

Alma Mater Studiorum - Università di Bologna

DOTTORATO DI RICERCA IN
DATA SCIENCE AND COMPUTATION

Ciclo 33

Settore Concorsuale: 03/D1 - CHIMICA E TECNOLOGIE FARMACEUTICHE, TOSSICOLOGICHE
E NUTRACEUTICO - ALIMENTARI

Settore Scientifico Disciplinare: CHIM/08 - CHIMICA FARMACEUTICA

COMPUTATIONAL INVESTIGATION OF THE MOLECULAR MECHANISMS
REGULATING NUCLEIC ACID PROCESSING METALLOENZYMES

Presentata da: Elisa Donati

Coordinatore Dottorato

Andrea Cavalli

Supervisore

Marco De Vivo

Co-supervisore

Andrea Cavalli

Esame finale anno 2022

*“e sun chi a miä
tréi camixe de vellûu
dui cuverte u mandurlin
e 'n cãmà de legnu dûu*

*e 'nte 'na beretta neigra
a teu fotu da fantinn-a
pe puèi baxâ ancún Zena
'nscià teu bucca in naftalin-a”*

Fabrizio De Andrè,
Crêuza de mã, D'ã mã Riva, 1983.

ACKNOWLEDGEMENTS

With the completion of this thesis a period of life is drawing to a close. However, new doors will open up soon. The words written here enclose the hard work done in the past four years as well as the knowledge that I have acquired during my PhD period. During this period many things have changed, including myself. These acknowledgements are written to give recognition to all the people that have assist such life path.

The first thanks go to my parents. I would not be here without your support and I am very grateful for all the efforts you made during these years. I moved away more than you wanted, but I always knew that I could rely on you, regardless to where I am.

I thank Marco, my supervisor, for giving me the opportunity to grow professionally and also personally. Thank you, because you have built an amazing research group, and I had the pleasure to see it evolving. It has been very important to be in a friendly, supporting and cooperating environment. Thanks also for all the opportunities you gave me to travel, meet different people and visit different places. I will keep those memories with me forever.

A huge thanks is reserved to all the colleagues who have been with me during these years. Sandro, Luca, Grazisa, Claudio, Josephine, Vito, Sergio, Mattia, Dorothea, Valentina, Tony, Pepe, Nicoletta, Marco, Amos, Federico, Michela, Francesco, Ilaria, Isabella, Jissy, Pietro, Andrea, Alessandro, Adam, Ankita, Inacrist, Matilde, Manuel, Maria Antonietta, David and Alessia. Thank you all for the useful discussion, the constructive advice, the time spent together laughing and supporting each other. You all have really made special my days here.

I would like to thank Prof. Ursula Roethlisberger, for hosting me in her laboratory at EPFL in Lausanne. I loved the city, the research and all the amazing people in her group: Ariadni, Polydefkis, Marko, Simon, Farzaneh, Viacheslav, François, Justin, Letizia, Guido, Mathias, Murat, Thibaud, Paramvir, Siri, Swarnendu and of course Karin. Thank you all for the amazing time spent together and the great discussion.

A special thought goes to Jacopo and Sebastian, who started the PhD with me. I couldn't ask for anyone better than you two as friends and colleagues. I consider myself very lucky and proud for all the experiences we lived together. Thank you!

Lastly, thanks to all the people of 'vicoli'. Genova will always be partly my home.

TABLE OF CONTENTS

<u>ACKNOWLEDGEMENTS</u>	IV
<u>TABLE OF CONTENTS</u>	VI
<u>ABSTRACT</u>	VIII
<u>Chapter I.</u> <u>Nucleic acids processing enzymes: Polymerases and Nucleases</u>	1
1. Polymerases	1
1.1. Classification, Properties and Structure of Polymerases	2
1.2. Catalytic cycle and Chemical Reaction Mechanism	4
1.3. Role of the Third Metal Ion in Polymerases	6
1.4. Strategically Located Basic Residues	8
1.5. Polymerase Applications: Therapeutics, Biotechnology and Synthetic Biology	8
2. Nucleases	11
2.1. Diversity and Classification of Nucleases	12
2.2. Reaction Mechanism and Catalysis	14
2.3. Expanded two-metal-ion architecture	15
2.4. Role of Metal Ions	16
2.5. Metal-Dependent and Independent Catalysis	16
2.6. A Third Metal Ion for Catalysis	18
2.7. RAD2/FEN superfamily of nucleases	20
2.8. Nuclease Applications: Genome Editing and Therapeutic Targets	22
3. Scope of the Thesis	24
<u>Chapter II.</u> <u>Computational Theory and Principles</u>	25
1. Molecular Dynamics Simulations	25
1.1. Challenges of Molecular Dynamics Simulations for Nucleic Acid Processing Metalloenzymes	27
2. Enhanced Sampling Techniques	29
1.2. Metadynamics	30
1.3. Well-Tempered Metadynamics with a Confined Procedure	32
1.4. Path Collective Variable	33
<u>Chapter III.</u> <u>Recruiting Mechanism and Functional Role of a Third Metal Ion in the Enzymatic Activity of 5' Structure-Specific Nucleases</u>	35
Abstract	35
1. Introduction	35
2. Results	38

3. Discussion	47
4. Conclusions	52
5. Material and Methods	52
<u>Chapter IV.</u> <u>Molecular mechanism of phosphate steering for DNA binding, cleavage localization, and substrate release in nucleases</u>	<u>55</u>
Abstract	55
1. Introduction	55
2. Results and Discussion	58
3. Conclusions	69
4. Material and Methods	70
<u>Chapter V.</u> <u>Quercetin and luteolin are single-digit micromolar inhibitors of the SARS-CoV-2 RNA-dependent RNA polymerase</u>	<u>74</u>
Abstract	74
1. Introduction	74
2. Results and Discussion	76
3. Conclusions	82
4. Material and Methods	83
<u>Chapter VI.</u> <u>Final Remarks and Perspective</u>	<u>87</u>
<u>Appendix A</u>	<u>89</u>
<u>Appendix B</u>	<u>108</u>
<u>Appendix C</u>	<u>124</u>
<u>Bibliography</u>	<u>132</u>

ABSTRACT

Polymerases and nucleases are enzymes processing DNA and RNA. They are involved in crucial processes for cell life by performing the extension and the cleavage of nucleic acid chains during genome replication and maintenance. Additionally, both enzymes are often associated to several diseases, including cancer. In order to catalyze the reaction, most of them operate via the two-metal-ion mechanism. For this, despite showing relevant differences in structure, function and catalytic properties, they share common catalytic elements, which comprise the two catalytic ions and their first-shell acidic residues. Notably, recent studies of different metalloenzymes revealed the recurrent presence of additional elements surrounding the active site, thus suggesting an extended two-metal-ion-centered architecture. However, whether these elements have a catalytic function and what is their role is still unclear. In this work, using state-of-the-art computational techniques, second- and third-shell elements are showed to act in metallonucleases favoring the substrate positioning and leaving group release. In particular, in hExo1 a transient third metal ion is recruited and positioned near the two-metal-ion site by a structurally conserved acidic residue to assist the leaving group departure. Similarly, in hFEN1 second- and third-shell Arg/Lys residues operate the phosphate steering mechanism through (i) substrate recruitment, (ii) precise cleavage localization, and (iii) leaving group release. Importantly, structural comparisons of hExo1, hFEN1 and other metallonucleases suggest that similar catalytic mechanisms may be shared by other enzymes. Overall, the results obtained provide an extended vision on parallel strategies adopted by metalloenzymes, which employ divalent metal ion or positively charged residues to ensure efficient and specific catalysis. Furthermore, these outcomes may have implications for de novo enzyme engineering and/or drug design to modulate nucleic acid processing.

Chapter I. Nucleic acids processing enzymes: Polymerases and Nucleases

1. Polymerases

Nucleic acid polymerases (Pols) are crucial for cell survival and life propagation.¹⁻³ These enzymes play a major role in gene expression, regulation, transcription (DNA → RNA), reverse transcription (RNA → DNA) and DNA damage repair.⁴⁻⁶ Therefore, Pols have been implicated in various human diseases, including viral and bacterial infection, neurodegenerative diseases, and cancer.⁷⁻⁹ Specifically, mutations on the gene encoding DNA Pol β (i.e. POLB) are linked to adenocarcinoma of the colon,¹⁰ while overexpression of POLK and POLQ (genes encoding DNA Pol κ and Pol θ) are associated to lung cancer and colorectal, breast, and nonsmall cell lung cancers, respectively.¹¹⁻¹³ Additionally, POLH (DNA Pol η) mutations are linked to a variant type of xeroderma pigmentosum¹⁴ and POLG mutations to several types of mitochondrial diseases.^{15,16} On the other hand, RNA Pols are essential for replication of viral RNA, like in hepatitis C virus, poliovirus and severe acute respiratory syndrome (SARS) infections. In the last couple of years, many efforts have been directed for searching an effective treatment for coronavirus disease 2019 (COVID-19), with a significant focus on the RNA-dependent RNA Pol (RdRp) of SARS coronavirus 2 (SARS-CoV-2). Consequently, DNA and RNA Pols are an attractive pharmacological target for small-molecule inhibitors.¹⁷⁻¹⁹ Nevertheless Pols are of particular interest also for synthetic biology²⁰ to develop a wide variety of applications ranging from targeted therapies to storage and propagation of genetic information.²¹⁻²³ Moreover, Pols are also utilized in biotechnological applications, such as the polymerase chain reaction (PCR) method.²⁴

Hence, a deeper and more extensive understanding of molecular mechanisms and functional properties that regulate DNA and RNA polymerase catalysis is pivotal to further push forward the research towards new frontiers positively impacting human health. In this regard, the computational approach significantly contributed to this end. The constant and substantial advance in the performance of molecular modeling techniques makes them a

powerful and useful tool in Pol research,^{25–27} providing a molecular-level understanding of the mechanisms that regulate the catalytic cycle and functional properties of polymerases.

1.1. Classification, Properties and Structure of Polymerases

Pol enzymes synthesize long chains of nucleic acids by adding consecutively a deoxyribonucleoside triphosphate (dNTP) or a ribonucleoside triphosphate (NTP), to the 3'-end of the growing primer. Pols can be classified based on their function, sequence and structure.

