature string simulations. Subsequent DFTB3/MM metadynamics simulations by Roston, Demapan, and Cui[119] further indicated that this mode of proton transfer occurs concertedly with nucleophilic attack, with the overall process being the rate-limiting step of nucleotidyl transfer. However, in the case of RNA Pol II, Carvalho, Fernandes, and Ramos showed by ONIOM QM/MM calculations that this concerted mechanism is unfavorable and that proton transfer to a bulk OH^- (doubly H-bonded to the RNA terminus and near Mg_A) is the lowest-energy pathway.[120] Notably, it was further found that, unlike the case of DNA Pols, the most stable configuration of the RNA Pol II active site is the one where 3′-OH is only weakly coordinated to Mg_A. Moreover, 3′-OH needs to be dissociated from Mg_A to be able to attack the P_α atom of NTP. Here, it must be noted that, compared with the case in DNA Pol crystal structures (e.g., those discussed in Section 3), M_A is far from the RNA terminus and almost directly below the P_α atom of the incoming NTP in the RNA Pol II crystal structure (PDB ID 2E2H).[61] A later QM/MM study by Roßbach and Ochsenfeld, using the nudged elastic band approach and a different crystal structure (PDB ID 4A3F[121]) as the starting point, corroborated this mechanism for RNA Pol II.[56]

4.1.2 | Alternative mechanisms

As an alternative to these classic mechanisms, a water-mediated and substrate-assisted (WMSA) mechanism was proposed by Zhang et al. on the basis of QM/MM FEP studies of DNA Pol IV (Dpo4),[122] T7 DNA Pol,[123] and DNA Pol κ (Figure 3).[124] In the originally proposed WMSA mechanism for Dpo4, the α-phosphate of the incoming dNTP acts as the general base;[122] however, in contrast to the previous studies discussed in Section 4.1.1, the proton from 3′-OH is not directly transferred to it but rather through a water molecule. This indirect process was found to be the rate-limiting step and a lower-energy pathway compared with direct proton transfer to α-phosphate. The proton is then relayed to the γ-phosphate of dNTP via a solvent water and finally, directly to the β-phosphate of dNTP during nucleotidyl transfer. Thus, the WMSA mechanism accounts for not only the deprotonation of 3′-OH, but also the protonation of PPi, which is believed to be required for its release.[80] A variant of this mechanism was proposed for T7 DNA Pol[123] and DNA Pol κ,[124] wherein the initial base is γ-phosphate and not α-phosphate. For T7 DNA Pol, using the same QM/MM

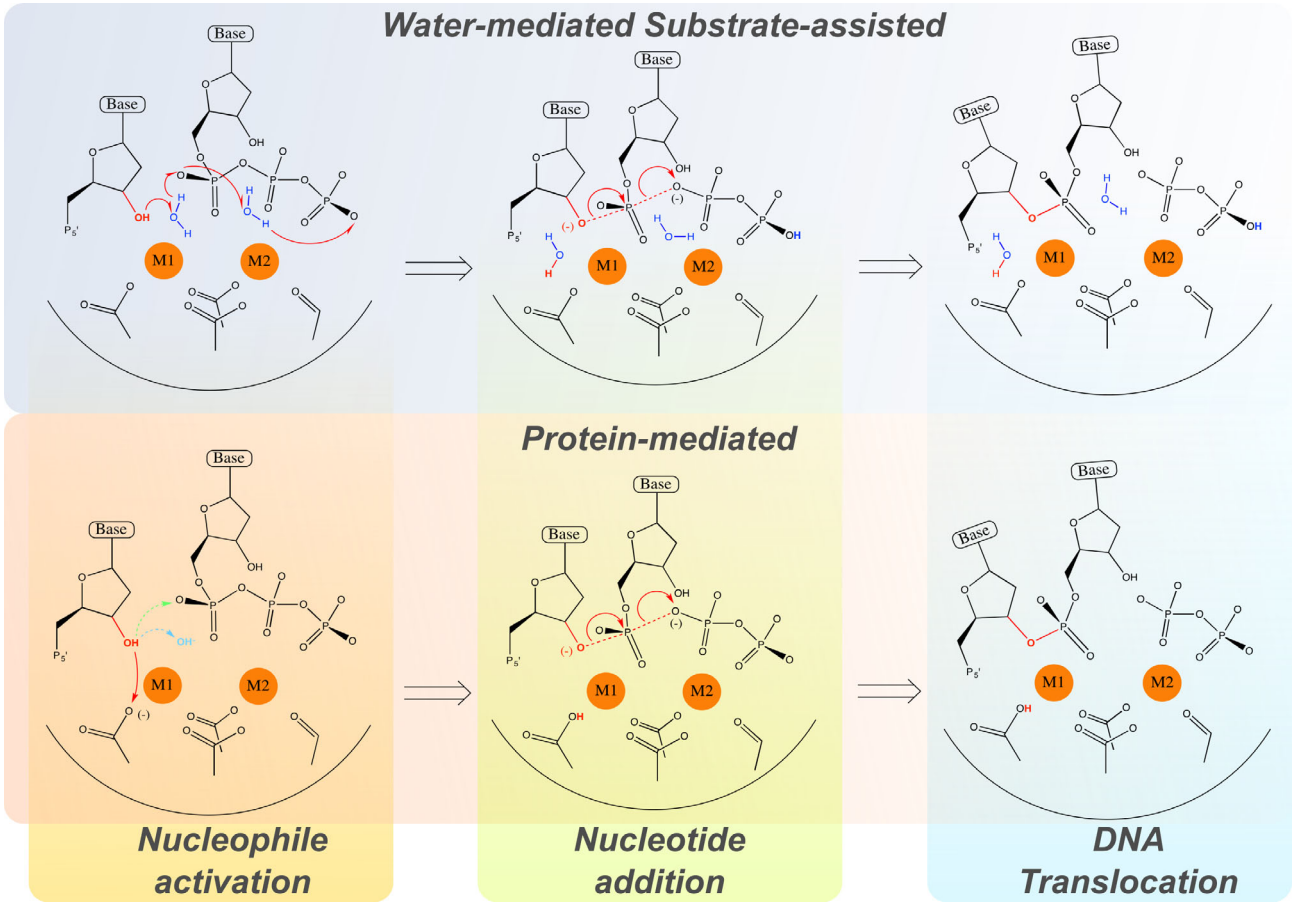

FIGURE 3 Proposed mechanisms for nucleotidyl transfer catalyzed by polymerases. In the protein-mediated mechanism, the 3′-OH group of the primer terminus is deprotonated by a conserved catalytic residue (Asp or Glu). In the water-mediated and substrate-assisted mechanism, the 3′-OH proton is initially transferred to the α-phosphate of the incoming nucleotide via a water molecule and ultimately relayed to the β-phosphate prior to pyrophosphate release and translocation. Reprinted with permission from Reference 126 Copyright 2018 American Chemical Society. https://doi.org/10.1021/acscatal.8b03363. Further permissions related to the material excerpted should be directed to the American Chemical Society

method/theory level, it was shown that the WMSA mechanism has a much lower barrier[123] than direct proton transfer to D654, which was proposed earlier by Florian, Goodman, and Warshel.[112] Indirect proton transfer to the α-phosphate via water and subsequent migration to the $O_{\alpha\beta}$ atom of NTP was also shown to be a viable pathway for RNA Pol II by Salahub et al. using QM/MM calculations.[125] However, indirect proton transfer via D483 and direct proton transfer to the α-phosphate were found to be equally favorable pathways. This result differs from the earlier one by Carvalho, Fernandes, and Ramos,[120] although both studies suggested that the RNA 3′-OH is weakly coordinated to or unbound from Mg_A during nucleotidyl transfer.

Alberts, Wang, and Schlick[127] proposed a similar water-mediated mechanism for DNA Pol β that, however, challenges those previously proposed by Wilson et al. (proton transfer to D256[110,111]) and Matute, Yoon, and Warshel (proton transfer to bulk solvent via Mg_A-bound OH^-).[118] The QM/MM calculations showed that the proton from 3′-OH is initially transferred to a water molecule and, during nucleotidyl transfer, migrates to D190 and finally to the γ-phosphate of dNTP as in the WMSA mechanism. In comparison, direct proton transfer to D256 was found to have a much higher barrier, and the proton returns to O3′ when the resulting intermediate was subjected to unconstrained minimization.

Finally, our group proposed a unique self-activated mechanism (SAM),[91] which significantly differs from the ones discussed above (Figure 4). The mechanism is "self-activated" because it is initiated by the intramolecular H-bond between the 3′-OH and β-phosphate groups of the newly incorporated dNTP, which facilitates the in situ deprotonation of 3′-OH. Importantly, the presence of this intramolecular H-bond upon the formation of the Michaelis complex was demonstrated to be a conserved feature in all available crystal structures of (deoxy)ribonucleotides complexed with

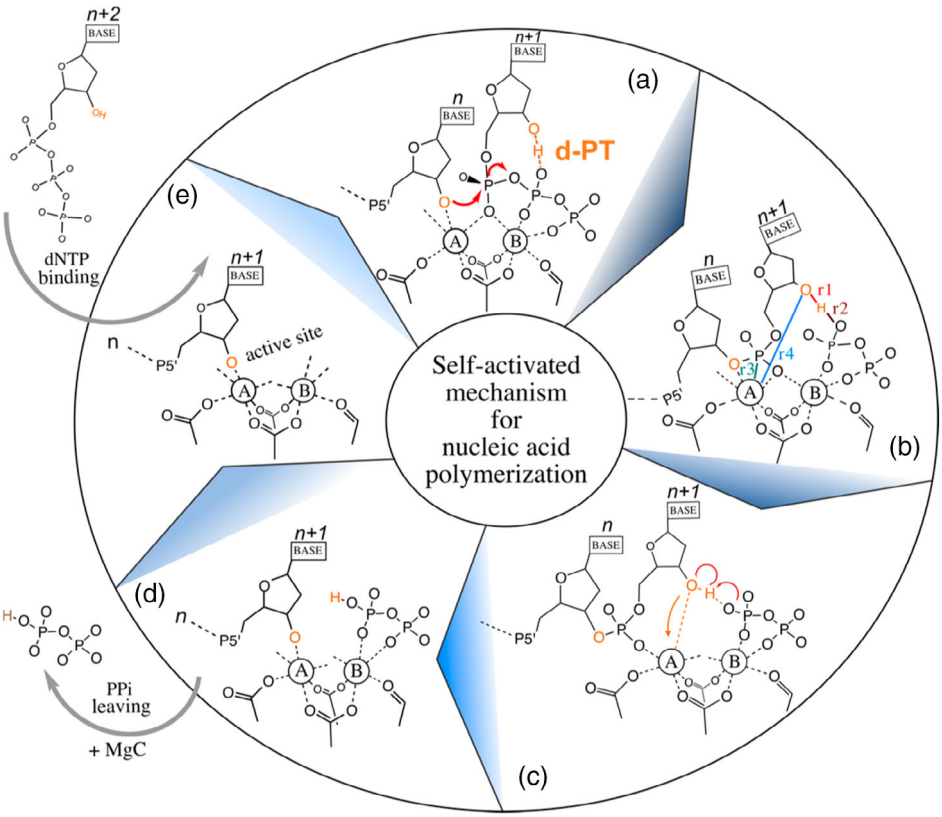

FIGURE 4 Proposed self-activated mechanism for nucleotidyl transfer catalyzed by polymerases. (a) The deprotonated 3′-OH group of the primer terminus attacks the P_α atom of the incoming nucleotide. (b) The primer strand has been extended by one nucleotide. The collective variables (r1–r2) and (r3–r4) were used to investigate (c) the subsequent intramolecular proton transfer and translocation by QM/MM metadynamics simulation. (d) The 3′-OH group of the newly incorporated nucleotide has been deprotonated, translocation has occurred, and pyrophosphate has been released. (e) The enzyme is ready for the binding of the next nucleotide. Reprinted with permission from Reference 91 Copyright 2016 American Chemical Society. https://doi.org/10.1021/jacs.6b05475. Further permissions related to the material excerpted should be directed to the American Chemical Society

Pol/D(R)NA binary complexes in the Protein Data Bank (Figure 5).[91] This unprecedented structural observation was indeed the basis of the proposed SAM. Car-Parrinello QM/MM simulations of DNA Pol η showed that direct proton transfer from 3′-OH to the pro-*S* oxygen of the β-phosphate occurs simultaneously with both the formation of the leaving PPi group and partial translocation (Figure 4). As a result, the newly incorporated dNTP is already deprotonated and positioned above M_A prior to the binding of the next dNTP, rendering the enzyme ready for a new catalytic cycle of dNTP incorporation. Intriguingly, this mechanism has been recently compared, via DFTB3/MM metadynamics simulations, to the other more conservative ones.[119] Roston, Demapan, and Cui argued that SAM is disfavored because their calculated pK_a for 3′-OH (lower limit of 9.0 for the posttranslocated state) implies that 3′-OH is not likely to remain deprotonated long enough during catalysis. This argument seems, however, marginal since SAM implies a level of synchronicity of a concerted sequence of chemical and physical steps that cannot be easily tested, especially with semiquantitative semiempirical methods. As a matter of fact, SAM remains the only conceptually novel alternative to the classical mechanisms. In fact, as eventually recognized also by Roston, Demapan, and Cui,[119] we emphasize that, unlike the previously proposed mechanisms for nucleotide incorporation (Figure 3), SAM is the only mechanism that enables a synergistic interconnection of the chemical (3′-OH deprotonation and nucleotidyl transfer) and physical (translocation) steps for a closed-loop catalytic cycle (Figure 4).[91] Notably, in this case, ab initio QM/MM simulations are essential in capturing the concerted structural rearrangement for nucleophile activation and leaving group formation in SAM during the partial nucleic acid translocation for Pol catalysis. Thus, the coupling between the chemical and physical steps form a closed-loop cycle for Pol catalysis. This is the reason why SAM remains a unique mechanistic hypothesis for the way the chemical and physical steps are linked together, returning the most favored energy path for nucleotide addition in Pols, so far.