The classification by chemical function (i.e. the type of template – DNA vs RNA, and the nature of the incoming nucleotide – dNTP vs NTP) divided Pols into five classes: i) the DNA-dependent DNA polymerases (DdDp), which add dNTP guided by a DNA template, ii) the reverse transcriptase (RT) or RNA-dependent DNA polymerases (RdDp), which incorporate dNTP complementary to RNA template, iii) the DNA-dependent RNA polymerases (DdRp), which in contrast add NTP complementary to DNA template, iv) RNA replicase or RNA-dependent RNA polymerases (RdRp), which add NTP opposite to a RNA template, and v) the template-less RNA elongation, which add nucleotide in a template-independent manner.

Within the first functional class, i.e. DdDp, the Pols can be further subdivided on the basis of sequence homology into six families: i) family A (such as Pol I, Pol γ , θ , ν), ii) family B (e.g. Pol II, Pol α , δ , ϵ , ζ), iii) family C (e.g. Pol III), iv) family D (e.g. Pol D), v) family X (e.g. Pol IV, Pol β , λ , μ), and vi) family Y (such as Pol V, Pol η , ι , κ).²⁸ The DNA Pols differ significantly in biochemical properties like fidelity and processivity, which supply DNA Pols for a specific biological function. Fidelity is quantified as the ratio of the catalytic efficiencies for incorrect and correct nucleotide incorporation, while processivity is measure as the average number of nucleotides (nt) incorporated during a single binding event.²⁸ Hence, A- and B-family DNA Pols, which play a major role in DNA replication, have the highest fidelity (i.e. 10^{-5} to 10^{-6} without exonuclease activity)^{29,30} and processivity (>100 nt/s).²³ Conversely, X- and Y-family DNA Pols, which take part to DNA repair and damage response mechanisms,^{31,32} have a lower fidelity (e.g. for Y-family 10^{-1} to 10^{-3})^{29,30} and processivity (e.g. for Y-family <20 nt/s).³³

On the other hand, RNA Pols in eukaryotes are classified into 3 groups based on their cellular function: i) RNA Pol I, which transcribes ribosomal RNA genes, ii) RNA Pol II,

which synthesizes messenger RNA and a subset of small noncoding RNAs, and iii) RNA Pol III, which synthesizes transfer RNAs, 5S RNA, and the majority of small noncoding RNAs.³⁴ Similar differences in fidelity and processivity are found for RNA Pols. Indeed, while Pol II is reported to have high fidelity, with a transcription error rate of 10^{-5} ,³⁵ viral RNA Pols are medium-fidelity Pols (i.e. 10^{-3} to 10^{-5}), thus allowing RNA viruses easy adaption to different host cell environments.³⁶

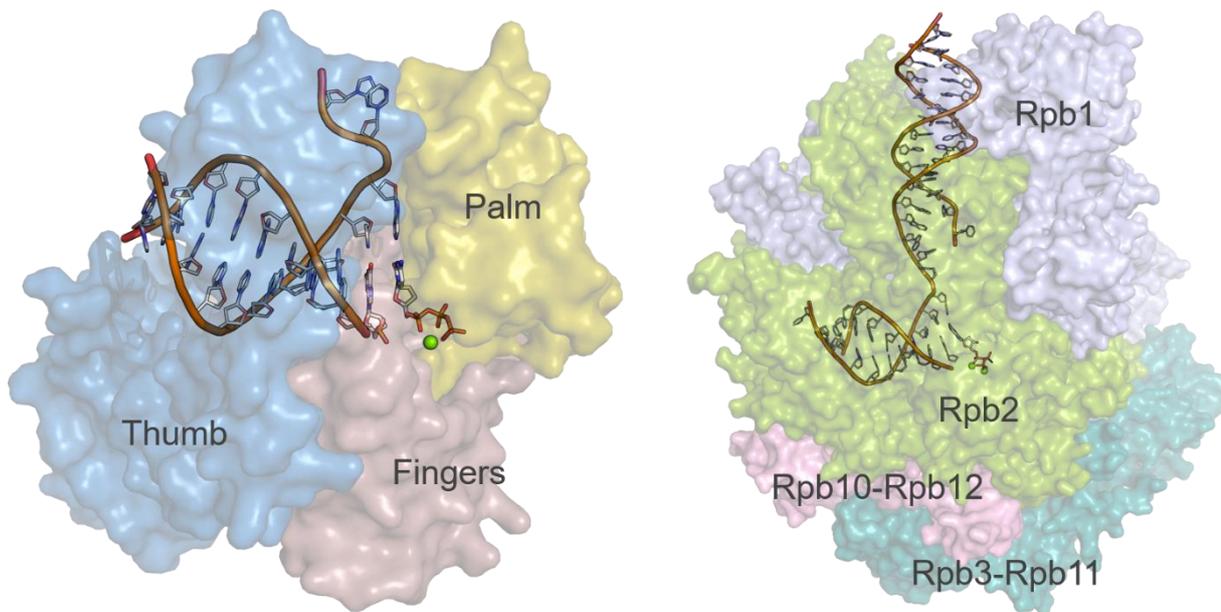


Figure I.1 Crystal structures of (left) DNA polymerase (Pol) I (PDB ID 2HVI³⁷) and (right) multiple subunits RNA Pol II (PDB ID 2E2H³⁸) in complex with catalytic metal ions (in sphere, colored in green), the incoming nucleotide and DNA/RNA substrate, respectively. DNA Pols adopt a right-hand architecture, which consists of palm (yellow), fingers (light brown), and thumb (blue) subdomains. The minimal configuration capable to perform polymerization in multiple subunits RNA Pols is formed by the catalytic subunits, i.e. Rpb1 (in light blue) and Rpb2 (in green), and the assembly platform, i.e. Rpb3-Rpb11 (in teal) and Rpb10-Rpb12 (in pink).

Taken together, all these data suggest a degree of variance for biochemical properties among Pols, which in turn can be modulated through different mechanisms. These diversities are evident also in Pols structures, depending on their biological function. For example, A- and B-family possess a 3'-5' proofreading domain that improves the essential fidelity, in X-family a lyase domain is present and Y-family is characterized by a smaller structure that can establish only few contacts with the incoming dNTP, thus contributing to the lower processivity and fidelity. Commonly, the catalytic domain can be divided based on the architecture into palm (containing the catalytic residues), fingers, and thumb (interacting with the incoming dNTP and DNA substrates, respectively) subdomains

(Figure I.1). However, some exceptions are possible. For instance, DNA Pol X from African swine fever virus has only the palm and fingers subdomains, and in the X-family DNA Pol a different nomenclature for the subdomains is adopted, i.e. D (DNA binding), C (catalytic), and N (nascent base-pair binding) subdomains.

Comparatively, RNA Pols share a more conserved structure. An initial differentiation concerns the single subunit (from virus) and multiple subunits RNA Pols (from Archaea, Bacteria, and Eukarya). Within the single subunit, RNA Pols adopt the right-hand architecture with seven conserved structural motifs: motifs A–E (in the palm subdomain), and motifs F and G (in the fingers subdomain). Multiple subunits RNA Pols are named based on the domain of life they belong to: bacterial RNA Pols subunits use Greek letter, archaeal and eukaryotic RNA Pols subunits are named Rpo/Rpb, respectively, followed by a number (Figure I.1).

1.2. Catalytic cycle and Chemical Reaction Mechanism

The catalytic cycle (Figure I.2) whereby Pols elongate the nascent D(R)NA strand consists of a multifaceted stepwise process with both chemical and physical steps. During each catalytic cycle, Pols add one nucleotide at the 3' primer terminus, thus extending the growing primer strand (Figure I.2). This occurs via S_N2 -like phosphoryl-transfer reaction assisted by the well-characterized two-metal-ion mechanism (Figure I.3).^{39–41}

In order to allow an efficient nucleotide selection, binding, and chemical reaction, the enzyme may undergo functional conformational motions (such as open \rightarrow closed conformational change mainly of the fingers subdomain) and structural rearrangements (e.g. the active site rearrangement). Once the reaction has been performed and the nucleotide incorporated, the pyrophosphate (PPi) group is released and the enzyme undergoes a reverse conformational change (such as closed \rightarrow open).³³ However, the causality and the order of these last two events is still debated.^{42,43} Eventually, either the Pol translocates by one base pair along the DNA, leaving the active site free for the binding and incorporation of the next nucleotide, or DNA/Pol complex dissociates thus allowing Pol to bind a new substrate.³³

As mentioned above, Pols operate via the two-metal-ion mechanism in order to catalyze the nucleotidyl transfer reaction.⁴⁴ The two divalent metal ions, i.e. M_A and M_B , (typically Mg^{2+} or Mn^{2+}) facilitate nucleophile formation, stabilize the transition state, and assist the

leaving group release.⁴⁵ Specifically, M_A enhances 3'-OH deprotonation by lowering its pK_a , to favor the negatively charged 3'-O⁻ form, while in parallel M_B facilitates products formation (Figure I.3).⁴⁶

Pols Catalytic Cycle

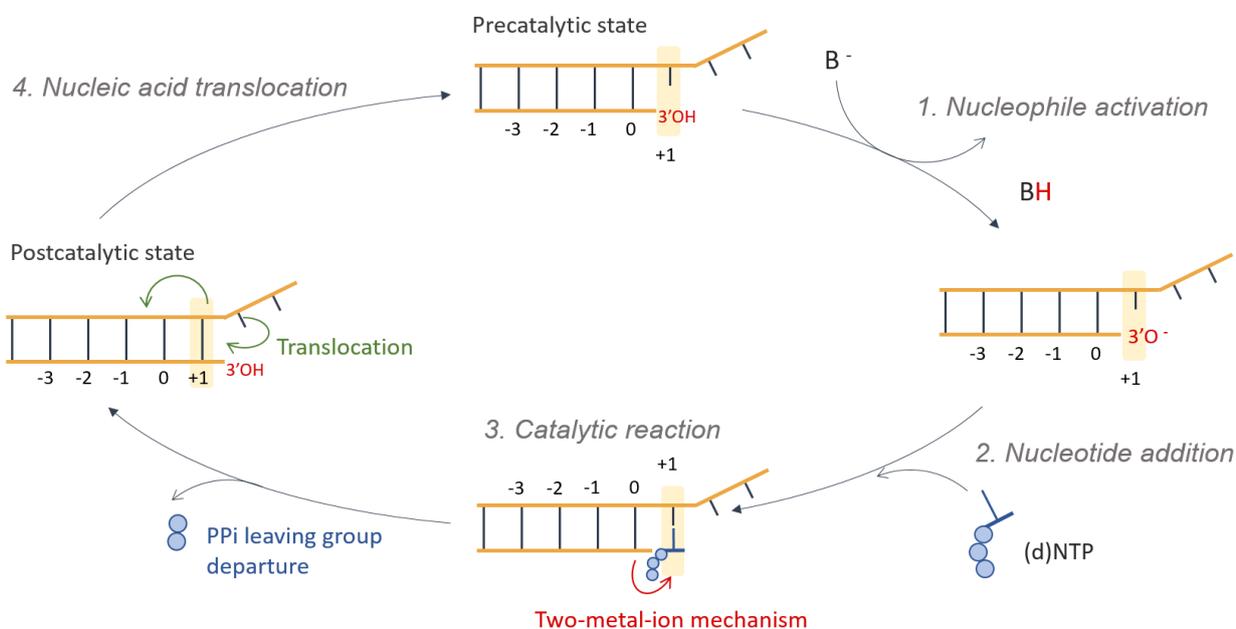


Figure I.2 Schematic representation of nucleotide addition cycle of DNA/RNA polymerases. The cycle is composed of four main steps: 1) the nucleophile activation, in which the 3'-OH end (in red) of the DNA/RNA substrate in the precatalytic state is deprotonated; 2) the nucleotide addition, i.e. (d)NTP (in stick and sphere, colored in blue) is placed in the active site (highlighted in yellow) to form base pair interactions with the complementary base (i.e. +1); 3) the catalytic reaction, with the consequent formation of the postcatalytic state and the release of the PPi leaving group; and 4) the nucleic acid translocation to restore the active site for a new catalytic cycle.