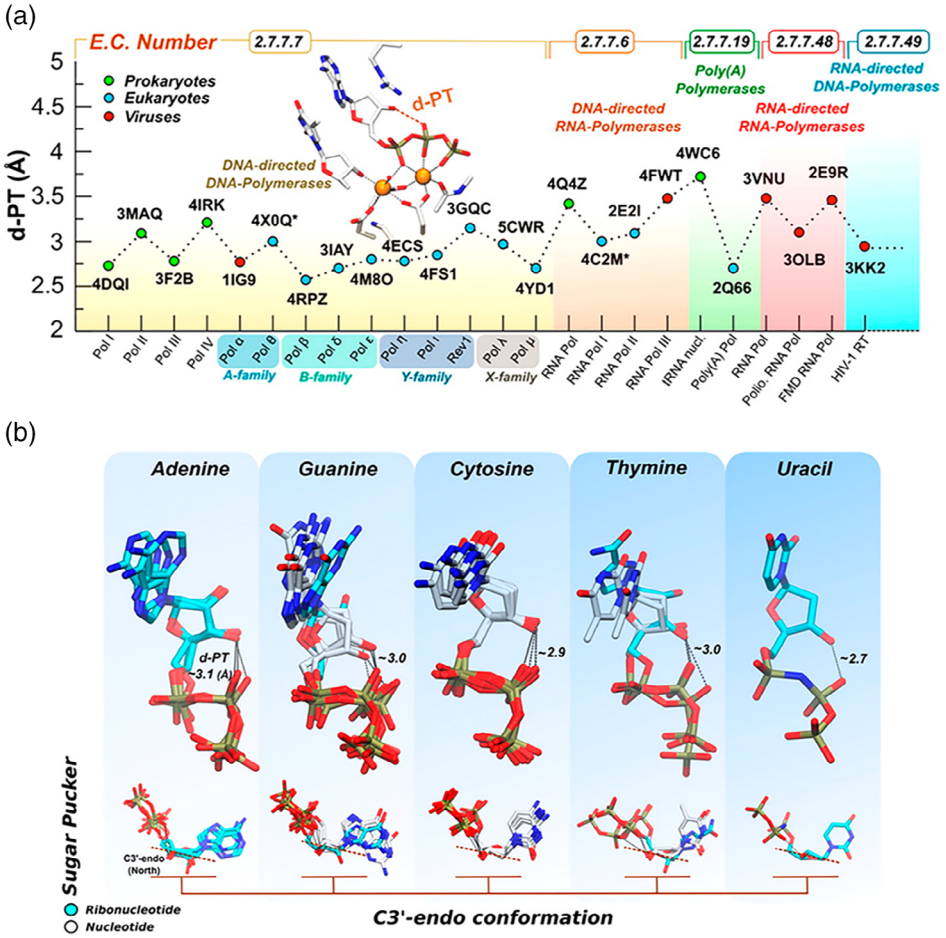

FIGURE 5 (a) Intramolecular H-bond distance between the 3′-OH and α-phosphate groups of the incoming nucleotide (d-PT, Å) in crystal structures of ternary Pol/D(R)NA/nucleotide complexes. (b) Superimposed crystal structures of ribonucleotides (cyan) and deoxyribonucleotides (white). The intramolecular H-bond and C3′-endo sugar conformation are preserved in all structures. Reprinted with permission from Reference 91 Copyright 2016 American Chemical Society. https://doi.org/10.1021/jacs.6b05475. Further permissions related to the material excerpted should be directed to the American Chemical Society

We close this section with a final consideration. As outlined above, different pathways have been proposed based on computational results for the activation of the 3′-OH nucleophile, even for the same Pol enzyme. Unfortunately, at present, it is not possible to discern experimentally which mechanism is operative in Pols at physiological conditions. Actually, some of these mechanisms may even coexist. Furthermore, we note that specific proton transfer pathways may contribute only marginally to the overall catalytic effect as long as the nucleophile is formed.[128] That is, while it is intellectually interesting to discriminate one deprotonation pathway from another, such pathways remain speculative if not corroborated by experimental data. A specific pathway for nucleophile activation (i.e., deprotonation) is actually significant only if proposed as the rate-limiting step, or if it involves a specific and conserved residue (i.e., irreplaceable residue for catalysis) as the proton acceptor.

4.2 | Role of the third metal ion in DNA polymerases

Time-resolved crystallography studies of DNA Pol η[86,129] and DNA Pol μ[130] showed that a third metal ion (Mg^{2+} or Mn^{2+}) appears near the α-phosphate during phosphodiester bond formation, suggesting that it plays a role in transition-state stabilization. For instance, in another group of nucleotide-processing enzymes, homing endonucleases, the third metal ion was found to be essential to achieving the proper geometry for phosphodiester bond hydrolysis.[131] However, subsequent computational studies of DNA Pol η gave diverging results on the effect of this third metal ion on the chemical barrier. This is likely because these studies used different mechanisms for the chemical

reaction.[116,119,132–135] Stevens and Hammes-Schiffer employed SAM, our proposed mechanism for 3′-OH deprotonation,[91] in modeling the chemical reaction in the two-Mg^{2+} system.[116] QM/MM finite temperature string simulations showed that nucleotidyl transfer has a higher barrier than 3′-OH deprotonation via SAM. On the other hand, for the three-Mg^{2+} system, only nucleotidyl transfer could be modeled because the third Mg^{2+} prevents the proton transfer from 3′-OH through electrostatic effects, which agrees with our earlier finding.[91] Nucleotidyl transfer was found to be a thermodynamically downhill process without a significant barrier[116] because the third Mg^{2+} stabilizes the PPi product through electrostatic interactions.[91,116,134]

Roston, Demapan, and Cui, on the other hand, modeled 3′-OH deprotonation by Mg_A-bound OH^- and nucleotidyl transfer as a concerted process.[119] QM/MM metadynamics simulations showed that the third Mg^{2+} lowers the barrier for this concerted process, lowers the pK_a of the Mg_A-coordinated H_2O, and stabilizes the Mg_A-bound OH^- and negative charge accumulated on the leaving group at the transition state. Interestingly, when a catalytic aspartate (D115) was used as the general base in the proton transfer step, the inclusion of the third Mg^{2+} in the model had no significant effect on the calculated barrier. In contrast to these two studies, Yoon and Warshel modeled the chemical reaction as a stepwise process of proton transfer, whose mechanism was unspecified, and nucleophilic attack.[132] FEP/EVB calculations showed that the third Mg^{2+} has no effect on the barrier for nucleophilic attack, although the overall reaction is more exothermic than the same reaction in the two-Mg^{2+} system. Their study is consistent with that of Wilson et al.[133] on DNA Pol β, wherein ONIOM QM/MM calculations along the P_α–$O_{\alpha\beta}$ bond reaction coordinate showed that the barriers for nucleotidyl transfer in the two- and three-Mg^{2+} systems are the same.

Unlike the case of DNA Pol η and DNA Pol μ, the third metal ion only appeared at the product state in the time-resolved crystallography study of DNA Pol β.[84] Wilson et al. thus postulated that it is likely involved in the reverse reaction (pyrophosphorolysis) and subsequently showed through ONIOM QM/MM calculations that the third Mg^{2+}, which is coordinated to the O_α and $O_{\alpha\beta}$ atoms, prevents the reformation of the P_α–$O_{\alpha\beta}$ bond.[135] On the other hand, our metadynamics study of DNA Pol η indicated that the third Mg^{2+}, in cooperation with R61, acts as an exit shuttle for PPi.[134] The mechanism involves two steps: (1) partial unbinding of the Mg_B–Mg_C–PPi complex and sidechain conformational change of R61 to interact (along with Y52 and R55) with this leaving group and (2) return of R61 to its original conformation to facilitate the release of the Mg_B–Mg_C–PPi complex. This proposed role of the third Mg^{2+} in PPi release was supported by subsequent FEP/EVB calculations of the overall reaction in DNA Pol η by Yoon and Warshel.[132] Notably, we observed a similar role for such a third transient ion in product release in another nucleotide-processing metalloenzyme, human exonuclease 1.[136] The importance of transient positive ions in nucleic acid synthesis and scission has been demonstrated for two-metal-ion enzymes and ribozymes (Figure 6).[137,138]

4.3 | Translocation mechanism of polymerases

Translocation is an important step in the catalytic cycle because it frees the active site for the addition of a new nucleotide. However, its mechanism and timing (i.e., before, after, or concurrent with active site reopening and PPi release) are not well defined by experimental studies. MD simulations of multi-subunit RNA Pols[139,140] and single-subunit T7 RNA Pol[141] suggested that translocation occurs after PPi release. Translocation in RNA Pol was demonstrated to follow the Brownian ratchet mechanism, wherein the system interconverts between the pre- and posttranslocated states until the binding of the next NTP inhibits backward translocation and locks the system in the posttranslocated state.[142,143] Yu et al.[144,145] and Huang et al.[146–148] provided a more detailed picture of the RNA Pol translocation mechanism through Markov state models, which enable the modeling of long-timescale dynamics (milliseconds for translocation) from many short MD simulations initiated from different parts of the free energy landscape. In T7 RNA Pol, the translocation of RNA and template DNA was initially unsynchronized because the latter was temporarily hindered by stacking interaction with F644 in the Y helix.[145] Opening of the O helix resulted in the insertion of Y639 into the active site, pushing the RNA–DNA hybrid upstream. In contrast to this system, RNA Pol II was simulated with the active site (i.e., trigger loop) already open prior to translocation[147] on the basis of earlier evidence from fluorescence[149] and MD simulation studies.[143] RNA and template DNA were observed to translocate simultaneously, facilitated by bending motions of the bridge helix (part of which is structurally analogous to the Y helix). The bridge helix was also shown to play a role in the backtracking process in RNA Pol II, which enables the nucleolytic cleavage of an RNA dinucleotide containing the wrong base.[148] Through its bending motions, it was observed to promote the fraying of the RNA terminus away from the template DNA if the base pairing is unstable, which is the case in a mispair.

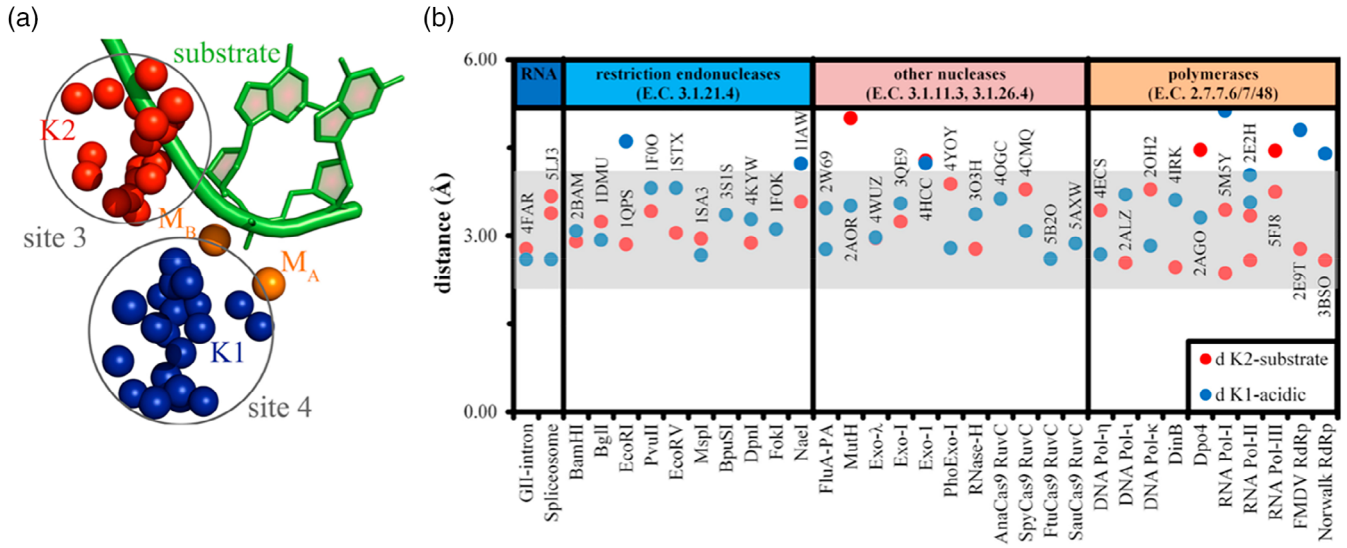

FIGURE 6 (a) Overlap of crystal structures of two-metal-ion nucleic-acid-processing enzymes. The blue and red spheres represent cations or basic amino residues in similar positions as the two catalytically important potassium ions (K1 and K2) in self-splicing group II intron ribozymes. (b) Distances (Å) between the K1-like elements and acidic residues coordinating M_A–M_B (d K1-acidic, blue dots) and between K2-like elements and the substrate (d K2-substrate, red dots). Reprinted with permission from Reference 137 Copyright 2017 Elsevier Ltd. https://doi.org/10.1016/j.str.2017.11.008. Further permissions related to the material excerpted should be directed to Elsevier Ltd

In comparison, there have been fewer computational studies of the specific steps in DNA Pol translocation. In the case of DNA Pol η, which does not show distinct conformational states during the catalytic cycle, we proposed the coupling of DNA Pol translocation with the deprotonation of the 3′-OH group of the newly incorporated dNTP (see Section 4.1.2 and Figure 4, SAM for nucleotide incorporation).[91] On the other hand, in a computational study of the closed conformation of HIV RT employing locally enhanced sampling, steered MD, and Milestoning, translocation was observed to begin as PPi exits the protein.[150] Similarly, a restricted-perturbation targeted MD study of DNA Pol I suggested that PPi release enables the coupled closed → open conformational transition and translocation.[151]

Modeling D(R)NA translocation in DNA Pols and RNA Pols is certainly highly challenging given the global nature (i.e., many atoms and structural motifs involved, all moving in concert) of the structural rearrangements required to advance the Pol along the template strand. In this regard, methodological advances aimed at identifying collective motions in complex molecular systems will contribute to unravel the critical features and molecular determinants of Pol translocation.

4.4 | Origin of polymerase fidelity

4.4.1 | Opposing views on the contribution of prechemistry steps

Because global (i.e., open → closed transition) and local conformational changes in Pols are a prerequisite for D(R)NA synthesis, the prechemistry steps of the nucleotide addition cycle have been hypothesized to contribute to catalysis (i.e., k_{cat} or k_{pol}) and consequently, to the fidelity of Pols, regardless of whether they are rate determining. This is in line with the induced-fit mechanism, wherein the binding of the correct substrate brings the enzyme to the proper orientation for the catalytic reaction.[152] This hypothesis has been investigated through comparative computational studies of the conformational transition mechanism in matched and mismatched complexes. For instance, transition path sampling calculations of DNA Pol β revealed that the active site residues around the Mg^{2+} ions do not rearrange to a catalytically competent configuration in the mismatched G:A complex during the closing of the N-subdomain, unlike in the matched G:C complex.[153–155] Furthermore, free energy calculations indicated that the closed conformation is only a metastable state for the mismatched complex. A more recent MD study on DNA Pol I suggested that the binding of the

incorrect dNTP (dTTP opposite G) disrupts the interaction networks between the 3′ → 5′ exonuclease domain and palm and thumb subdomains, causing the conformational equilibrium to shift toward the open conformation.[156] Because of the differences between the prechemistry steps of matched and mismatched complexes revealed by these computational studies, as well as experiments,[157,158] it has been proposed that prechemistry steps can act as kinetic "checkpoints" for the selection of the correct (d)NTP.[159]

Along the same lines, Johnson and his collaborators[160,161] proposed a three-step mechanism for (d)NTP incorporation

$$ED_n + N \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ED_nN \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} FD_nN \underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} FD_{n+1}PPi \overset{\text{fast}}{\rightleftharpoons} ED_{n+1} + PPi \tag{1}$$

where E and F are the open and closed forms of the Pol, respectively, D is D(R)NA, and N is the incoming nucleotide. k_{cat}/K_M can then be derived as

$$\frac{k_{cat}}{K_M} = \frac{k_1 k_2 k_3}{k_2 k_3 + k_{-1}(k_2 + k_3)} \tag{2}$$

Subsequently, their MD simulations of HIV RT using the Milestoning approach showed that the closed conformation is more thermodynamically favorable than the open conformation for the matched A:T complex, but not for the mismatched A:A complex.[162] Moreover, the reacting groups were properly aligned for the chemical reaction in the matched complex, but disordered in the mismatched complex. From these results, it was inferred that for correct dNTP incorporation, $k_3 \gg k_{-2}$ and $k_{cat}/K_M \approx K_1 k_2$ (where $K_1 = k_1/k_{-1}$), while for misincorporation, $k_3 \ll k_{-2}$ and $k_{cat}/K_M \approx K_1 K_2 k_3$ (where $K_2 = k_2/k_{-2}$). This implies that substrate affinity (K_1) and the rate of the open → closed transition (k_2) determine the specificity of HIV RT for the correct dNTP.

The concept of kinetic checkpoints has also been invoked to explain the fidelity of RNA Pols. Of the five kinetic checkpoints proposed by Feig et al. for multi-subunit RNA Pols, two occur prior to the chemical step: (1) rotation of the NTP as it moves from the entry site to the insertion site and (2) steric interaction between the closed trigger loop and NTP.[163] It was subsequently demonstrated that a kinetic model including these checkpoints could reproduce the experimental misincorporation rate.[164] On the other hand, for the viral single-subunit T7 RNA Pol, Yu et al. proposed three prechemistry kinetic checkpoints: (1) NTP dissociation from the preinsertion site, (2) NTP binding at the insertion site, and (3) NTP dissociation from the insertion site (i.e., the reverse of 2).[165,166] Furthermore, they calculated a "selection" free energy for each of these checkpoints by umbrella sampling simulations. The total selection free energy was then used to derive the theoretical error rate of T7 RNA Pol using the chemical master equation.[167,168] It was found that the total selection free energy from the prechemistry kinetic checkpoints is sufficient to reproduce the experimental error rate for dNTP incorporation but not that for incorrect NTP incorporation.[166] From this, it was inferred that the chemistry step must also contribute to the total selection free energy through a higher activation barrier in order to discriminate against the incorrect NTP. Additionally, these studies and other free energy studies of prechemistry complexes of RNA Pols[169,170] consistently demonstrated that dNTP or the incorrect NTP tend to dissociate from the preinsertion and/or insertion site, hinder the closing of the active site, and/or cause structural distortion.

Similarly to Yu et al., Salahub et al. devised a stochastic kinetic scheme for the entire nucleotide addition cycle to determine the origin of the sugar selectivity (i.e., NTP vs. dNTP) of RNA Pol II.[171] Unlike the studies above, the chemical step was found to have a more dominant contribution, that is, the experimental substrate selectivity of RNA Pol II could only be reproduced by assigning a different chemical reaction rate for each nucleotide. This conclusion was later corroborated by Roßbach and Ochsenfeld, who applied QM/MM calculations and the nudged elastic band method to show that ATP incorporation by RNA Pol II has a lower reaction barrier than dATP incorporation.[56] The lower reaction barrier was due to an active site Arg (R446) that bridges the RNA terminus with ATP (via H-bonds) but not with dATP.

On the other hand, Warshel et al. posited that unless they are rate limiting (i.e., have the highest free energy barrier), these so-called checkpoints represented by prechemistry steps do not contribute to fidelity.[172] In other words, unlike the case in the kinetic models proposed for RNA Pols,[165,166,171] fidelity is not considered as the cumulative effect of all kinetic checkpoints during the catalytic cycle. According to this hypothesis, if the chemistry step is the rate-limiting step (as experimentally shown in the case of, for example, DNA Pol β[173–175] and DNA Pol η[176]), only the chemical barrier, and not the barriers of the prechemistry steps, contribute to catalysis and fidelity. However, Warshel et al. clarified that prechemistry conformational changes do determine the final active site preorganization, as reflected in

the ground-state Michaelis complex.[172] In this way, prechemistry conformational changes by themselves, and not their associated barriers, have an indirect impact on the chemical barrier and catalytic rate. For example, for DNA Pol β, structural data showed that the fully folded mismatched G:dATP/DNA/Pol ternary complex has a closed conformation but the primer 3′-OH is neither coordinated to M_A nor properly aligned with the P_α atom of dATP for nucleophilic attack, unlike the matched G:dCTP/DNA/Pol complex.[84] Two independent computational studies, employing ONIOM QM/MM[177] and FEP/EVB,[178] indicated that bringing the mismatched complex from this ground state to a reactant state entails an additional energy cost. The resulting reactant state and subsequent transition state for the nucleophilic attack on the incorrect dNTP (dATP) were shown to lie at higher energies than the corresponding states in the matched complex. These studies somehow support the point of Warshel et al.[172] that active site preorganization, irrespective of the preceding conformational change, contributes to the fidelity of a Pol if the chemical reaction is the rate-determining step. Both the matched and mismatched complexes of DNA Pol β were shown to react in the closed conformation; however, the optimal active site preorganization of the former facilitated transition state stabilization, leading to a lower chemical barrier, whereas the poor active site preorganization of the latter diminished transition state stabilization, leading to a higher chemical barrier. This overall point on the mechanistic origin of fidelity in Pols remains however quite intriguing, with mechanistic aspects related to correct ligand selection and its subsequent efficient catalytic processing that will certainly benefit from additional careful computational studies.

4.4.2 | Contribution of (d)NTP binding affinity to polymerase fidelity

The relative binding energies of the correct and incorrect (d)NTPs ($\Delta\Delta G_{\text{bind}}$) is related to K_M or K_d[179]

$$\frac{(K_d)_{\text{incorrect}}}{(K_d)_{\text{correct}}} = \exp(\Delta\Delta G_{\text{bind}}/\text{RT}) \tag{3}$$

and therefore, also contributes to Pol fidelity (see Section 2.1). Binding free energy studies of dNTP/DNA/Pol ternary complexes have been mainly reported by Warshel et al., who employed the linear response approximation (LRA) method. As an approximation, only the potential energy due to the nucleobase moiety was monitored, and all possible combinations of the four bases (limited to the neutral anti-anti configuration) were considered.[180,181] The binding free energy contribution to fidelity was evaluated for DNA Pol β and T7 DNA Pol. The calculations showed that $\Delta\Delta G_{\text{bind}}$ is higher for T7 DNA Pol, consistent with the higher fidelity of this DNA Pol. The higher $\Delta\Delta G_{\text{bind}}$ for T7 DNA Pol was attributed to the larger displacement of the incorrect dNTP toward the major groove compared with the case in DNA Pol β. Warshel et al. subsequently modified their approach to include the contribution of the triphosphate group of the incoming dNTP and calculated the binding energies at the transition state ($\Delta G^{\text{TS}}_{\text{bind}}$) instead of the ground state.[182] A high correlation ($R = 0.97$) was obtained between the calculated $\Delta G^{\text{TS}}_{\text{bind}}$ and ΔG_{bind} derived from presteady-state kinetic data. They also compared the contributions of both the ΔG_{bind} and chemical barrier to the fidelities of T7 DNA Pol,[179] DNA Pol β,[178] and DNA Pol η[132] and showed that ΔG_{bind} has a slightly larger contribution in all cases. Ucisik and Hammes-Schiffer also reported binding free energy calculations using the more computationally expensive thermodynamic integration and reproduced the specificity of DNA Pol η for dATP versus dGTP opposite template T.[183] On the other hand, to rationalize the sugar selectivity of RB69 DNA Pol and T7 RNA Pol, Yoon and Warshel evaluated the change in $\Delta G^{\text{TS}}_{\text{bind}}$ upon mutating CTP to dCTP.[184] The FEP calculations suggested that the discrimination against CTP by RB69 DNA Pol is due to steric interaction (van der Waals contribution) between the 2′-OH group and an active site Tyr (Y416), while the discrimination against dCTP by T7 RNA Pol is due to electrostatic effects. The results of these computational studies therefore suggest that the binding affinity for the incoming (d)NTP also contributes to Pol fidelity.