Such efficient mechanism is shared with other enzymes, like type II topoisomerases and several nucleases, which instead perform the opposite reaction of D(R)NA cleavage.⁴⁷⁻⁴⁹ Each of the catalytic ion is coordinated by two or three conserved acidic residues (typically Asp or Glu), the 3'-OH of the primer strand coordinates M_A while the triphosphate group of the incoming nucleotide coordinates M_B .^{29,34,50} Once the active site is assembled and the reactant state formed, the deprotonation of the 3'-OH initiates the reaction forming the nucleophile which has to attack the P α atom of the nucleotidyl triphosphate. It follows a S_N2-like reaction where a pentacovalent phosphate transition state is formed leading to the typical inversion of the umbrella of the scissile phosphate.⁴⁴ Concurrently, the PPi

molecule is produced and subsequently released with the consequent nucleic acid translocation to restore the active site for a new catalytic cycle.^{46,51}

Two-metal-ion mechanism : Polymerization

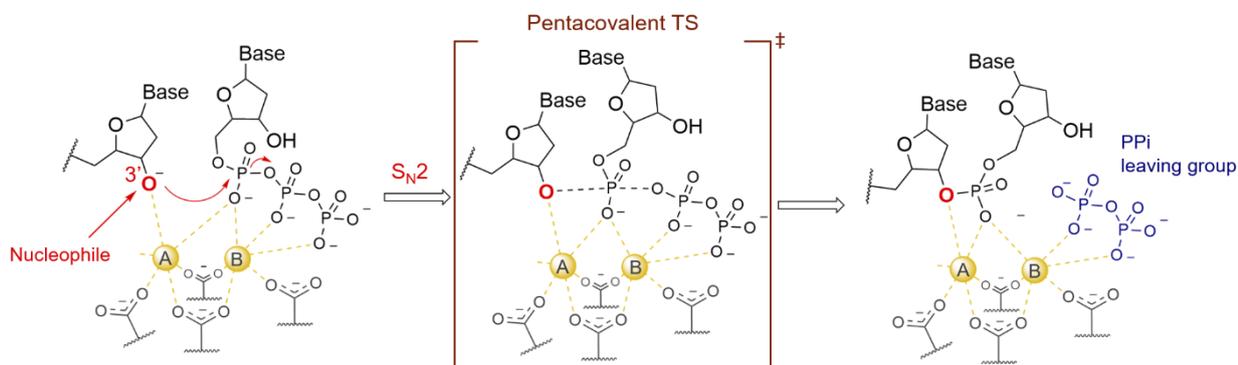


Figure I.3 Catalytic reaction scheme for the two-metal-ion mechanism adopted by Pols during nucleic acids polymerization. The chemical reaction is a S_N2-like phosphoryl-transfer reaction catalyzed by two divalent metal ions, i.e. A and B (in sphere, colored in yellow). Both metals actively participate in catalysis by i) facilitating the nucleophile, i.e. 3'O⁻ (in red), ii) stabilizing the pentacovalent transition state, i.e. TS, and iii) assisting the PPi leaving group departure in product state.

The rapid evolution of experimental techniques, such as crystallography and the growing interest on Pols have contributed to reveal many key aspects of Pols catalysis.⁵² Time-resolved X-ray crystallography and cryogenic electron microscopy (cryo-EM) enabled researchers to capture and establish the main points of the catalytic cycle and reaction mechanism of Pols. Intriguingly, several ternary Pol/D(R)NA/d(r)NTP crystal structures have revealed the presence of a transient third metal ion bound to the catalytic site (Figure I.4). Thus, it suggests that additional metal ion may be actively involved in catalysis, although its exact role is unclear.

1.3. Role of the Third Metal Ion in Polymerases

Several examples of DNA Pols crystallographic data resolved a third divalent metal ion (M_C) close and above the catalytic active site, thus showing an additional new player in the two-metal-ion mechanism. In Figure I.4 are reported the available DNA Pols structures resolved in the presence of the third ion.² It includes ternary Pol η,^{51,53} Pol β,^{17,54} Pol μ^{55,56} complexes, and also the engineered Pol Kod-RI, an artificial threose nucleic acid (TNA) polymerase.⁵⁷ The surprising finding of a third metal ion in DNA Pols has opened a stimulating and unresolved debate about the specific role of this additional ion for Pol catalysis. Notably, the presence of such third solvent-exposed cation is recurrent in other

nucleic acids processing enzymes. For example, it was found that in *D. mobiliz* homing endonuclease (I-DmoI), the third metal ion is essential to achieving the proper geometry for phosphodiester bond hydrolysis.⁵⁸ More recently, a third metal ion was found in time-resolved crystal structures of the human exonuclease 1 (hExo-1)⁵⁹ while *in crystallo* reaction intermediates of RNA ribonuclease H1 (RNase H1) captured multiple metal ions (K^+ and Mg^{2+}) transiently bound in the vicinity of the two-metal-ion active site.⁶⁰ Although the presence of the third ion seems to be functional, there is debate on its mechanistic role. Indeed, for Pols, there are three non-mutually exclusive hypothesis: i) it may have a catalytic role in the chemical reaction, stabilizing the highly negative charge formed on the scissile phosphate during the nucleotidyl transfer, and/or (ii) it may favor products formation, inhibiting the backward reaction from the products to the reagents, and/or (iii) it may assist the leaving group PPi departure.

Species	Pols	Pol Family	PDB ID
<i>s. cerevisiae</i>	DNA Pol δ	B family	3IAY
archaeal	KOD DNA Pol		5OMF
archaeal	9°N DNA Pol		5OMQ
engineered	KOD-RI TNA Pol		5VU8
human	DNA Pol β	X family	4UAY, 4UB3, 4UB5, 4UBB, 4UBC, 4RPY, 4RPZ, 4RQ0, 4RQ2, 4RQ4, 4RQ5, 4RQ6, 4RQ8, 4KLG, 4KLH, 4KLI, 4KLJ, 4KLO, 4KLQ, 3RH4, 3RH5, 3RH6, 3JPP
	DNA Pol μ		4MOA, 5TYX, 5TYW, 5TYV, 5TYU, 5VZ9, 5VZC, 5VZF, 5VZI
human	DNA Pol η	Y family	4ECT, 4ECU, 4ECV, 4ECW, 4ECX, 5KFH, 5KFI, 5KFJ, 5KFK, 5KFL, 5KFN, 5KFP, 5KFW, 5KFX, 5KFZ, 5KG0, 5KG1, 5KG2, 5KG3, 5KG4, 5KG5, 5KG6, 5KG7, 5L9X

PDB ID 3IAY

PDB ID 4ECV

Figure I.4 (Left) Table reporting the crystal structures of ternary Pols/DNA/dNTP complexes in which three divalent metal ions are found in their active sites. (Right) Representations of two different architectures formed by the three metal ion, i.e. M_A , M_B and M_C (in sphere, colored in orange). dNTP and acidic residues (Asp/Glu) forming the active site (in blue) are in licorice.

A low metal ion concentration is seen in experiments to hamper catalysis, indeed when this condition are used the products are not experimentally detected.⁶¹ Due to this evidence Yang and collaborators suggested that M_C is essential for nucleotidyl transfer.⁵³ However, different computational studies of DNA Pol η have shown diverging results on the M_C effect on the chemical barrier. At the same time, the role in facilitating leaving group

departure played by the third ion remains possible. This hypothesis was supported by quantum mechanics/molecular mechanics (QM/MM) simulations from Wilson and co-workers for Pol β and by De Vivo and co-workers for Pol η using either force-field based molecular dynamics (MD) simulations⁶² and Car-Parrinello (CP) QM/MM.⁶³ Likewise, the stabilization of reaction products was demonstrated by Yoon and Warshel using empirical valence-bond (EVB) calculations on Pol- η structures,⁶⁴ and subsequently it was further confirmed via QM/MM simulations by Stevens and Hammes-Schiffer.⁶⁵

Taken together the available structural, kinetics, and computational evidences show the challenge to establish whether the third ion holds one or more of these roles during catalysis.

1.4. Strategically Located Basic Residues

Polymerases elongate nucleic acids primer strand incorporating the new nucleotide in the Watson-Crick (W-C) base pairing with the template strand. Structural and bioinformatics studies suggest a functional role during nucleotide incorporation for strategically located Arg/Lys residue. Such positively-charged residue is located in the second-shell of two-metal-ion active site and is conserved in a wide variety of Pols.⁶⁶ Notably, crystal structures of Pol/DNA/(d)NTP ternary complexes showed the nascent base-pair in an uncommon Hoogsteen (HG) base pairing conformation when Arg/Lys was found displaced, missing or mutated. Additionally, such residue is missing in some error-prone Pols, such as DNA Pol X. Through the use of MD simulations coupled with free-energy calculations, Arg/Lys residue is then suggested to establish specific interactions with the incoming nucleotide thus promoting and assisting the correct W-C base pairing during Pols catalysis. These results define a larger two-metal-ion-centered enzymatic structure which potentially reflects a common strategy of Pols in DNA and RNA processing.