4.4.3 | Role of conserved residues in correct (d)NTP selection

Residues that are potentially critical to the pre-chemistry conformational change, chemical reaction, and fidelity have been identified qualitatively by comparing their interactions during MD simulations of matched/mismatched complexes[185–187] or wild-type/mutant Pols.[185,188,189] For a more quantitative approach, energy decomposition analysis has been employed.[111,114,190,191] For example, Pedersen et al. used ONIOM QM/MM calculations to determine residue

interactions in the active sites of DNA Pol β[111] and DNA Pol λ[114] that stabilize the transition state and lower the chemical barrier for correct dNTP incorporation. Graham, Syeda, and Cisneros performed energy decomposition analysis, along with electrostatic free energy response and noncovalent interaction analyses, using only energies obtained from MD simulations.[190] Interactions were calculated at a purported fidelity-checking site (two bases downstream from the active site) of DNA Pol I to which the enzyme transiently translocates after dNTP incorporation. These calculations allowed the identification of residues that are potentially involved in fidelity checking at the postinsertion site. This would be based on the change in the interaction energy of such residues with a mispaired base in the DNA substrate. On the other hand, Warshel et al. employed the PDLD/S-LRA method to reproduce the effect of mutations on the experimental activation energy of DNA Pol β and compare the contributions of ionized residues near the active site to the ΔG_{bind}^{TS} of matched and mismatched complexes.[192] However, to assess the contribution of residues far from the active site, they used a different approach, the hybrid FEP/linear interaction energy (LIE) approximation method.[193] The results were then rationalized by comparing residue interactions in the binary and transition state complexes of wild-type and mutant Pol β.

Unsurprisingly, most of the important residues identified by these methods are conserved residues. Importantly, using bioinformatics analysis of a large set of DNA/Pol structures, we found a conserved positively charged residue (Arg or Lys) located in the fingers subdomain of DNA Pols that always interacts with the incoming dNTP (Figure 7).[194] MD and metadynamics simulations of DNA Pol η indicated that π-stacking and H-bond interactions of the conserved residue, R61, with the base and triphosphate groups of the incoming dNTP, respectively, facilitate canonical WC base pairing, which is lower in energy than the alternative Hoogsteen base pairing. However, for the R61A mutant, the Hoogsteen base pairing was found to be slightly lower in energy, which is consistent with the observed Hoogsteen base pairing in dNTP/DNA/Pol crystal structures where the conserved Lys or Arg is missing, mutated, or displaced. We also identified through bioinformatics analysis two conserved Lys or Arg residues in the second coordination shell of the two-metal-ion center of Pols (Figure 6).[137] MD simulations of DNA Pol η indicated a possible role in active site stabilization on the basis of the observed partial unfolding of the DNA substrate and consequent distortion of the Michaelis–Menten complex upon mutation of these residues. Mutagenesis[195] and computational[196] studies on another nucleotide-processing metalloenzyme, all-α dimeric deoxyuridine triphosphate nucleotidohydrolase, suggested that such residues contribute to specificity by stabilizing the correct substrate.

5 | COMPUTATIONAL STUDIES OF POLYMERASE APPLICATIONS

Computational methods can be utilized not only to elucidate the Pol catalytic mechanism, but also to discover inhibitors against Pols and guide the protein engineering of Pols for various applications. This will reduce the need for more costly traditional experimental approaches such as inhibitor screening assays and directed evolution. Here, we discuss a few examples of common and emerging applications of computational methods in these research areas.

5.1 | Therapeutics: Polymerases as targets for disease treatment and prevention

One of the well-known Pol drug targets is HIV RT, which catalyzes the reverse transcription of the viral single-stranded (+) RNA genome of the HIV into double-stranded DNA.[16] Computational approaches, including docking, de novo design, and FEP, have been successfully applied in the discovery of nonnucleoside inhibitors with picomolar and low-nanomolar activities against wild-type and mutant HIV RT.[197] Currently, DNA Pols involved in translesion synthesis (e.g., Rev1, DNA Pol η, DNA Pol ι, DNA Pol κ, and DNA Pol ξ) have gained attention as drug targets because of their involvement in the onset and development of cancer.[17,18] Additionally, these DNA Pols can decrease the efficacy of chemotherapy by bypassing the nucleotide adducts formed by anticancer agents. One of the reported drug candidates is an indole thiobarbituric acid derivative that inhibits human DNA Pol η activity.[198] Kinetic experiments and chemical footprinting assays, with the aid of docking calculations, indicated that the inhibitor likely binds to a pocket between the fingers and little finger subdomains of the dNTP/DNA/Pol ternary complex, thereby preventing the proper positioning of the template DNA strand. Another timely target is the RNA-dependent RNA Pol of SARS-CoV-2, which causes COVID-19.[24] Because of the urgency to develop therapeutics, one of the approaches that has been adopted is the drug repurposing of known RNA-dependent RNA Pol inhibitors such as remdesivir. Several MD studies have already been performed to understand the mechanism of inhibition of remdesivir, which will hopefully aid the discovery of new and more active inhibitors.[199–202]

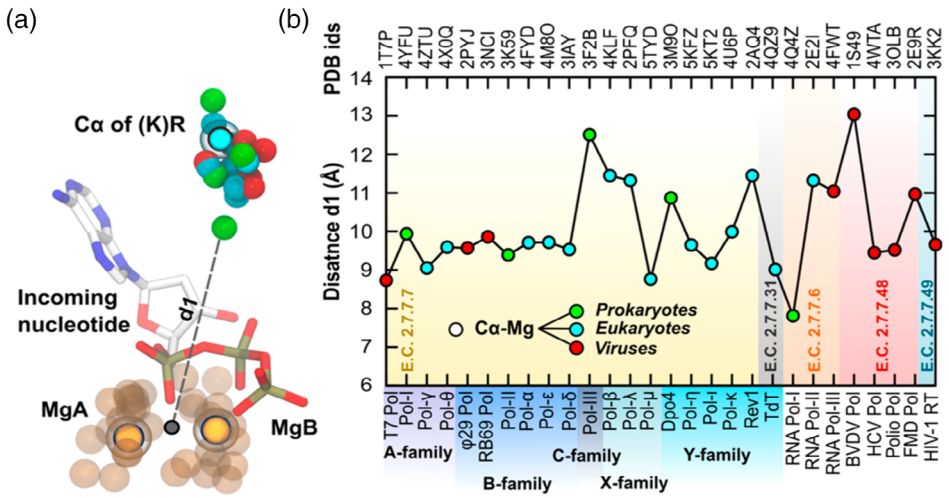

FIGURE 7 Distance between the α-carbon of a positively charged residue (Arg or Lys) and the two catalytic metals (d1, Å) in crystal structures of ternary Pol/D(R)NA/nucleotide complexes. The positively charged residue is believed to facilitate Watson–Crick base pairing between the incoming nucleotide and templating base. Reprinted with permission from Reference 194 Copyright 2018 American Chemical Society. https://doi.org/10.1021/jacs.7b12446. Further permissions related to the material excerpted should be directed to the American Chemical Society

5.2 | Biotechnology: DNA Polymerases for PCR amplification of damaged DNA

A major drawback of DNA Pols used in PCR amplification, such as Taq DNA Pol I, is their inability to amplify damaged DNA samples for medical and forensic analysis.[203] Thus, an attractive solution to this problem is to use thermostable Y-family DNA Pols, which specialize in lesion bypass. However, the poor processivity and activity of these DNA Pols hinder their application. Nevertheless, several studies have demonstrated that these limitations can be circumvented by strengthening DNA binding through mutations and/or fusion with a nonspecific DNA-binding protein such as *Sulfolobus solfataricus* nonspecific DNA-binding protein 7d (Sso7d).[204–208] In this respect, computational approaches can be used to rationally design DNA Pol variants with improved processivity and catalytic activity. This has been demonstrated for the Sso7d-Dbh (Din B homologue from *S. acidocaldarius*) fusion protein.[204] Potential mutation sites (nonconserved residues) around the DNA substrate in Dbh were identified by bioinformatics analysis, and the ΔG_{bind} of DNA to the selected mutants were calculated using the molecular mechanics Poisson–Boltzmann surface area (MM-PBSA) method. The designed variants were experimentally confirmed to have lower K_d (i.e., higher affinity), higher processivity, and higher activity for dCTP incorporation opposite both G and 8-oxoG than wild-type Sso7-Dbh.[204] The same computational strategy was also applied to improve the processivity and catalytic efficiency (dCTP incorporation opposite 8-oxoG) of Dpo4 from *S. solfataricus*.[205]

5.3 | Synthetic biology: DNA Polymerases for incorporation of unnatural base pairs

The expansion of the genetic alphabet, through the creation of UBPs from modified nucleotides, enables the storage of additional information in DNA.[209,210] Furthermore, the increased chemical and structural diversity of the modified DNA and RNA would broaden their molecular biological and biotechnological applications to, for example, production of aptamers that bind to proteins and cells, incorporation of unnatural amino acids into proteins, and generation of semisynthetic organisms. However, the major challenge is to identify or develop DNA Pols that can incorporate UBPs with high catalytic efficiency and fidelity and maintain high extension efficiency after the incorporation of the UBP. In this respect, computational approaches can complement structural studies in understanding the mechanisms of UBP insertion and postinsertion elongation by DNA Pols. For instance, MD simulations have been used to investigate protein dynamics in a variant of the Klenow fragment of Taq DNA Pol I (KlenTaq), which incorporates the "hachimoji" **P**: **Z** base pair (**P**: 2-amino-imidazo[1,2-*a*]-1,3,5-triazin-4(8*H*)one; **Z**: 6-amino-5-nitro-2(1*H*)-pyridone) with high catalytic efficiency.[210] The calculations demonstrated that the mutations (located far from the active site) increase the enzyme's

flexibility, allowing it to interact with the **P**:**Z**-containing DNA as optimally as with a purely WC DNA. MD simulations have also aided in understanding the incorporation of P-alkyl phosphonate nucleic acids, which have a noncanonical uncharged backbone, by B-family DNA Pols. Particularly, the selective incorporation of the $(S)_p$-diastereoisomer of P-alkyl phosphonate nucleic acids over the $(R)_p$-diastereoisomer was shown to be due to the formation of a stable Michaelis–Menten complex with a catalytically ready configuration.[211] On the other hand, QM/MM energy minimization has been used to understand how the protein environment of wild-type KlenTaq modulates the conformation of a 1,4-diethynylbenzene-modified nucleotide once it is positioned at the 3′-primer terminus.[212] Density functional theory calculations also indicated a low energy barrier for the rotation around the diethynyl axis of the modified nucleotide, which explained its twisted conformation when it is at positions 2, 4, and 6 upstream from the 3′-primer terminus.[212]

6 | CONCLUSION

Computational studies have built upon the wealth of existing structural and kinetic data on Pols to significantly advance our understanding of the Pol catalytic mechanism. Importantly, computational approaches have provided a means to study several key aspects of Pol function. We have reviewed numerous representative computational investigations on Pol function, including mechanistic events that are difficult to address at the molecular level with experimental means. Such events range from physical steps, such as conformational change and translocation, to chemical steps, such as proton transfer and reaction with transient ions. Another relevant aspect is misincorporation in Pols, which is challenging to examine in atomic detail through experiments owing to the inherent instability of mismatched complexes. In this case, molecular simulations can help in the direct comparison of the catalytic cycles of matched and mismatched complexes and the formulation of mechanistic hypotheses regarding the origin of Pol fidelity. Yet, there remains a lack of consensus on the primer 3′-OH deprotonation mechanism and role of the third metal ion in DNA Pols. Thus, further computational and experimental studies are required to clarify these aspects of the catalytic cycle, particularly on whether they are generally applicable to all Pols or specific to only the Pol under investigation. To resolve these issues in future computational work, we consider it critical to address: (1) extensive configurational sampling, (2) consistency in theoretical method and active site models employed (e.g., either with or without the third metal ion in DNA Pols) when evaluating all possible mechanisms, and (3) consideration of the pK_a changes of the residues involved due to the microenvironment.[74,113,132]

In addition to the investigation of the fundamental mechanism for Pol function, we also touched upon computational investigations in drug discovery targeting Pols; docking to predict binding poses, FEP to calculate binding affinity, and QM/MM and MD simulations[213,214] to study the mechanism of inhibition and perform dynamic docking can indeed find good use also in the context of Pol drug discovery efforts. The use of newer methods such as τ-random acceleration MD[215] and smoothed potential MD[216] to study inhibitor binding kinetics (e.g., drug residence time) has gained traction in recent years. Machine learning, which is commonly used for de novo molecular design and quantitative structure–activity relationship modeling, can also be utilized for other stages of drug discovery and development such as biomarker discovery.[217] In the case of Pols, we note that inhibitors can be designed to bind to either the active site and so terminate nucleic acid synthesis or bind to an allosteric site and hinder the formation of a reactive complex (e.g., prevent enzyme closing).[16,17,23] In this context, drug discovery targeting Pols faces two main challenges, namely, inhibitor selectivity due to the structural similarities among Pols and drug resistance due to mutations in Pols, specifically and often in the case of viral RNA Pols. Optimistically, we believe these drug discovery challenges will be overcome also with the aid of the variety of computational approaches discussed above and, in this regard, we look forward to future efforts and findings achieved via computational modeling integrated with experiments.