1.5. Polymerase Applications: Therapeutics, Biotechnology and Synthetic Biology

As reported in Section 1, Pols have gained attention as pharmacological target for small-molecules inhibition. Indeed, many Pols are involved in the onset of several diseases, such as cancer, neurodegenerative diseases and viral and bacterial infections. Thanks to the increasing capabilities of computational techniques, such as molecular docking, de novo

design and free-energy perturbation (FEP), it has been possible to apply such approaches to give a relevant contribution in discovering inhibitors against Pols. This allowed reducing the need for more expensive experimental studies. A representative example of Pol drug target is the HIV reverse transcriptase (RT). HIV RT is able to reverse transcribes (RNA → DNA) the viral RNA genome into double-strand DNA having three sequential biochemical activities: i) RNA-dependent DNA polymerase, ii) ribonuclease H (RNase H), and iii) DNA-dependent DNA polymerase activities.¹⁹ Thanks to computational techniques, it has been possible to discover non-nucleoside inhibitors with pico- and low-nanomolar activities against wild-type and mutant HIV RT.⁶⁷ Other interesting drug targets are DNA Pols involved in translesion synthesis, such as Y-family members, because of their capability to bypass the nucleotide adducts formed by anticancer agents, thus reducing the efficacy of chemotherapy medications.

Currently, RNA-dependent RNA Pol from SARS-CoV-2, which caused COVID-19 has become a valid drug target given the crucial role during viral replication and transcription processes. Given the urgent and unmet medical need, computational and experimental studies have taken advantages for drug repurposing to suggest several inhibitors for RdRp SARS-CoV-2 target. For example, Remdesivir, Favipiravir, Galidesivir Ribavirin are all nucleotide/nucleoside analog proposed as SARS-CoV-2 treatment. However, despite the promising inhibitory effects,⁶⁸ clinical trials showed no statistically significant benefits on hospitalized patients together with the possible occurrence of adverse effects in relation with the administration of such drugs.⁶⁹ In the last two years, several computational studies have contributed to understand the mechanism of inhibition of such drugs, in order to aid the discovery and development of new and more suitable inhibitors.^{70–72}

DNA polymerases also play central roles in modern molecular biology and biotechnology, such as DNA cloning, the polymerase chain reaction (PCR), DNA sequencing, single nucleotide polymorphism (SNP) detection, whole genome amplification (WGA), synthetic biology, and molecular diagnostics.⁷³ One of the earliest and more important DNA polymerase-based biotechnology applications is PCR, which employ Taq DNA Pol I. However, some limitations come from the DNA Pol I inability of replicate and amplify damaged DNA.⁷⁴ Indeed, this is a major limitation in forensic science or the analysis of ancient DNAs. Thermostable Y-family Pols would offer an attractive

solution, nevertheless the low processivity hamper their usage for PCR. Several studies have shown how to circumvent such drawback through DNA Pol mutations and/or fusion with a nonspecific DNA-binding protein, in order to strengthen the DNA binding.^{75–77} Here, computational techniques, such as bioinformatics analysis and molecular mechanics Poisson–Boltzmann surface area (MM- PBSA) method have been usefully applied to rationally design DNA Pol variants with the desired properties.

Another recent fast-growing field is the artificial Pols design. For example, the development of new unnatural nucleotides expanding the genetic alphabet, enables the storage of additional information in DNA.⁷⁸ Through the modifications of sugar and nucleobases, it is possible to build modified D(R)NA that can be used for biological and biotechnological applications.⁷⁷ Some of these applications are the synthesis of new aptamers able to bind specifically to a target, the incorporation of unnatural amino acids into proteins, and the generation of semisynthetic organisms. In this context, the development of synthetic xeno nucleic acids (XNA) is an active area. XNA polymers can store information and be immune to endogenous nuclease activity, furthermore they may perform catalysis (xenzymes).⁷⁶ On the other hand, a challenging area is the development of DNA Pols able to process unnatural base pairs (UBP). Here, computational studies are applied along with structural data to give a better understanding of the molecular mechanism of UBP processing Pols. For example, MD simulations have been employed to give deeper insight on the molecular mechanism of a variant of the Klenow fragment of Taq DNA Pol I, which incorporates efficiently P:Z (i.e. 2-amino-imidazo[1,2-a]-1,3,5-triazin-4(8H)one and 6-amino-5-nitro-2(1H)-pyridone, respectively). The results suggested that mutations far from the active site allow better interaction of the enzyme with the P:Z-containing DNA.

Taken together, all these results support the relevance of Pols for several aspect of scientific advances, such as target therapy, biotechnology and synthetic biology. Importantly, computational approaches have provided a means to study several key aspects of Pol function.

2. Nucleases

Nucleases are enzymes that perform a wide range of activity essential for life.⁷⁹ They play crucial roles for genome replication, acting as 3'-5' nuclease proofreading as well as 5'-3' RNA primer removal.^{80,81} Nucleases are also involved in recombination and repair processes^{82,83} such as base and nucleotide excision repair (BER and NER, respectively), which involve the removal of damaged base (in BER) or short strands of oligonucleotides (in NER).⁸⁴ Alternatively to DNA repair, cells can respond to internal signals causing the programmed cell death.⁸⁵ This mechanism also implies nuclease activity. Other nuclease activities are involved in modifications and reorganization of nucleic acids topology, like topoisomerases,⁸⁶ which participate in the winding of DNA and site-specific recombinases acting during the rearrangements of DNA segments.⁸⁷ Additionally, RNA splicing, processing and maturation required nuclease activities.⁸⁸⁻⁹¹

Because of the central role of nucleases in such a variety of functional processes for genome maintenance and cell life cycle, these enzymes have been implicated in various human diseases. For example, mutations on the gene encoding Dna2 (i.e. DNA2) – a nuclease-helicase essential for DNA synthesis and repair – are associated with sensitivity to DNA replication stress, genome instability, mitochondrial myopathy, and the primordial dwarfism disorder Seckel syndrome.^{92,93} Concurrently, overexpression of DNA2 is linked to different cancer cell such as breast and ovarian cancers.^{93,94} Another example is represented by FAN1, a DNA repair nuclease. Genetic deficiencies, copy number variants, and single nucleotide variants of FAN1 (i.e. the gene encoding FANCD2/FANCI-associated nuclease 1) is associated to chronic kidney diseases, neurological conditions such as schizophrenia, cancer, and repeat expansion diseases including Huntington's disease, fragile X and autism syndromes.⁹⁵ Additionally, mutations on flap endonuclease 1 (Fen1), lead to autoimmune and chronic inflammation diseases together with different type of cancers such as breast, lung, skin, kidney, colorectal, ovarian and testicular tumors.⁹⁶⁻⁹⁹ Fen1 related diseases include also Werner syndrome and Vitelliform macular dystrophy.^{100,101} Therefore, inhibition and/or modulation of nucleases have become an attractive therapeutic strategy.^{102,103} Additionally, in the era of drug resistance, a new challenge is to expand the therapeutic targets in order to expand mechanistic drug diversity. An emerging approach is to target ribonucleases that mediate RNA mechanisms.¹⁰⁴

Notably, recent advances in the development of programmable site-specific nucleases have revolutionized the idea of genome editing.¹⁰⁵ In this context, clustered regularly interspaced short palindromic repeats (CRISPR)-associated (Cas) 9 is the most widely used protein for genome editing. CRISPR-Cas9 enables efficient, site-specific genome engineering in cells and whole organisms and together with other CRIPR systems are already employed to alleviate genetic disorders in animals and are now entering the realm of human diseases treatment.¹⁰⁶

Given the diversity of nuclease architectures and domains, the different activities contributing to specificity, target range, and biased outcomes, a constant and effective understanding of nuclease molecular mechanisms and functional properties is required. In this regard, computational methods are a powerful resource to further advance on the nuclease research, giving deep insight at atomistic level on the catalytic mechanism, thus leading to benefits for human health and progress.

2.1. Diversity and Classification of Nucleases

Nucleases are enzymes able to perform the cleavage of DNA and RNA. They have a wide variety of activities, catalytic architectures and include both protein and catalytic RNA, i.e. ribozymes. Thus, nucleases can be classified by different properties and functions (Figure I.5).

One way to classifying them is based on processed substrate (Figure I.5): i) DNase recognize and cleave DNA substrates (e.g. Dna2),¹⁰⁷ ii) RNase are specific for RNA substrates (e.g. RNaseH)¹⁰⁸ and iii) sugar non-specific nucleases perform the cleavage of both DNA and RNA (e.g. EndoG).^{109,110} Another classification method is based on the recognition properties (Figure I.5), thus: i) sequence-specific nucleases target to a specific D(R)NA sequence, e.g. CRISPR-Cas,¹¹¹ zinc finger nucleases (ZFNs)¹¹² and transcription activator-like effector nucleases (TALENs)¹¹³, ii) structure-specific nucleases recognize specific nucleic acids secondary structures, e.g. FENs family members, xeroderma pigmentosum group F-complementing protein (XPF) family, meiotic recombination protein 11 (MRE11) family, synthetic lethal of unknown function 1 (SLX1)¹¹⁴. In contrast to Pols, nucleases can cleave the substrates with a different polarity (Figure I.5). For this, they can be divided in: i) 3'-5' nucleases, which require 3' end for substrate recognition

and ii) 5'-3' nucleases, which require a 5'-end for substrate recognition. Additionally, nucleases can further be divided based on the cleavage products (Figure I.5).

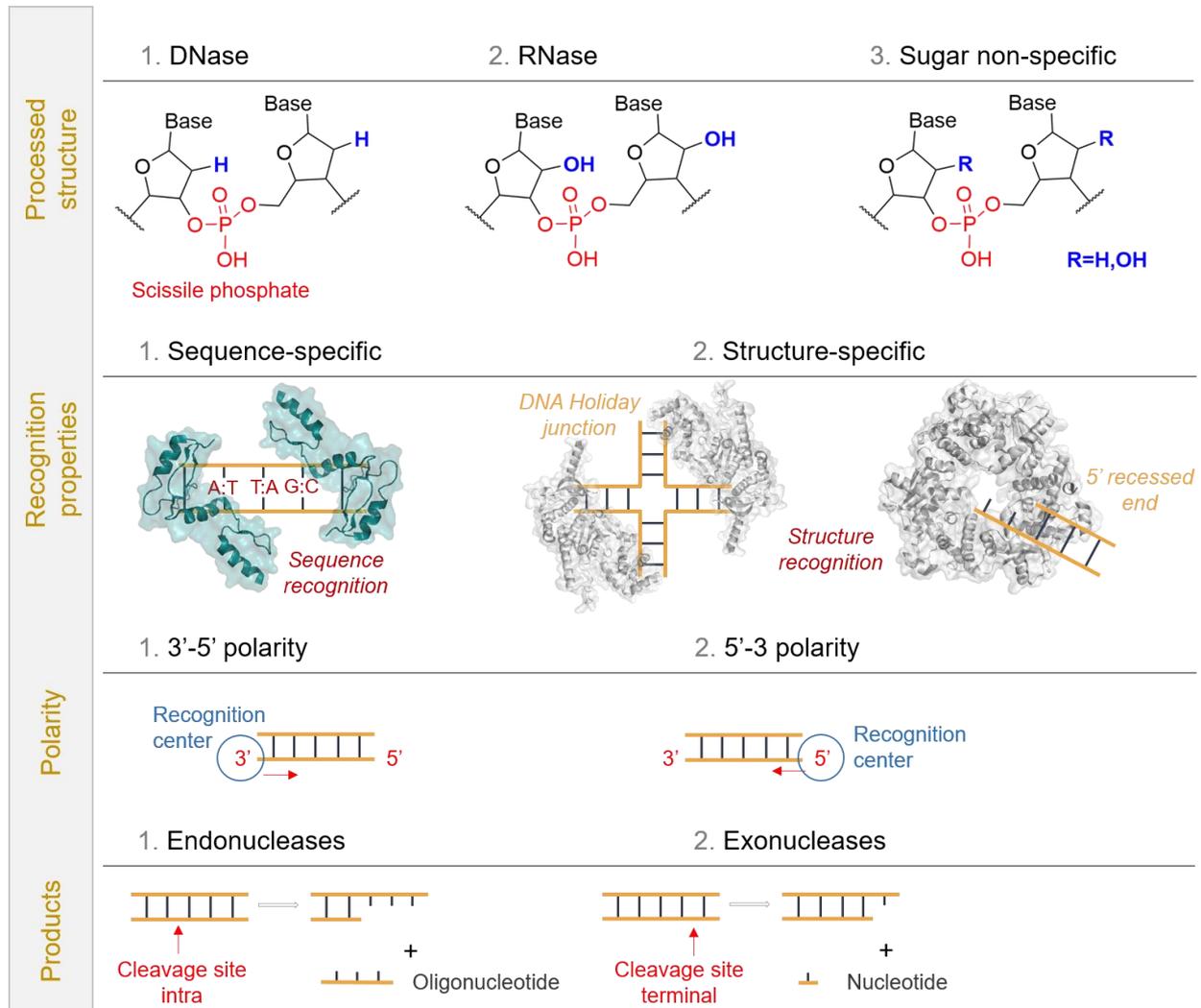


Figure I.5 Schematic representation of classification methods for nucleases based on: i) the chemical structure of the processed substrate, namely DNA, RNA or sugar non-specific; ii) the properties of recognition method, such as sequence-specific or structure-specific; iii) the polarity, depending on whether the enzyme recognizes the 3' end or the 5' end; iv) the type of products formed after the reaction took place, thus endonuclease reaction hydrolyzes the phosphodiester bond producing an oligonucleotide fragment, while exonuclease reaction hydrolyzes the 3' or 5' terminal nucleotide.

Indeed, exonucleases cleave one nucleotide at a time from the end strand (such as Exo1 and Exoλ), while endonucleases are able to incise inside the strand (such as Fen1 and group II introns), thus leading to an oligonucleotide product. Eventually, according to the catalytic mechanism, there are three major classes: i) two-metal-ion mechanism (e.g. type IA, type II topoisomerases and group I, group II ribozymes), ii) one-metal-ion mechanism (e.g. NucA and EndoG), and iii) metal-independent mechanism (e.g. Cas6 and RNaseA). Due

to such great diversity, nucleases cannot be sufficiently identified using any single criterion. To complicate even more the rationalization, similar biological functions are often achieved by enzymes with unrelated tertiary structures and drastically different mechanisms, thus making difficult to predict molecular mechanism and properties of unknown nucleases.⁷⁹ Conversely, similar tertiary structure nucleases may employ different catalytic mechanisms.⁷⁹ Overall, all these evidences show the challenging research area represented by nucleases, in which a wide range of different architectures and catalytic mechanisms are engaged for the same catalytic reaction.

2.2. Reaction Mechanism and Catalysis

Nucleases cleave phosphodiester bonds in nucleic acid polymers on the 5' or 3' side (i.e. O5'-P or P-O3' bond cleaved, respectively). The chemical reaction involves the deprotonation by a general base of a nucleophile, which in this way is activated for direct hydrolysis. After the nucleophilic attack on the phosphorus, a penta-covalent bipyramid intermediate is formed followed by the cleavage of the scissile bond and the inversion of the phosphorus stereo configuration (i.e. umbrella inversion). This occurs via in-line S_N2-like reaction and it can take place either assisted by one or multiple metal ions¹¹⁵ (Figure I.6) or in a metal-independent manner^{116,117} (this is discussed in more details in the next paragraph). Subsequently, a general acid facilitates product formation by protonating the leaving group.⁷⁹ Depending on the side of nucleophile for the in-line attack, two different phosphodiester bond are cleaved. For example, if the nucleophile is positioned on the 5' side, P-O3' bond is cut thus generating 5'-phosphate and 3'OH products (Figure I.6), while the 3'-phosphate and 5'OH products are formed when the attack of the nucleophile occurs from the 3' side and the O5'-P bond is cleaved. However, given that 5'-end phosphate is a ready substrate for DNA ligation and the 3'OH can be further used as nucleophile by many enzymes (e.g. DNA/RNA polymerases, spliceosomes and DNA ligases), those are the products most often generated. Differently to polymerases, nucleases can employ different nucleophiles to cleave the scissile phosphate bond. A water molecule turns out to be the most common nucleophilic species, however other examples of nucleophiles can be: i) hydroxyl group at the 3'end of RNA or DNA used during RNA splicing, DNA strand transfer or hairpin formation^{118,119} ii) aminoacids side chains such as tyrosine, serine and histidine may form a covalent DNA phosphoryl-protein intermediate during DNA

recombination and topoisomerization,^{87,120} iii) 2'OH group or free ribonucleotides can be employed by RNases, forming a labile 2',3'-cyclic phosphate,^{90,121} iv) inorganic phosphate used by polynucleotide phosphorylases (PNPases) and RNase PH to degrade ss-RNA.¹²²

Two-metal-ion mechanism : Hydrolysis

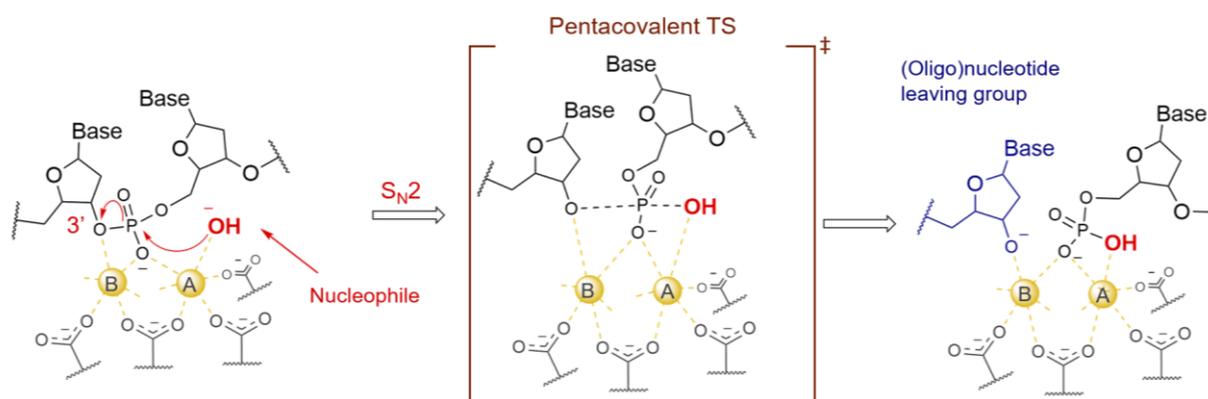


Figure I.6 Catalytic reaction scheme for the two-metal-ion mechanism adopted by nucleases during nucleic acids hydrolysis. By analogy with two-metal-ion mechanism for Pols, the chemical reaction is a S_N2 -like reaction catalyzed by two divalent metal ions, i.e. A and B (in sphere, colored in yellow). Both metals actively participate in catalysis by i) facilitating the nucleophile, i.e. 3'-OH (in red), ii) stabilizing the pentacovalent transition state, i.e. TS, and iii) assisting the (oligo)nucleotide leaving group departure.

2.3. Expanded two-metal-ion architecture

The two-metal-ion catalytic reaction is employed by several enzymes including nucleases and polymerases. Independent of their structural and functional differences, these enzymatic machineries display a remarkable degree of structural similarity. For example, a first-shell structural architecture of strictly conserved acidic groups chelating the catalytic ions is crucial for efficient DNA and RNA processing.⁴⁵ However, recent structural investigations have identified in different enzyme classes two positively charged elements with a conserved spatial locality in the second coordination shell of the catalytic ions. In this respect, crystallographic structures of self-splicing group II intron ribozyme have revealed two catalytic potassium ions (namely, K1 and K2) located close to the active site.^{123,124} Subsequent analysis of six classes of ribozymes and protein enzymes identified two second-shell structural elements (recognized as K1- and K2-like), which i) occupy similar positions, ii) contribute to modulate both the electrostatics of the catalytic site and substrate stability, iii) are evolutionarily conserved, and iv) their mutation is linked to functional defects. Overall, these structural evidences outline a larger two-metal-ion

architecture. This in turn may reveal a common strategy of different two-metal-ion enzymes, such as group II introns, nucleases, and DNA/RNA polymerases to ensure fidelity, substrate specificity, and catalytic efficiency for nucleic acids processing.¹²⁵ Additionally, these outcomes stimulate the research of finding additional elements to reveal a more complex two-metal-ion architecture.