We have also highlighted promising applications of computations in the protein engineering of Pols for biotechnology and synthetic biology. Computational methods that can be used to predict the effect of mutation on catalytic efficiency, fidelity, and processivity include free energy calculations[218,219] and machine learning.[220] The knowledge of the catalytic mechanism and determinants of the nucleotide specificity of Pols gained through computational studies can aid in the rational design of Pols. In this regard, one property that has not yet been sufficiently explored by computational studies is Pol processivity. This is particularly challenging and important because PCR and synthetic applications require highly processive DNA Pol variants that can efficiently extend DNA after the incorporation of modified or synthetic nucleotides.

In conclusion, computational studies not only provide testable hypotheses and models of the Pol catalytic cycle and chemical properties that can be subsequently validated by experimental methods such as mutagenesis, but also generate predictions that can greatly expedite the discovery of Pol-targeting drugs and protein engineering of Pols. Computational studies are certainly expected to continue playing a significant role in Pol research and we are eager to witness the computer-aided advances of this fascinating research field.

ACKNOWLEDGMENTS

Marco De Vivo thanks the Italian Association for Cancer Research (AIRC) for financial support (IG 23679).

AUTHOR CONTRIBUTIONS

Inacrist Geronimo: Conceptualization; writing-original draft. **Pietro Vidossich:** Conceptualization; writing-original draft. **Elisa Donati:** Writing-original draft. **Marco De Vivo:** Conceptualization; writing-original draft.

CONFLICT OF INTEREST

The authors have declared no conflicts of interest for this article.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Marco De Vivo https://orcid.org/0000-0003-4022-5661

RELATED WIREs ARTICLES

Recent advances in dynamic docking for drug discovery

REFERENCES

1. Crick F. Central dogma of molecular biology. *Nature*. 1970;*227*:561–3.
2. Shapiro JA. Revisiting the central dogma in the 21st century. *Ann N Y Acad Sci*. 2009;*1178*:6–28.
3. The Nobel Prize in Physiology or Medicine 1959. https://www.nobelprize.org/prizes/medicine/1959/summary/. Accessed 27 Nov 2020.
4. The Nobel Prize in Chemistry 2006. https://www.nobelprize.org/prizes/chemistry/2006/summary/.Accessed 27 Nov 2020.
5. The Nobel Prize in Chemistry 2015. https://www.nobelprize.org/prizes/chemistry/2015/summary/.Accessed 27 Nov 2020.
6. Loeb LA, Monnat RJ. DNA polymerases and human disease. *Nat Rev Genet*. 2008;*9*:594–604.
7. Lange SS, Takata K, Wood RD. DNA polymerases and cancer. *Nat Rev Cancer*. 2011;*11*:96–110.
8. Masutani C, Kusumoto R, Yamada A, Dohmae N, Yokoi M, Yuasa M, et al. The XPV (xeroderma pigmentosum variant) gene encodes human DNA polymerase η. *Nature*. 1999;*399*:700–4.
9. Lang T, Maitra M, Starcevic D, Li S-X, Sweasy JB. A DNA polymerase β mutant from colon cancer cells induces mutations. *Proc Natl Acad Sci U S A*. 2004;*101*:6074–9.
10. Nurminen A, Farnum GA, Kaguni LS. Pathogenicity in POLG syndromes: DNA polymerase gamma pathogenicity prediction server and database. *BBA Clin*. 2017;*7*:147–56.
11. Rahman S, Copeland WC. POLG-related disorders and their neurological manifestations. *Nat Rev Neurol*. 2019;*15*:40–52.
12. O-Wang J, Kawamura K, Tada Y, Ohmori H, Kimura H, Sakiyama S, et al. DNA polymerase κ, implicated in spontaneous and DNA damage-induced mutagenesis, is overexpressed in lung cancer. *Cancer Res*. 2001;*61*:5366–9.
13. Pillaire M-J, Selves J, Gordien K, Gouraud P-A, Gentil C, Danjoux M, et al. A 'DNA replication' signature of progression and negative outcome in colorectal cancer. *Oncogene*. 2010;*29*:876–87.
14. Higgins GS, Harris AL, Prevo R, Helleday T, McKenna WG, Buffa FM. Overexpression of POLQ confers a poor prognosis in early breast cancer patients. *Oncotarget*. 2010;*1*:175–84.
15. Kawamura K, Bahar R, Seimiya M, Chiyo M, Wada A, Okada S, et al. DNA polymerase θ is preferentially expressed in lymphoid tissues and upregulated in human cancers. *Int J Cancer*. 2004;*109*:9–16.
16. Berdis AJ. DNA polymerases as therapeutic targets. *Biochemistry*. 2008;*47*:8253–60.
17. Korzhnev DM, Hadden MK. Targeting the translesion synthesis pathway for the development of anti-cancer chemotherapeutics. *J Med Chem*. 2016;*59*:9321–36.
18. Zafar MK, Eoff RL. Translesion DNA synthesis in cancer: molecular mechanisms and therapeutic opportunities. *Chem Res Toxicol*. 2017;*30*:1942–55.
19. Nirwan S, Kakkar R. Chapter 11: rhinovirus RNA polymerase—structure, function, and inhibitors. In: Gupta SP, editor. *Viral polymerases: structures, functions and roles as antiviral drug targets*. London: Academic Press; 2019. p. 301–31.

20. Dash S, Aydin Y, Stephens CM. Chapter 8: hepatitis C virus NS5B RNA-dependent RNA polymerase inhibitor—an integral part of HCV antiviral therapy. In: Gupta SP, editor. *Viral polymerases: structures, functions and roles as antiviral drug targets*. London: Academic Press; 2019. p. 211–35.
21. Dyall J, Gross R, Kindrachuk J, Johnson RF, Olinger GG, Hensley LE, et al. Middle east respiratory syndrome and severe acute respiratory syndrome: current therapeutic options and potential targets for novel therapies. *Drugs*. 2017;*77*:1935–66.
22. Castro C, Arnold JJ, Cameron CE. Incorporation fidelity of the viral RNA-dependent RNA polymerase: a kinetic, thermodynamic and structural perspective. *Virus Res*. 2005;*107*:141–9.
23. Ferrer-Orta C, Arias A, Escarmís C, Verdaguer N. A comparison of viral RNA-dependent RNA polymerases. *Curr Opin Struct Biol*. 2006;*16*:27–34.
24. Wang Y, Anirudhan V, Du R, Cui Q, Rong L. RNA-dependent RNA polymerase of SARS-CoV-2 as a therapeutic target. *J Med Virol*. 2021;*93*:300–10.
25. Nikoomanzar A, Chim N, Yik EJ, Chaput JC. Engineering polymerases for applications in synthetic biology. *Q Rev Biophys*. 2020;*53*:e8.
26. Coulther TA, Stern HR, Beuning PJ. Engineering polymerases for new functions. *Trends Biotechnol*. 2019;*37*:1091–103.
27. Kranaster R, Marx A. Engineered DNA polymerases in biotechnology. *Chembiochem*. 2010;*11*:2077–84.
28. Houlihan G, Arangundy-Franklin S, Holliger P. Engineering and application of polymerases for synthetic genetics. *Curr Opin Biotechnol*. 2017;*48*:168–79.
29. Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, et al. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*. 1988;*239*:487–91.
30. The Nobel Prize in Chemistry 1993. https://www.nobelprize.org/prizes/chemistry/1993/summary/. Accessed 27 Nov 2020.
31. Gibbs RA. DNA amplification by the polymerase chain reaction. *Anal Chem*. 1990;*62*:1202–14.
32. Hemsley A, Arnheim N, Toney MD, Cortopassi G, Galas DJ. A simple method for site-directed mutagenesis using the polymerase chain reaction. *Nucleic Acids Res*. 1989;*17*:6545–51.
33. Fredricks DN, Relman DA. Application of polymerase chain reaction to the diagnosis of infectious diseases. *Clin Infect Dis*. 1999;*29*:475–86.
34. Valones MAA, Guimarães RL, Brandão LAC, de Souza PRE, Carvalho AAT, Crovela S. Principles and applications of polymerase chain reaction in medical diagnostic fields: a review. *Brazilian J Microbiol*. 2009;*40*:1–11.
35. Lynch C, Fleming R. A review of direct polymerase chain reaction of DNA and RNA for forensic purposes. *WIREs Forensic Sci*. 2019;*1*:e1335.
36. Porteous LA, Armstrong JL, Seidler RJ, Watrud LS. An effective method to extract DNA from environmental samples for polymerase chain reaction amplification and DNA fingerprint analysis. *Curr Microbiol*. 1994;*29*:301–7.
37. Ichida JK, Horhota A, Zou K, McLaughlin LW, Szostak JW. High fidelity TNA synthesis by Therminator polymerase. *Nucleic Acids Res*. 2005;*33*:5219–25.
38. Herdewijn P. Nucleic acids with a six-membered 'carbohydrate' mimic in the backbone. *Chem Biodivers*. 2010;*7*:1–59.
39. Veedu RN, Wengel J. Locked nucleic acid nucleoside triphosphates and polymerases: on the way towards evolution of LNA aptamers. *Mol Biosyst*. 2009;*5*:787–92.
40. Wang Z, Xu W, Liu L, Zhu TF. A synthetic molecular system capable of mirror-image genetic replication and transcription. *Nat Chem*. 2016;*8*:698–704.
41. Jiang W, Zhang B, Fan C, Wang M, Wang J, Deng Q, et al. Mirror-image polymerase chain reaction. *Cell Discov*. 2017;*3*:17037.
42. Yang Z, Hutter D, Sheng P, Sismour AM, Benner SA. Artificially expanded genetic information system: a new base pair with an alternative hydrogen bonding pattern. *Nucleic Acids Res*. 2006;*34*:6095–101.
43. Yamashige R, Kimoto M, Takezawa Y, Sato A, Mitsui T, Yokoyama S, et al. Highly specific unnatural base pair systems as a third base pair for PCR amplification. *Nucleic Acids Res*. 2012;*40*:2793–806.
44. Kutzner C, Páll S, Fechner M, Esztermann A, de Groot BL, Grubmüller H. More bang for your buck: improved use of GPU nodes for GROMACS 2018. *J Comput Chem*. 2019;*40*:2418–31.
45. Lee T-S, Cerutti DS, Mermelstein D, Lin C, LeGrand S, Giese TJ, et al. GPU-accelerated molecular dynamics and free energy methods in Amber18: performance enhancements and new features. *J Chem Inf Model*. 2018;*58*:2043–50.
46. Phillips JC, Hardy DJ, Maia JDC, Stone JE, Ribeiro JV, Bernardi RC, et al. Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J Chem Phys*. 2020;*153*:44130.
47. Yang W, Gao Y. Translesion and repair DNA polymerases: diverse structure and mechanism. *Annu Rev Biochem*. 2018;*87*:239–61.
48. Raper AT, Reed AJ, Suo Z. Kinetic mechanism of DNA polymerases: contributions of conformational dynamics and a third divalent metal ion. *Chem Rev*. 2018;*118*:6000–25.
49. Hoitsma NM, Whitaker AM, Schaich MA, Smith MR, Fairlamb MS, Freudenthal BD. Structure and function relationships in mammalian DNA polymerases. *Cell Mol Life Sci*. 2020;*77*:35–59.
50. Wu W-J, Yang W, Tsai M-D. How DNA polymerases catalyse replication and repair with contrasting fidelity. *Nat Rev Chem*. 2017;*1*:68.
51. Kunkel TA. DNA replication fidelity. *J Biol Chem*. 2004;*279*:16895–8.
52. Saribasak H, Rajagopal D, Maul RW, Gearhart PJ. Hijacked DNA repair proteins and unchained DNA polymerases. *Philos Trans R Soc B*. 2009;*364*:605–11.
53. Chaney SG, Campbell SL, Bassett E, Wu Y. Recognition and processing of cisplatin- and oxaliplatin-DNA adducts. *Crit Rev Oncol Hematol*. 2005;*53*:3–11.

54. Svetlov V, Nudler E. Basic mechanism of transcription by RNA polymerase II. *Biochim Biophys Acta, Gene Regul Mech*. 2013;*1829*:20–8.
55. Ninio J. Connections between translation, transcription and replication error-rates. *Biochimie*. 1991;*73*:1517–23.
56. Roßbach S, Ochsenfeld C. Quantum-chemical study of the discrimination against dNTP in the nucleotide addition reaction in the active site of RNA polymerase II. *J Chem Theory Comput*. 2017;*13*:1699–705.
57. Smith EC, Sexton NR, Denison MR. Thinking outside the triangle: replication fidelity of the largest RNA viruses. *Annu Rev Virol*. 2014; *1*:111–32.
58. Beard WA, Wilson SH. Structure and mechanism of DNA polymerase β. *Biochemistry*. 2014;*53*:2768–80.
59. Sale JE, Lehmann AR, Woodgate R. Y-family DNA polymerases and their role in tolerance of cellular DNA damage. *Nat Rev Mol Cell Biol*. 2012;*13*:141–52.
60. Warren JJ, Forsberg LJ, Beese LS. The structural basis for the mutagenicity of O6-methyl-guanine lesions. *Proc Natl Acad Sci U S A*. 2006;*103*:19701–6.
61. Wang D, Bushnell DA, Westover KD, Kaplan CD, Kornberg RD. Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis. *Cell*. 2006;*127*:941–54.
62. Johnson SJ, Taylor JS, Beese LS. Processive DNA synthesis observed in a polymerase crystal suggests a mechanism for the prevention of frameshift mutations. *Proc Natl Acad Sci U S A*. 2003;*100*:3895–900.
63. Kettenberger H, Armache K-J, Cramer P. Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. *Mol Cell*. 2004;*16*:955–65.
64. Boehr DD, Arnold JJ, Moustafa IM, Cameron CE. Structure, dynamics, and fidelity of RNA-dependent RNA polymerases. In: Murakami KS, Trakselis MA, editors. *Nucleic acid polymerases. Nucleic acids and molecular biology*. Berlin, Heidelberg: Springer; 2014. p. 309–33.
65. Werner F. Molecular mechanisms of transcription elongation in Archaea. *Chem Rev*. 2013;*113*:8331–49.
66. Yang W. An overview of Y-family DNA polymerases and a case study of human DNA polymerase η. *Biochemistry*. 2014;*53*:2793–803.
67. Patel SS, Wong I, Johnson KA. Pre-steady-state kinetic analysis of processive DNA replication including complete characterization of an exonuclease-deficient mutant. *Biochemistry*. 1991;*30*:511–25.
68. Reed AJ, Vyas R, Raper AT, Suo Z. Structural insights into the post-chemistry steps of nucleotide incorporation catalyzed by a DNA polymerase. *J Am Chem Soc*. 2017;*139*:465–71.
69. Yin YW, Steitz TA. The structural mechanism of translocation and helicase activity in T7 RNA polymerase. *Cell*. 2004;*116*:393–404.
70. Belogurov GA, Artsimovitch I. The mechanisms of substrate selection, catalysis, and translocation by the elongating RNA polymerase. *J Mol Biol*. 2019;*431*:3975–4006.
71. Vassylyev DG, Vassylyeva MN, Zhang J, Palangat M, Artsimovitch I, Landick R. Structural basis for substrate loading in bacterial RNA polymerase. *Nature*. 2007;*448*:163–8.
72. Westover KD, Bushnell DA, Kornberg RD. Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center. *Cell*. 2004;*119*:481–9.
73. Shu B, Gong P. Structural basis of viral RNA-dependent RNA polymerase catalysis and translocation. *Proc Natl Acad Sci U S A*. 2016; *113*:E4005–14.
74. Wang B, Feig M, Cukier RI, Burton ZF. Computational simulation strategies for analysis of multisubunit RNA polymerases. *Chem Rev*. 2013;*113*:8546–66.
75. Steitz TA. A mechanism for all polymerases. *Nature*. 1998;*391*:231–2.
76. Palermo G, Cavalli A, Klein ML, Alfonso-Prieto M, Dal Peraro M, De Vivo M. Catalytic metal ions and enzymatic processing of DNA and RNA. *Acc Chem Res*. 2015;*48*:220–8.
77. De Vivo M, Dal Peraro M, Klein ML. Phosphodiester cleavage in ribonuclease H occurs via an associative two-metal-aided catalytic mechanism. *J Am Chem Soc*. 2008;*130*:10955–62.
78. Rosta E, Nowotny M, Yang W, Hummer G. Catalytic mechanism of RNA backbone cleavage by ribonuclease H from quantum mechanics/molecular mechanics simulations. *J Am Chem Soc*. 2011;*133*:8934–41.
79. Casalino L, Nierzwicki Ł, Jinek M, Palermo G. Catalytic mechanism of non-target DNA cleavage in CRISPR-Cas9 revealed by ab initio molecular dynamics. *ACS Catal*. 2020;*10*:13596–605.
80. Castro C, Smidansky E, Maksimchuk KR, Arnold JJ, Korneeva VS, Götte M, et al. Two proton transfers in the transition state for nucleotidyl transfer catalyzed by RNA- and DNA-dependent RNA and DNA polymerases. *Proc Natl Acad Sci U S A*. 2007;*104*:4267–72.
81. Sawaya MR, Pelletier H, Kumar A, Wilson SH, Kraut J. Crystal structure of rat DNA polymerase beta: evidence for a common polymerase mechanism. *Science*. 1994;*264*:1930–5.
82. Beard WA, Shock DD, Batra VK, Pedersen LC, Wilson SH. DNA polymerase β substrate specificity: side chain modulation of the "A-rule". *J Biol Chem*. 2009;*284*:31680–9.
83. Batra VK, Beard WA, Shock DD, Krahn JM, Pedersen LC, Wilson SH. Magnesium-induced assembly of a complete DNA polymerase catalytic complex. *Structure*. 2006;*14*:757–66.
84. Freudenthal BD, Beard WA, Shock DD, Wilson SH. Observing a DNA polymerase choose right from wrong. *Cell*. 2013;*154*:157–68.
85. Sawaya MR, Prasad R, Wilson SH, Kraut J, Pelletier H. Crystal structures of human DNA polymerase β complexed with gapped and nicked DNA: evidence for an induced fit mechanism. *Biochemistry*. 1997;*36*:11205–15.
86. Nakamura T, Zhao Y, Yamagata Y, Hua Y, Yang W. Watching DNA polymerase η make a phosphodiester bond. *Nature*. 2012;*487*: 196–201.

87. Rich A. The double helix: a tale of two puckers. *Nat Struct Mol Biol*. 2003;*10*:247–9.
88. Evich M, Spring-Connell AM, Germann MW. Impact of modified ribose sugars on nucleic acid conformation and function. *Heterocycl Commun*. 2017;*23*:155–65.
89. Davies DB, Danyluk SS. Nuclear magnetic resonance studies of 5′-ribo- and deoxyribonucleotide structures in solution. *Biochemistry*. 1974;*13*:4417–34.
90. Feng N, Qi C, Hou Y-J, Zhang Y, Wang D-C, Li D-F. The C2′- and C3′-endo equilibrium for AMP molecules bound in the cystathionine-beta-synthase domain. *Biochem Biophys Res Commun*. 2018;*497*:646–51.
91. Genna V, Vidossich P, Ippoliti E, Carloni P, De Vivo M. A self-activated mechanism for nucleic acid polymerization catalyzed by DNA/RNA polymerases. *J Am Chem Soc*. 2016;*138*:14592–8.
92. Borer PN, LaPlante SR, Kumar A, Zanatta N, Martin A, Hakkinen A, et al. 13C-NMR relaxation in three DNA oligonucleotide duplexes: model-free analysis of internal and overall motion. *Biochemistry*. 1994;*33*:2441–50.
93. Nikolova EN, Bascom GD, Andricioaei I, Al-Hashimi HM. Probing sequence-specific DNA flexibility in A-tracts and pyrimidine-purine steps by nuclear magnetic resonance 13C relaxation and molecular dynamics simulations. *Biochemistry*. 2012;*51*:8654–64.
94. Steffen FD, Khier M, Kowerko D, Cunha RA, Börner R, Sigel RKO. Metal ions and sugar puckering balance single-molecule kinetic heterogeneity in RNA and DNA tertiary contacts. *Nat Commun*. 2020;*11*:104.
95. Doublié S, Tabor S, Long AM, Richardson CC, Ellenberger T. Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution. *Nature*. 1998;*391*:251–8.
96. Li Y, Korolev S, Waksman G. Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of Thermus aquaticus DNA polymerase I: structural basis for nucleotide incorporation. *EMBO J*. 1998;*17*:7514–25.
97. Franklin MC, Wang J, Steitz TA. Structure of the replicating complex of a pol α family DNA polymerase. *Cell*. 2001;*105*:657–67.
98. Berman AJ, Kamtekar S, Goodman JL, Lázaro JM, de Vega M, Blanco L, et al. Structures of phi29 DNA polymerase complexed with substrate: the mechanism of translocation in B-family polymerases. *EMBO J*. 2007;*26*:3494–505.
99. Wang F, Yang W. Structural insight into translesion synthesis by DNA pol II. *Cell*. 2009;*139*:1279–89.
100. Garcia-Diaz M, Bebenek K, Krahn JM, Pedersen LC, Kunkel TA. Role of the catalytic metal during polymerization by DNA polymerase lambda. *DNA Repair*. 2007;*6*:1333–40.
101. Cavanaugh NA, Beard WA, Wilson SH. DNA polymerase β ribonucleotide discrimination: insertion, misinsertion, extension, and coding. *J Biol Chem*. 2010;*285*:24457–65.
102. Williams AA, Darwanto A, Theruvathu JA, Burdzy A, Neidigh JW, Sowers LC. Impact of sugar pucker on base pair and mispair stability. *Biochemistry*. 2009;*48*:11994–2004.
103. Ortín J, Martín-Benito J. The RNA synthesis machinery of negative-stranded RNA viruses. *Virology*. 2015;*479–480*:532–44.
104. Pflug A, Guilligay D, Reich S, Cusack S. Structure of influenza a polymerase bound to the viral RNA promoter. *Nature*. 2014;*516*:355–60.
105. Reguera J, Gerlach P, Cusack S. Towards a structural understanding of RNA synthesis by negative strand RNA viral polymerases. *Curr Opin Struct Biol*. 2016;*36*:75–84.
106. Kouba T, Drncová P, Cusack S. Structural snapshots of actively transcribing influenza polymerase. *Nat Struct Mol Biol*. 2019;*26*:460–70.
107. Wandzik JM, Kouba T, Karuppasamy M, Pflug A, Drncova P, Provaznik J, et al. A structure-based model for the complete transcription cycle of influenza polymerase. *Cell*. 2020;*181*:877–893.e21.
108. Batra VK, Beard WA, Shock DD, Pedersen LC, Wilson SH. Structures of DNA polymerase beta with active-site mismatches suggest a transient abasic site intermediate during misincorporation. *Mol Cell*. 2008;*30*:315–24.
109. Izatt RM, Rytting JH, Hansen LD, Christensen JJ. Thermodynamics of proton dissociation in dilute aqueous solution. V. an entropy titration study of adenosine, pentoses, hexoses, and related compounds. *J Am Chem Soc*. 1966;*88*:2641–5.
110. Batra VK, Perera L, Lin P, Shock DD, Beard WA, Pedersen LC, et al. Amino acid substitution in the active site of DNA polymerase β explains the energy barrier of the nucleotidyl transfer reaction. *J Am Chem Soc*. 2013;*135*:8078–88.
111. Lin P, Pedersen LC, Batra VK, Beard WA, Wilson SH, Pedersen LG. Energy analysis of chemistry for correct insertion by DNA polymerase β. *Proc Natl Acad Sci U S A*. 2006;*103*:13294–9.
112. Florián J, Goodman MF, Warshel A. Computer simulation of the chemical catalysis of DNA polymerases: discriminating between alternative nucleotide insertion mechanisms for T7 DNA polymerase. *J Am Chem Soc*. 2003;*125*:8163–77.
113. Wilson KA, Fernandes PA, Ramos MJ, Wetmore SD. Exploring the identity of the general base for a DNA polymerase catalyzed reaction using QM/MM: the case study of human translesion synthesis polymerase η. *ACS Catal*. 2019;*9*:2543–51.
114. Cisneros GA, Perera L, García-Díaz M, Bebenek K, Kunkel TA, Pedersen LG. Catalytic mechanism of human DNA polymerase λ with Mg2+ and Mn2+ from ab initio quantum mechanical/molecular mechanical studies. *DNA Repair*. 2008;*7*:1824–34.
115. Rungrotmongkol T, Mulholland AJ, Hannongbua S. QM/MM simulations indicate that Asp185 is the likely catalytic base in the enzymatic reaction of HIV-1 reverse transcriptase. *MedChemComm*. 2014;*5*:593–6.
116. Stevens DR, Hammes-Schiffer S. Exploring the role of the third active site metal ion in DNA polymerase η with QM/MM free energy simulations. *J Am Chem Soc*. 2018;*140*:8965–9.
117. Miller BR, Beese LS, Parish CA, Wu EY. The closing mechanism of DNA polymerase I at atomic resolution. *Structure*. 2015;*23*:1609–20.
118. Matute RA, Yoon H, Warshel A. Exploring the mechanism of DNA polymerases by analyzing the effect of mutations of active site acidic groups in polymerase β. *Proteins*. 2016;*84*:1644–57.