2.4. Role of Metal Ions

Metallonucleases are enzymes which catalyze the D(R)NA phosphodiester bond hydrolysis availing of metal-ion-dependent catalysis.^{126,127} The most common and abundant divalent metal ions are Mg^{2+} and Ca^{2+} ,¹²⁸ nevertheless Zn^{2+} , Fe^{2+} and Mn^{2+} can also be found essential and widespread. The main advantage of having such ions within the active site is the high density of positive charge, which balances the highly negative charges of phosphodiester along the D(R)NA backbone. Additionally, metal ions have specific coordination shell that varies in geometry and stiffness depending on the ion properties. The most common coordination geometry is octahedral,¹²⁹ as it is for Mg^{2+} and Fe^{2+} , which have six ligands in their first coordination shell. Ca^{2+} and Zn^{2+} , on the other hand can have different geometries in addition to octahedral, such as tetrahedral or even more complex with seven to nine ligands.^{130,131} The ligands coordinating metal ions can be either water molecules or protein residues. These latter usually constitute the conserved active site of metalloenzymes, such as aspartate, glutamate, histidine or serine residues.

Nucleases can be grouped into three main classes based on the role of the metal ions during catalysis: i) two-metal-ion, ii) one-metal-ion, and iii) metal-independent nucleases. There is also a further group of nucleases where a third metal ion (or even more than one) is suggested to be involved in the catalysis,^{127,132,133} in a similar way found for Pols.^{2,3} This additional case is discussed separately.

2.5. Metal-Dependent and Independent Catalysis

X-ray crystallographic structures of nucleases often captured the presence of two metal ions in the active site, thus revealing a recurrent two-metal-ion mechanism for metal-aided phosphodiester bond hydrolysis (Figure I.6).¹³⁴ Such mechanism was first proposed for 3'-5' DnaQ-like exonuclease,¹³⁵ and over years further extended to several other enzymes such as nucleases, polymerases, topoisomerases and even ribozymes. It consists of two

divalent metal ions (i.e. M_A and M_B , usually Mg^{2+}) $\sim 4 \text{ \AA}$ apart bound in the enzyme active site. Here, conserved residues (commonly Asp and Glu) coordinate the two catalytic ions together with water molecules and the scissile phosphate, which is placed on top of the two-metal-ion center. M_A is generally coordinated by the nucleophile, thus facilitating the nucleophile deprotonation while M_B stabilizes the leaving group during product formation. Additionally, the high negative charge formed in the pentacovalent transition state (TS) is balanced by the positive charge of metal ions, which are likely to move closer than 3.5 \AA to stabilize the TS complex.¹³⁶ The majority of nucleases employ the two-metal-ion catalysis.

An alternative mechanism adopted by metallonucleases is the one-metal-ion catalysis, where M_A is absent while M_B is conserved.¹¹⁵ The latter is supposed to destabilize the scissile bond and thus facilitate the nucleophilic attack by involving two oxygen atoms from the scissile phosphate with a non-optimal coordination angle (O- M_B -O). The most common nucleases adopting such mechanism are $\beta\beta\alpha$ -Me superfamily and HUH nucleases. The active site arrangement is quite different from the two-metal-ion architecture and in both classes the nucleases involve at least a conserved His residue to coordinate the metal ion. Additionally, in HUH enzymes a second histidine is required for metal-ion coordination. Notably, both metal-ion selection and substrate specificity are less stringent than two-metal-ion catalysis. Indeed, several $\beta\beta\alpha$ -Me superfamily members can equally process both DNA and RNA substrates.¹³⁷

DNA and RNA metal-independent nucleases are also available in nature. In contrast with metal-dependent mechanism where commonly the nucleophile is a water molecule, metal-independent catalysis appear to use protein side chains or 2'-OH as nucleophilic species. Specifically, for DNase a phospho-enzyme covalent intermediate is formed via Tyr, Ser or His side chain.^{87,138,139} For RNase instead, the nucleophilic attack of 2'-OH generates 2',3' cyclic phosphate intermediate. This requires for catalysis to proceed a local backbone distortion together with the base unstacking and unpairing.¹⁴⁰⁻¹⁴³ However, despite sharing metal-independent mechanism, the active site and tertiary structure of such nucleases are quite different.

2.6. A Third Metal Ion for Catalysis

Outstanding exceptions among nucleases concern the presence of a third divalent metal ion within the active site. Structural and biochemical data suggest that such additional ion have a role during catalysis. For example, time-resolved crystal structures captured three metal ions in different intermediates of homing endonucleases I-DmoI during DNA double-strand break (DSB) reaction (Figure I.7).⁵⁸ During the catalysis, specific hydrolysis of two phosphodiester bonds within both strands takes place. Here, both structural and computational studies showed that a third transient metal ion is pivotal for the reaction to be accomplished. Indeed, it provides the proper geometry and chemical environment for catalysis, additionally it triggers the sequential cleavage of both DNA strands.⁵⁸ This proposed mechanism seems to be shared with other homing endonucleases, such as I-SceI where three metal ions were captured by crystallographic structures.¹⁴⁴

Three Zn^{2+} ions have been found in the active site of two unrelated nucleases: i) AP endonuclease IV (Endo IV) that cleaves dsDNA, and ii) nuclease P1/S1 that cleaves ssDNA.¹⁴⁵ High resolution crystallographic structures (Figure I.7) of Endo IV highlighted the resemblance of two over three Zn^{2+} ions to M_A and M_B while the third Zn^{2+} locates in a position similar to a histidine residue present in other AP endonucleases, such as APE1.⁷⁹ Here, the third ion is supposed to stabilize a twisted conformation of the scissile phosphate in the active site. Structural data analysis together with *ab initio* QM/MM MD further suggested that such additional ion may be responsible for stabilizing the developing charge on the leaving group toward the end of the reaction.¹³² Although P1 nuclease is unrelated in amino acid sequence and tertiary structure to Endo IV, both shared a similar active site in which Asp and His residues coordinate the three metal ions. Experimental results on both P1 and Endo IV, suggested a similar role for the third zinc ion during catalysis.¹⁴⁶

Other nucleases with three metal ions in the active site also exist. For example, kinetic data further supported by crystallographic structures suggest a possible involvement of three magnesium ions during T5FEN catalysis (Figure I.7).¹⁴⁷ Belonging to the same FEN superfamily, human Exonuclease 1 (hExo1) has been recently co-crystallized in complex with DNA substrate and metal ions. In these recently published time-resolved crystal structures, the presence of a third divalent metal ion in the close proximity of the active site, is recurrent in all the pre-reactive structures (four in total, PDB ID 5V06-9).¹⁴⁸ In the

Results chapter, it is presented a computational investigation on the recruitment mechanism and functional role of such ion. Furthermore, three metal ions mechanism is also suggested for EcoRV restriction endonuclease, where the additional metal ion may play a structural role.¹⁴⁹ Eventually, a third metal ion involved during catalysis in ribonuclease H (RNase H) was first proposed starting from MD simulations and recently supported by several crystallographic data (Figure I.7), showing a complex monovalent- and divalent-cation trafficking during the catalytic process.¹⁵⁰ Subsequent MD simulations and free energy calculations gave insight into the role of such ions. Indeed, it is suggested that the correct substrate positioning is carried out by monovalent transient ions, while the third transient Mg^{2+} triggers the subsequent release of the leaving group.¹⁵¹

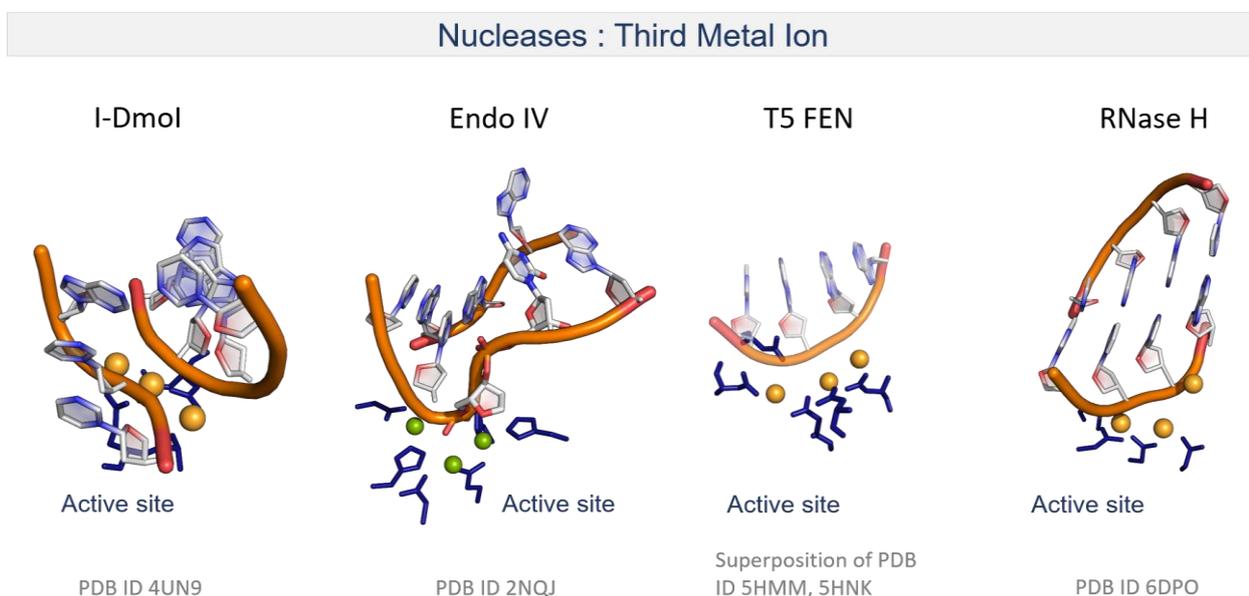


Figure I.7 Close view of the active sites (in licorice, colored in blue), DNA/RNA substrates and the three divalent metal ions (in spheres) of different nucleases. In particular, crystal structures of I-Dmol (PDB ID 4UN9⁵⁸), Endo IV (PDB ID 2NQJ¹⁴⁵), T5 FEN (PDB ID 5HMM¹⁴⁷, 5HNK¹⁴⁷) and RNase H (PDB ID 6DPO¹⁵⁰) are represented. Catalytic Mg^{2+}/Mn^{2+} are colored in orange, while Zn^{2+} ions are in green. Representation of T5 FEN is obtained superimposing the active site and divalent ions from PDB ID 5HMM and the DNA substrate (in transparent cartoon) from PDB ID 5HNK.

All together these evidences suggested that the recurrent presence in crystallographic structures together with bot experimental and computational results indicate an intriguing relevance of such additional ion not only for Pols but also for nucleases catalysis.

2.7. RAD2/FEN superfamily of nucleases

The cleavage of phosphodiester bond of nucleic acids is an essential function in cellular processes, such as genome maintenance, replication and repair, RNA processing, maturation, interference and eventually apoptosis. The origin for the complexity and variety of nucleases is to carry out all these functions and for these they have developed diverse strategies for recognizing and cleaving DNA/RNA based on specific: i) sequence (e.g., type II restriction endonucleases),¹⁵² ii) structure (e.g., FEN superfamily members),¹⁵³ or less commonly iii) length (e.g., endoribonuclease Dicer).¹⁵⁴ Flap endonucleases (FENs) superfamily belongs to the 5' structure-specific nucleases and comprises both endo- and exonucleases. The eukaryotic members (Figure I.8) of the superfamily are Flap endonuclease 1 (FEN1), Exonuclease 1 (Exo1), Gap endonuclease 1 (GEN1) and Xeroderma pigmentosum complementation group G protein (XPG). All these enzymes are able to bind and cleave specific 3-dimensional structures of DNA substrates with a 5'-strand polarity.^{155,156} For example, FEN1 primary substrate is 5'-3' double flap DNA, Exo1 recognizes 3' overhang DNA substrates, GEN1 is specific for Holliday junctions formed during recombination, and XPG recognizes bubble substrates generated during nucleotide excision repair of DNA lesions (Figure I.8).¹⁵⁷ Despite the diversity of preferred substrate structures these enzymes share some similarities. They all perform the phosphodiester bond hydrolysis using the two-metal-ion mechanism (Figure I.6), where the scissile phosphate is located in between the first and second nucleotides within dsDNA region. Additionally, FEN superfamily members share a similar enzyme architecture with common domains (Figure I.8), i.e. i) the nuclease domain where the two-metal-ion catalytic site is located, ii) the mobile arch, which overhangs the nuclease domain and it is formed by the gateway and the cap regions, iii) the wedge domain, iv) the H2TH, which has a monovalent ion binding site and it is partially responsible for dsDNA binding, and v) the β -pin domain.¹⁵⁷ Misplaced phosphodiester bond hydrolysis can compromise genetic information, function and other sensitive processes. For this reason, structure-based nucleolytic activity is often strictly regulated. Regulation is achieved through the cooperation with other proteins in multistep transactions, which influence the substrate selection and the moment to cleave it.¹⁵⁶ How the first approach of enzyme-cognate

substrate happens, how specificities for substrate recognition are defined, and how this is regulated are central open points that require deeper understanding.

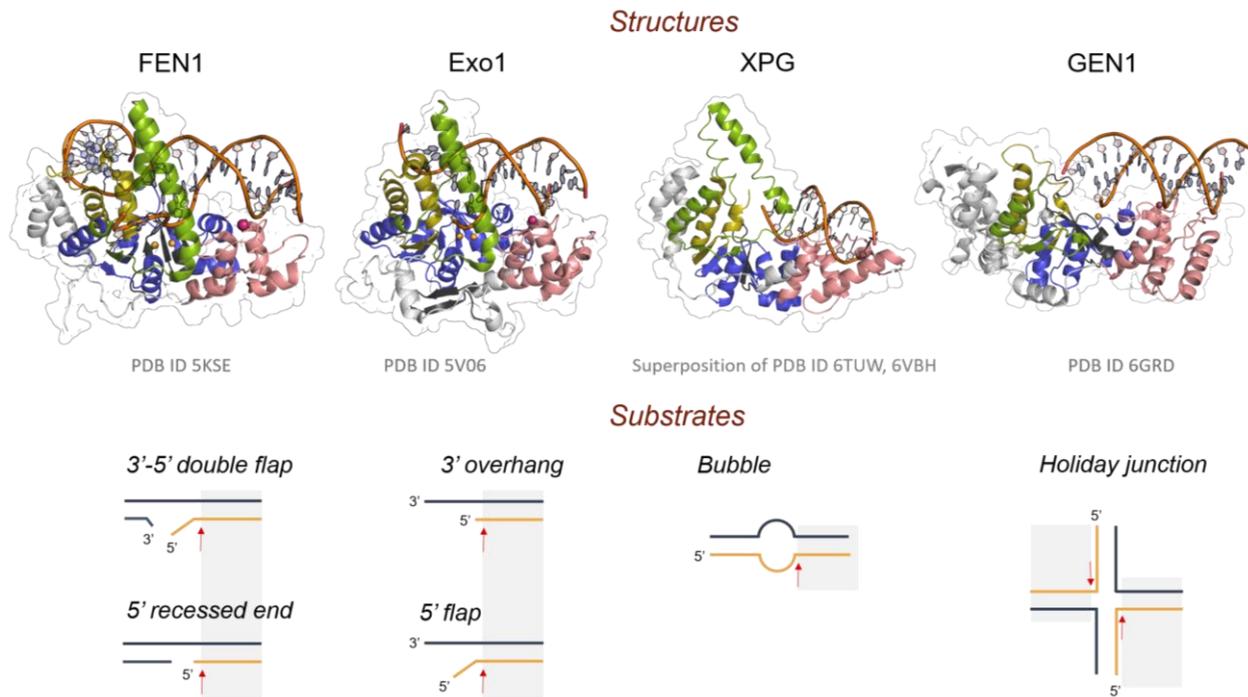


Figure I.8 On top, crystallographic structures of eukaryotic FEN superfamily members, i.e. FEN1 (PDB ID 5KSE¹⁵⁸), Exo1 (PDB ID 5V06¹⁴⁸), XPG (PDB ID 6TUV¹⁵⁹, 6VBH¹⁶⁰) and GEN1 (PDB ID 6GRD¹⁶¹) in complex with their DNA substrate. The enzymes are colored by the common domains: i) the nuclease domain (in blue) where the two-metal-ion catalytic site is located; ii) the mobile arch (in green), which overhangs the nuclease domain; iii) the wedge domain (in yellow); iv) the H2TH (in pink), which has a monovalent ion (in sphere, colored in magenta) binding site and it is partially responsible for dsDNA binding. For XPG, the image is obtained by superimposing the enzyme structure from PDB 6TUV and the DNA substrate from PDB 6VBH. On bottom, the schematic representations of the main substrates recognized and hydrolyzed by the corresponding enzyme. Cleavage site is indicated by the red arrow.

Here, structural data together with experimental and computational studies contributed to give useful insight on the molecular basis for human FENs catalysis. In particular, hFEN1 and its paralogue hExo1 have been recently characterized to shed light on key structural features by which the similar active site can specifically accommodate the appropriate substrate.¹⁶²⁻¹⁶⁴ Notably, while hFEN1 has a primarily endonucleolytic activity and cleaves 5'-flap substrates, hExo1 preferentially cleaves one nucleotide at a time in 5'-recessed end substrates. However both enzymes are able to perform both reactions, although with different catalytic rates.¹⁶⁵ Biochemical and structural studies have shown how the mobile arch in hExo1 and hFEN1 plays an active role in the exclusion of incorrect substrates. Firstly, the dimensions of the mobile arch is wide enough to allow through only

a single strand of DNA, thus discriminating a 5'-flap single strand from a 5' double strand.^{162,163,166} This evidence rules out a process, the threading mechanism, where both enzymes thread 5'-flap DNA through a structured arch. Intriguingly, several crystallographic structures of hFEN1 have additionally underscored a disorder-to-order mechanism that precedes the DNA threading. Such mechanism is characterized by a significant flexibility of the mobile arch, which gradually goes from disorder to order conformation.¹⁵³ Importantly, this mechanism is not observed in hExo1. Additionally, recent *in crystallo* reaction intermediates together with mutagenesis data have contributed to suggest a further strategy to enhance the specificity and catalysis in FENs members.¹⁵⁸ The proposed mechanism, the so-called phosphate steering mechanism, consists in an electrostatically driven steering of the 5'-flap phosphates, to properly position the scissile bond at the two-metal-ion site. Four second- and third-shell positively-charged residues (i.e. Arg and Lys) are proposed to take part and promote such mechanism. Notably, these residues are (semi)conserved in the FEN superfamily, thus suggesting a possible extension of the phosphate steering mechanism over the family members.¹⁵⁸

Despite the extended characterization different questions remain to be clarified. For example, which is the molecular origin for different catalytic activity preferred by hFEN1 and hExo1. Again, how the recurrent presence in both Pols and nucleases of strategically located positively charged elements governs substrate positioning and the catalytic process.

2.8. Nuclease Applications: Genome Editing and Therapeutic Targets

Among all nucleases, in the last decade CRISPR-Cas systems have drawn attention as genome editing technology for the advantages of low cost, high efficiency and simple design.¹⁶⁷ However, zinc finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs) are also mainstream tools for such application. Genome editing is referred to the modification of genomic DNA at a specific target site. Over the years, such technology has been found to be an effective method applicable in various fields, ranging from basic research to applied biotechnology and biomedical research.¹⁶⁸ It is a useful tool to study the pathogenesis of hereditary diseases, gene function, and to develop novel targets for gene therapy. An interesting and innovative example, concerns designer nucleases. In contrast to small-molecules drug approach, designer nucleases allows a permanent inactivation of disease-causing targets thus leading to the inhibition of virus-associated

cancers.¹⁶⁹ The binding site of designer nucleases recognizes a specific DNA site and in these the active site cleaves through a double stranded break (DSB) mechanism. This induces the activation of non-homologous DNA end joining (NHEJ) pathway, which is highly error-prone, thus causing deleterious mutations or gene inactivation. Examples in which designer nucleases have been applied to inactivate integrated and viral DNA genomes are the cases of human immunodeficiency virus (HIV)¹⁷⁰ and hepatitis B virus (HBV),¹⁷¹ respectively. Although genome editing and similar techniques are broadly applied nowadays, particular attention and further studies are required in order to overcome some possible side effects such as off-target effects, delivery methods, immunogenicity and potential risk of cancer.^{172–175}