119. Roston D, Demapan D, Cui Q. Extensive free-energy simulations identify water as the base in nucleotide addition by DNA polymerase. *Proc Natl Acad Sci U S A*. 2019;*116*:25048–56.
120. Carvalho ATP, Fernandes PA, Ramos MJ. The catalytic mechanism of RNA polymerase II. *J Chem Theory Comput*. 2011;*7*:1177–88.
121. Cheung ACM, Sainsbury S, Cramer P. Structural basis of initial RNA polymerase II transcription. *EMBO J*. 2011;*30*:4755–63.
122. Wang L, Yu X, Hu P, Broyde S, Zhang Y. A water-mediated and substrate-assisted catalytic mechanism for Sulfolobus solfataricus DNA polymerase IV. *J Am Chem Soc*. 2007;*129*:4731–7.
123. Wang L, Broyde S, Zhang Y. Polymerase-tailored variations in the water-mediated and substrate-assisted mechanism for nucleotidyl transfer: insights from a study of T7 DNA polymerase. *J Mol Biol*. 2009;*389*:787–96.
124. Lior-Hoffmann L, Wang L, Wang S, Geacintov NE, Broyde S, Zhang Y. Preferred WMSA catalytic mechanism of the nucleotidyl transfer reaction in human DNA polymerase κ elucidates error-free bypass of a bulky DNA lesion. *Nucleic Acids Res*. 2012;*40*:9193–205.
125. Zhang R, Bhattacharjee A, Field MJ, Salahub DR. Multiple proton relay routes in the reaction mechanism of RNAP II: assessing the effect of structural model. *Proteins*. 2015;*83*:268–81.
126. Genna V, Donati E, De Vivo M. The catalytic mechanism of DNA and RNA polymerases. *ACS Catal*. 2018;*8*:11103–18.
127. Alberts IL, Wang Y, Schlick T. DNA polymerase β catalysis: are different mechanisms possible? *J Am Chem Soc*. 2007;*129*:11100–10.
128. Ivanov I, Tainer JA, McCammon JA. Unraveling the three-metal-ion catalytic mechanism of the DNA repair enzyme endonuclease IV. *Proc Natl Acad Sci U S A*. 2007;*104*:1465–70.
129. Gao Y, Yang W. Capture of a third Mg2+ is essential for catalyzing DNA synthesis. *Science*. 2016;*352*:1334–7.
130. Jamsen JA, Beard WA, Pedersen LC, Shock DD, Moon AF, Krahn JM, et al. Time-lapse crystallography snapshots of a double-strand break repair polymerase in action. *Nat Commun*. 2017;*8*:253.
131. Molina R, Stella S, Redondo P, Gomez H, Marcaida MJ, Orozco M, et al. Visualizing phosphodiester-bond hydrolysis by an endonuclease. *Nat Struct Mol Biol*. 2015;*22*:65–72.
132. Yoon H, Warshel A. Simulating the fidelity and the three Mg mechanism of pol η and clarifying the validity of transition state theory in enzyme catalysis. *Proteins*. 2017;*85*:1446–53.
133. Perera L, Freudenthal BD, Beard WA, Pedersen LG, Wilson SH. Revealing the role of the product metal in DNA polymerase β catalysis. *Nucleic Acids Res*. 2017;*45*:2736–45.
134. Genna V, Gaspari R, Dal Peraro M, De Vivo M. Cooperative motion of a key positively charged residue and metal ions for DNA replication catalyzed by human DNA polymerase-η. *Nucleic Acids Res*. 2016;*44*:2827–36.
135. Perera L, Freudenthal BD, Beard WA, Shock DD, Pedersen LG, Wilson SH. Requirement for transient metal ions revealed through computational analysis for DNA polymerase going in reverse. *Proc Natl Acad Sci U S A*. 2015;*112*:E5228–36.
136. Donati E, Genna V, De Vivo M. Recruiting mechanism and functional role of a third metal ion in the enzymatic activity of 5′ structure-specific nucleases. *J Am Chem Soc*. 2020;*142*:2823–34.
137. Genna V, Colombo M, De Vivo M, Marcia M. Second-shell basic residues expand the two-metal-ion architecture of DNA and RNA processing enzymes. *Structure*. 2018;*26*:40–50.
138. Manigrasso J, Chillón I, Genna V, Vidossich P, Somarowthu S, Pyle AM, et al. Visualizing group II intron dynamics between the first and second steps of splicing. *Nat Commun*. 2020;*11*:2837.
139. Da L-T, Wang D, Huang X. Dynamics of pyrophosphate ion release and its coupled trigger loop motion from closed to open state in RNA polymerase II. *J Am Chem Soc*. 2012;*134*:2399–406.
140. Da L-T, Pardo Avila F, Wang D, Huang X. A two-state model for the dynamics of the pyrophosphate ion release in bacterial RNA polymerase. *PLoS Comput Biol*. 2013;*9*:e1003020.
141. Da L-T, E C, Duan B, Zhang C, Zhou X, Yu J. A jump-from-cavity pyrophosphate ion release assisted by a key lysine residue in T7 RNA polymerase transcription elongation. *PLoS Comput Biol*. 2015; *11*: e1004624.
142. Woo H-J, Liu Y, Sousa R. Molecular dynamics studies of the energetics of translocation in model T7 RNA polymerase elongation complexes. *Proteins*. 2008;*73*:1021–36.
143. Feig M, Burton ZF. RNA polymerase II with open and closed trigger loops: active site dynamics and nucleic acid translocation. *Biophys J*. 2010;*99*:2577–86.
144. Long C, E. C, Da L-T, Yu J. A viral T7 RNA polymerase ratcheting along DNA with fidelity control. *Comput Struct Biotechnol J*. 2019; *17*: 638–644.
145. Da L-T, E C, Shuai Y, Wu S, Su X-D, Yu J. T7 RNA polymerase translocation is facilitated by a helix opening on the fingers domain that may also prevent backtracking. *Nucleic Acids Res*. 2017; *45*: 7909–7921.
146. Unarta IC, Zhu L, Tse CKM, Cheung PP-H, Yu J, Huang X. Molecular mechanisms of RNA polymerase II transcription elongation elucidated by kinetic network models. *Curr Opin Struct Biol*. 2018;*49*:54–62.
147. Silva D-A, Weiss DR, Pardo Avila F, Da L-T, Levitt M, Wang D, et al. Millisecond dynamics of RNA polymerase II translocation at atomic resolution. *Proc Natl Acad Sci U S A*. 2014;*111*:7665–70.
148. Da L-T, Pardo-Avila F, Xu L, Silva D-A, Zhang L, Gao X, et al. Bridge helix bending promotes RNA polymerase II backtracking through a critical and conserved threonine residue. *Nat Commun*. 2016;*7*:11244.
149. Malinen AM, Turtola M, Parthiban M, Vainonen L, Johnson MS, Belogurov GA. Active site opening and closure control translocation of multisubunit RNA polymerase. *Nucleic Acids Res*. 2012;*40*:7442–51.
150. Atis M, Johnson KA, Elber R. Pyrophosphate release in the protein HIV reverse transcriptase. *J Phys Chem B*. 2017;*121*:9557–65.

151. Golosov AA, Warren JJ, Beese LS, Karplus M. The mechanism of the translocation step in DNA replication by DNA polymerase I: a computer simulation analysis. *Structure*. 2010;*18*:83–93.
152. Koshland DE. Application of a theory of enzyme specificity to protein synthesis. *Proc Natl Acad Sci U S A*. 1958;*44*:98–104.
153. Radhakrishnan R, Schlick T. 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*. 2004;*101*:5970–5.
154. Radhakrishnan R, Schlick T. Fidelity discrimination in DNA polymerase β: differing closing profiles for a mismatched (G:A) versus matched (G:C) base pair. *J Am Chem Soc*. 2005;*127*:13245–52.
155. Radhakrishnan R, Arora K, Wang Y, Beard WA, Wilson SH, Schlick T. Regulation of DNA repair fidelity by molecular checkpoints: "gates" in DNA polymerase β's substrate selection. *Biochemistry*. 2006;*45*:15142–56.
156. Meli M, Sustarsic M, Craggs TD, Kapanidis AN, Colombo G. DNA polymerase conformational dynamics and the role of fidelity-conferring residues: insights from computational simulations. *Front Mol Biosci*. 2016;*3*:20.
157. Joyce CM, Benkovic SJ. DNA polymerase fidelity: kinetics, structure, and checkpoints. *Biochemistry*. 2004;*43*:14317–24.
158. Johnson KA. Role of induced fit in enzyme specificity: a molecular forward/reverse switch. *J Biol Chem*. 2008;*283*:26297–301.
159. Schlick T, Arora K, Beard WA, Wilson SH. Perspective: pre-chemistry conformational changes in DNA polymerase mechanisms. *Theor Chem Acc*. 2012;*131*:1287.
160. Tsai Y-C, Johnson KA. A new paradigm for DNA polymerase specificity. *Biochemistry*. 2006;*45*:9675–87.
161. Kellinger MW, Johnson KA. Nucleotide-dependent conformational change governs specificity and analog discrimination by HIV reverse transcriptase. *Proc Natl Acad Sci U S A*. 2010;*107*:7734–9.
162. Kirmizialtin S, Nguyen V, Johnson KA, Elber R. How conformational dynamics of DNA polymerase select correct substrates: experiments and simulations. *Structure*. 2012;*20*:618–27.
163. Wang B, Opron K, Burton ZF, Cukier RI, Feig M. Five checkpoints maintaining the fidelity of transcription by RNA polymerases in structural and energetic details. *Nucleic Acids Res*. 2015;*43*:1133–46.
164. Wang B, Sexton RE, Feig M. Kinetics of nucleotide entry into RNA polymerase active site provides mechanism for efficiency and fidelity. *Biochim Biophys Acta*. 2017;*1860*:482–90.
165. Chao E, Duan B, Yu J. Nucleotide selectivity at a preinsertion checkpoint of T7 RNA polymerase transcription elongation. *J Phys Chem B*. 2017;*121*:3777–86.
166. Long C, Chao E, Da L-T, Yu J. Determining selection free energetics from nucleotide pre-insertion to insertion in viral T7 RNA polymerase transcription fidelity control. *Nucleic Acids Res*. 2019;*47*:4721–35.
167. Yu J. Efficient fidelity control by stepwise nucleotide selection in polymerase elongation abstract: polymerases select nucleotides. *Comput Math Biophys*. 2014;*2*:141–60.
168. Long C, Yu J. Balancing non-equilibrium driving with nucleotide selectivity at kinetic checkpoints in polymerase fidelity control. *Entropy*. 2018;*20*:306.
169. Shen H, Li G. Bridging the missing link between structure and fidelity of the RNA-dependent RNA polymerase from poliovirus through free energy simulations. *J Chem Theory Comput*. 2014;*10*:5195–205.
170. Wu S, Wang J, Pu X, Li L, Li Q. T7 RNA polymerase discriminates correct and incorrect nucleoside triphosphates by free energy. *Biophys J*. 2018;*114*:1755–61.
171. Zhu R, de la Lande A, Zhang R, Salahub DR. Exploring the molecular origin of the high selectivity of multisubunit RNA polymerases by stochastic kinetic models. *Interdiscip Sci*. 2009;*1*:91–8.
172. Ram Prasad B, Kamerlin SCL, Florián J, Warshel A. Prechemistry barriers and checkpoints do not contribute to fidelity and catalysis as long as they are not rate limiting. *Theor Chem Acc*. 2012;*131*:1288.
173. Bakhtina M, Lee S, Wang Y, Dunlap C, Lamarche B, Tsai M-D. Use of viscogens, dNTPαS, and rhodium(III) as probes in stopped-flow experiments to obtain new evidence for the mechanism of catalysis by DNA polymerase β. *Biochemistry*. 2005;*44*:5177–87.
174. Bakhtina M, Roettger MP, Kumar S, Tsai M-D. A unified kinetic mechanism applicable to multiple DNA polymerases. *Biochemistry*. 2007;*46*:5463–72.
175. Bakhtina M, Roettger MP, Tsai M-D. Contribution of the reverse rate of the conformational step to polymerase β fidelity. *Biochemistry*. 2009;*48*:3197–208.
176. Oertell K, Florián J, Haratipour P, Crans DC, Kashemirov BA, Wilson SH, et al. A transition-state perspective on Y-family DNA polymerase η fidelity in comparison with X-family DNA polymerases λ and β. *Biochemistry*. 2019;*58*:1764–73.
177. Lin P, Batra VK, Pedersen LC, Beard WA, Wilson SH, Pedersen LG. Incorrect nucleotide insertion at the active site of a G:A mismatch catalyzed by DNA polymerase β. *Proc Natl Acad Sci U S A*. 2008;*105*:5670–4.
178. Ram Prasad B, Warshel A. Prechemistry versus preorganization in DNA replication fidelity. *Proteins*. 2011;*79*:2900–19.
179. Florián J, Goodman MF, Warshel A. Computer simulations of protein functions: searching for the molecular origin of the replication fidelity of DNA polymerases. *Proc Natl Acad Sci U S A*. 2005;*102*:6819–24.
180. Florián J, Goodman MF, Warshel A. Theoretical investigation of the binding free energies and key substrate-recognition components of the replication fidelity of human DNA polymerase β. *J Phys Chem B*. 2002;*106*:5739–53.
181. Florián J, Warshel A, Goodman MF. Molecular dynamics free-energy simulations of the binding contribution to the fidelity of T7 DNA polymerase. *J Phys Chem B*. 2002;*106*:5754–60.
182. Rucker R, Oelschlaeger P, Warshel A. A binding free energy decomposition approach for accurate calculations of the fidelity of DNA polymerases. *Proteins*. 2010;*78*:671–80.

183. Ucisik MN, Hammes-Schiffer S. Relative binding free energies of adenine and guanine to damaged and undamaged DNA in human DNA polymerase η: clues for fidelity and overall efficiency. *J Am Chem Soc*. 2015;*137*:13240–3.
184. Yoon H, Warshel A. The control of the discrimination between dNTP and rNTP in DNA and RNA polymerase. *Proteins*. 2016;*84*: 1616–24.
185. Tang K-H, Niebuhr M, Tung C-S, Chan H, Chou C-C, Tsai M-D. Mismatched dNTP incorporation by DNA polymerase β does not proceed via globally different conformational pathways. *Nucleic Acids Res*. 2008;*36*:2948–57.
186. Foley MC, Schlick T. Relationship between conformational changes in pol λ's active site upon binding incorrect nucleotides and mismatch incorporation rates. *J Phys Chem B*. 2009;*113*:13035–47.
187. Li Y, Schlick T. "Gate-keeper" residues and active-site rearrangements in DNA polymerase μ help discriminate non-cognate nucleotides. *PLoS Comput Biol*. 2013;*9*:e1003074.
188. Yang L, Beard WA, Wilson SH, Broyde S, Schlick T. Highly organized but pliant active site of DNA polymerase β: compensatory mechanisms in mutant enzymes revealed by dynamics simulations and energy analyses. *Biophys J*. 2004;*86*:3392–408.
189. Ucisik MN, Hammes-Schiffer S. Effects of active site mutations on specificity of nucleobase binding in human DNA polymerase η. *J Phys Chem B*. 2017;*121*:3667–75.
190. Graham SE, Syeda F, Cisneros GA. Computational prediction of residues involved in fidelity checking for DNA synthesis in DNA polymerase I. *Biochemistry*. 2012;*51*:2569–78.
191. Yeager A, Humphries K, Farmer E, Cline G, Miller BR. Investigation of nascent base pair and polymerase behavior in the presence of mismatches in DNA polymerase I using molecular dynamics. *J Chem Inf Model*. 2018;*58*:338–49.
192. Xiang Y, Oelschlaeger P, Florián J, Goodman MF, Warshel A. Simulating the effect of DNA polymerase mutations on transition-state energetics and fidelity: evaluating amino acid group contribution and allosteric coupling for ionized residues in human pol β. *Biochemistry*. 2006;*45*:7036–48.
193. Klvaňa M, Murphy DL, Jeřábek P, Goodman MF, Warshel A, Sweasy JB, et al. Catalytic effects of mutations of distant protein residues in human DNA polymerase β: theory and experiment. *Biochemistry*. 2012;*51*:8829–43.
194. Genna V, Carloni P, De Vivo M. A strategically located Arg/Lys residue promotes correct base paring during nucleic acid biosynthesis in polymerases. *J Am Chem Soc*. 2018;*140*:3312–21.
195. Mota CS, Gonçalves AMD, de Sanctis D. Deinococcus radiodurans DR2231 is a two-metal-ion mechanism hydrolase with exclusive activity on dUTP. *FEBS J*. 2016;*283*:4274–90.
196. Vidossich P, Castañeda Moreno LE, Mota C, de Sanctis D, Pietro MG, De Vivo M. Functional implications of second-shell basic residues for dUTPase DR2231 enzymatic specificity. *ACS Catal.* 2020;*10*:13825–33.
197. Jorgensen WL. Computer-aided discovery of anti-HIV agents. *Bioorg Med Chem*. 2016;*24*:4768–78.
198. Zafar MK, Maddukuri L, Ketkar A, Penthala NR, Reed MR, Eddy S, et al. A small-molecule inhibitor of human DNA polymerase η potentiates the effects of cisplatin in tumor cells. *Biochemistry*. 2018;*57*:1262–73.
199. Zhang L, Zhou R. Structural basis of the potential binding mechanism of remdesivir to SARS-CoV-2 RNA-dependent RNA polymerase. *J Phys Chem B*. 2020;*124*:6955–62.
200. Koulgi S, Jani V, Uppuladinne MVN, Sonavane U, Joshi R. Remdesivir-bound and ligand-free simulations reveal the probable mechanism of inhibiting the RNA dependent RNA polymerase of severe acute respiratory syndrome coronavirus 2. *RSC Adv*. 2020;*10*: 26792–803.
201. Wakchaure PD, Ghosh S, Ganguly B. Revealing the inhibition mechanism of RNA-dependent RNA polymerase (RdRp) of SARS-CoV-2 by remdesivir and nucleotide analogues: a molecular dynamics simulation study. *J Phys Chem B*. 2020;*124*:10641–52.
202. Aranda J, Orozco M. RNA-dependent RNA polymerase from SARS-CoV-2. Mechanism of reaction and inhibition by remdesivir. *bioRxiv*. 2020. https://doi.org/10.1101/2020.06.21.163592.
203. McDonald JP, Hall A, Gasparutto D, Cadet J, Ballantyne J, Woodgate R. Novel thermostable Y-family polymerases: applications for the PCR amplification of damaged or ancient DNAs. *Nucleic Acids Res*. 2006;*34*:1102–11.
204. Wu J, de Paz A, Zamft BM, Marblestone AH, Boyden ES, Kording KP, et al. DNA binding strength increases the processivity and activity of a Y-family DNA polymerase. *Sci Rep*. 2017;*7*:4756.
205. Wang L, Liang C, Wu J, Liu L, Tyo KEJ. Increased processivity, misincorporation, and nucleotide incorporation efficiency in Sulfolobus solfataricus Dpo4 thumb domain mutants. *Appl Environ Microbiol*. 2017;*83*:e01013–7.
206. Grúz P, Pisani FM, Shimizu M, Yamada M, Hayashi I, Morikawa K, et al. Synthetic activity of Sso DNA polymerase Y1, an archaeal DinB-like DNA polymerase, is stimulated by processivity factors proliferating cell nuclear antigen and replication factor C. *J Biol Chem*. 2001;*276*:47394–401.
207. Wang Y, Prosen DE, Mei L, Sullivan JC, Finney M, Vander Horn PB. A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance in vitro. *Nucleic Acids Res*. 2004;*32*:1197–207.
208. Akabayov B, Akabayov SR, Lee S-J, Tabor S, Kulczyk AW, Richardson CC. Conformational dynamics of bacteriophage T7 DNA polymerase and its processivity factor, *Escherichia coli* thioredoxin. *Proc Natl Acad Sci U S A*. 2010;*107*:15033–8.
209. Marx A, Betz K. The structural basis for processing of unnatural base pairs by DNA polymerases. *Chem A Eur J*. 2020;*26*:3446–63.
210. Ouaray Z, Singh I, Georgiadis MM, Richards NGJ. Building better enzymes: molecular basis of improved non-natural nucleobase incorporation by an evolved DNA polymerase. *Protein Sci*. 2020;*29*:455–68.
211. Arangundy-Franklin S, Taylor AI, Porebski BT, Genna V, Peak-Chew S, Vaisman A, et al. A synthetic genetic polymer with an uncharged backbone chemistry based on alkyl phosphonate nucleic acids. *Nat Chem*. 2019;*11*:533–42.

212. Kropp HM, Dürr SL, Peter C, Diederichs K, Marx A. Snapshots of a modified nucleotide moving through the confines of a DNA polymerase. *Proc Natl Acad Sci U S A*. 2018;*115*:9992–7.
213. De Vivo M, Masetti M, Bottegoni G, Cavalli A. Role of molecular dynamics and related methods in drug discovery. *J Med Chem*. 2016; *59*:4035–61.
214. De Vivo M, Cavalli A. Recent advances in dynamic docking for drug discovery. *WIREs Comput Mol Sci*. 2017;*7*:e1320.
215. Kokh DB, Amaral M, Bomke J, Grädler U, Musil D, Buchstaller H-P, et al. Estimation of drug-target residence times by τ-random acceleration molecular dynamics simulations. *J Chem Theory Comput.* 2018;*14*:3859–69.
216. Mollica L, Decherchi S, Zia SR, Gaspari R, Cavalli A, Rocchia W. Kinetics of protein-ligand unbinding via smoothed potential molecular dynamics simulations. *Sci Rep*. 2015;*5*:11539.
217. Vamathevan J, Clark D, Czodrowski P, Dunham I, Ferran E, Lee G, et al. Applications of machine learning in drug discovery and development. *Nat Rev Drug Discov*. 2019;*18*:463–77.
218. Seeliger D, de Groot BL. Protein thermostability calculations using alchemical free energy simulations. *Biophys J*. 2010;*98*:2309–16.
219. Beierlein FR, Kneale GG, Clark T. Predicting the effects of basepair mutations in DNA-protein complexes by thermodynamic integration. *Biophys J*. 2011;*101*:1130–8.
220. Yang KK, Wu Z, Arnold FH. Machine-learning-guided directed evolution for protein engineering. *Nat Methods*. 2019;*16*:687–94.

How to cite this article: Geronimo I, Vidossich P, Donati E, De Vivo M. Computational investigations of polymerase enzymes: Structure, function, inhibition, and biotechnology. *WIREs Comput Mol Sci*. 2021;e1534. https://doi.org/10.1002/wcms.1534