In addition to genome editing, many nucleases are associated to different diseases ranging from genetic, infectious diseases and cancers. For this they are potential therapeutic targets. For example, Dna2 nuclease is overexpressed in pancreatic cancers, which is among the more aggressive forms of human cancers.¹⁷⁶ Such nuclease is an essential enzyme in replication and homologous recombination (HR) repair and it is often employed by cancer cells to overcome replication stress and survive. Moreover, depletion of Dna2 significantly reduces pancreatic cancer cell survival, thus suggesting the therapeutic potential of its inhibition.¹⁷⁶ Mutations in other nucleases participating in HR are associated with diseases in a direct or indirect way (i.e. they lead to an accumulation of damage or to a predisposition to cancer).¹⁷⁷ For example MUS81, a member of the XPF family, is suggested to have a role in tumorigenesis and MUS81-deficient cells show chromosomal aberrations. Mutations in XPF and SLX4 proteins have been recognized, respectively: i) in patients with Xeroderma pigmentosum (XP) and a progeroid syndrome (XFE), and ii) in patients with Fanconi anemia complementation group P and hereditary breast cancer. Moreover, FEN1 and Exo1 nucleases have an essential role in genomic stability and cancer predisposition. Indeed, clinically relevant FEN1 mutations are clearly linked to compromised function, genomic instability and cancer,^{96,178,179} while FEN1-deficient mice show an increased susceptibility to cancers.¹⁷⁸ While Exo1 has been suggested to contribute to non-polyposis colorectal cancer (HNPCC) and sporadic colorectal cancers (CRC).^{177,180}

3. Scope of the Thesis

Polymerases and nucleases are metalloenzymes that process nucleic acids using metal ions to catalyze the extension and cleavage of nucleic acids, respectively. To do so, they mostly operate via the well-known two-metal-ion mechanism. Given that, despite exhibiting relevant differences in structure and catalytic properties, they share common catalytic elements that comprise the two catalytic ions and their coordinating acidic residues (i.e. the first coordination shell). Nevertheless, recent studies revealed the recurrent presence nearby the active site of different metalloenzymes of additional structural elements that interact with the substrates. This in turn suggest an extended two-metal-ion-centered architecture. In particular, recent *in crystallo* reaction intermediates of hExo1 have captured a third divalent metal ion intermittently bound close to the two-metal-ion active site. Evidences of this third ion has been observed in several polymerases and nucleases. Similarly, crystallographic structures together with mutagenesis data of hFEN1, suggest that four Arg/Lys residues operate the phosphate steering mechanism to favor specificity and catalysis. Interestingly such positively charged residues are located in the second- and third-shell of the two-metal-ion active site and moreover they are conserved among FENs superfamily members.

The aim of this thesis is to investigate whether these elements have a catalytic function and what is their role for nucleic acids processing using molecular dynamics simulations and enhanced sampling techniques. First, 0 discusses the functional role of the transient third metal ion during hExo1 catalysis. Additionally, the conformational switching of a structurally conserved acidic residue is investigated to define a Glu(Asp)-mediated mechanism for third ion recruitment and nucleic acid hydrolysis. Second, Chapter IV investigates the role of four basic residues in promoting catalysis and preventing off-target incision in hFEN1. Moreover, it is discussed the molecular mechanism for structure-based selection of specific DNA substrates controlled by such conserved Arg/Lys residues nearby the active site.

Chapter II. Computational Theory and Principles

1. Molecular Dynamics Simulations

Molecular Dynamics (MD) simulations are computational techniques employed to studying a microscopic unit (i.e. the simulation box) of macroscopic systems (e.g. enzymes). These tools are frequently applied in order to provide a wide variety of biomolecular processes, such as ligand binding/unbinding, conformational changes of biological macromolecules, as well as kinetic and thermodynamic properties. MD simulations predict time-dependent behavior of complex molecular systems revealing the positions of all the atoms at femtosecond temporal resolution. To do so, MD simulations exploit the Newton's equation for systems of N interacting atoms (i) [1].¹⁸¹

$$m_i \frac{d^2 r_i}{dt^2} = F_i, \quad i = 1, 2, \dots, N \quad [1]$$

Where, F_i is the force acting on atom i with mass m_i and position r_i . Given that the force can be expressed as the gradient of the potential energy (dV), the formula [1] can be written as:

$$m_i \frac{d^2 r_i}{dt^2} = - \frac{dV}{dr_i} \quad [2]$$

To reproduce the behavior of the atoms, the Newton's law of motion has to be simultaneously integrated for all the atoms i . However, there is not an analytical solution for that, thus numerical algorithms are employed. These are based on time discretization, where the value of Δt is dictated by the fastest degrees of freedom, such as bond vibration. For example, in MD simulations $\Delta t \sim 1-2$ fs. The most commonly used algorithm is the velocity-Verlet algorithm, which uses [3] to calculate the atoms position r_i :

$$r_i(t + \Delta t) = v_i(t)\Delta t + \frac{F_i(t)}{2m_i} \Delta t^2 \quad [3]$$

And the velocities are given by:

$$v_i(t + \Delta t) = v_i(t) + \frac{F_i(t + \Delta t) + F_i(t)}{2m_i} \Delta t \quad [4]$$

At each time step (i.e. Δt) coordinates and velocities are generated and stored in order to collect the time-dependent trajectory of the system. The global flow for MD simulations can be described as:

1. A set of initial coordinates (r_i) and velocities (v_i) for all the atoms is given.
2. The force (F_i) acting on each atom at time (t) is computed.
3. The acceleration for each atom is derived from F_i .
4. The coordinates and velocities at time $t + \Delta t$ are computed.
5. The configuration of the atoms is updated.

Here, the force acting on each on each atom is defined as the negative gradient of an empirical potential function (V), which takes into account the interactions of the atom i with the others. V is composed by two major terms, i.e. the bonded and the non-bonded interactions, and it has the following formula:

$$\begin{aligned}
 V = & \sum_{bonds} k_r (r - r_0)^2 + \sum_{angles} k_\theta (\theta - \theta_0)^2 + \sum_{dihedrals} k_\phi [1 \\
 & + \cos(n\phi + \phi_0)] \\
 & + \sum_i \sum_{j \neq i} 4\epsilon_{i,j} \left[\left(\frac{\sigma_{i,j}}{r_{i,j}} \right)^{12} - \left(\frac{\sigma_{i,j}}{r_{i,j}} \right)^6 \right] + \sum_i \sum_{j \neq i} 4\epsilon_{i,j} \frac{q_i q_j}{\epsilon_0 r_{i,j}}
 \end{aligned} \tag{5}$$

Where, the first three terms are the potential energy for bond stretching, angle bending and proper/improper torsions, respectively. While the last two terms refer to the potential energy for Van-der-Waals interactions, derived from the Lennard-Jones 6-12 potential,¹⁸² and electrostatic interactions, as the Coulomb's law. All five terms of [5] are parametrized to fit both experimental and computational data and the parameters are collected in the so-called 'force-field'. The most commonly used force-field for MD simulations are AMBER,¹⁸³ GROMOS¹⁸⁴ and CHARMM.¹⁸⁵

MD simulations is a type of deterministic simulation, contrary to Monte Carlo simulations, for example, which is a non-deterministic method. It means that given a particular input (e.g. a set of coordinates and velocities), the algorithm will always produce the same output (e.g. the new sets of coordinates and velocities).

In agreement with the ergodic hypothesis, when the MD simulation time approaches infinite the system is assumed to have sampled uniformly all the available phase space, and the ensemble average of a property P is equal to the time average:

$$P_{avg} = \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t dt' P(r, t') \quad [6]$$

Two main limitations come from MD simulations approach. Firstly, the reliability of the results obtained strictly depends on the accuracy of the force-field used. However, nowadays the trustworthiness of force-fields available for most of the biological systems - such as protein, nucleic acid, water molecules and metal ions - is surprisingly improved, thus reproducing reliable results. Secondly, the computational cost of simulating large systems on a relevant time scale. Indeed, some of the most interesting events, like protein conformational changes, ligand binding or protein folding require a timescale usually longer than the simulation time routinely accessible. Such limitation can lead to inadequate sampling of conformational states, which in turn limits the ability to capture functional properties of the simulated systems. Although the constant development of algorithm and hardware push out further such type of limitations, an effective idea to accelerate the thermodynamics calculation is using enhanced sampling techniques.

1.1. Challenges of Molecular Dynamics Simulations for Nucleic Acid Processing Metalloenzymes

Metal ions play an essential role in numerous vital processes.^{186,187} For example, they can screen electrostatic interactions that develop between charged biomolecules like nucleic acids and proteins. Significantly, by binding to the active site, divalent metal ions act as cofactors in metalloenzymes and ribozymes, thus increasing the rate of the chemical reaction to be performed. In this context, magnesium ions, Mg^{2+} , are particularly important for nucleic acid processing metalloenzymes.⁴⁵ Given the biological and chemical relevance of such ions, many efforts have been made in order to develop accurate Mg^{2+} force fields for reproducing their behavior with MD simulations.^{188,189} Currently, the most commonly used force fields are the non-polarizable forms.¹⁹⁰ One of the main advantages is that they are not computationally demanding.¹⁹¹ Moreover, by wisely adjusting the Lennard-Jones (LJ) parameters based on experimental properties, non-polarizable force fields may

implicitly account for polarizability.¹⁹⁰ Regarding this, two main strategies can be applied for constructing divalent metal ion force fields. Such strategies are based on bonded and non-bonded models.¹⁸⁸ The bonded model employs empirical bonding terms to impose the correct coordination structure, thus it do not allow the exchange of ligands in the first coordination shell.^{192,193} Such limitation may impede to capture and describe important events for biochemical systems, such as changes in coordination state or binding mode. Given that, most of the computational studies on biologically interesting systems use the non-bonded model.¹⁹⁴ At present, this approach represent the metal ion as a point charge surrounded by a non-electrostatic pairwise potential that most commonly is the LJ potential.¹⁹⁴ Such potential is defined by a repulsive and an attractive term that vary as r^{-12} and r^{-6} , respectively, in which r represent the internuclear distance (12-6 model).^{188,194} More recently, the further addition of r^{-4} term has improved the LJ potential description.¹⁹⁵ Importantly, while for monovalent ions Joung-Cheatham parameters¹⁹⁶ are now default in Amber force fields, for divalent metal ions is more complex. Indeed, ideal parameters that reproduce several experimental properties at once for divalent metal ions are very difficult to obtain with non-bonded model, due to the underestimation of metal ion-ligand interactions.¹⁹⁵ Currently, different parameters to simulate Mg^{2+} ions in water are available and optimized to reproduce experimental data, such as solvation free energy, activity derivative, and water residence time. Among them, Åqvist¹⁹⁷, Allnér-Villa¹⁹⁸ and Li-Merz¹⁹⁹ parameters are the most widely used to simulate Mg^{2+} ion in combination with TIP3P²⁰⁰ water model. The Åqvist parameters, for example, were obtained calibrating the calculated hydration free energy against the experimental data.¹⁹⁷ However, they fail to reproduce thermodynamic or ion specific effects,^{201,202} thus causing artifacts in simulations with nucleic acids and protein. On the other hand, Allnér-Villa parameters focuses on kinetic features of the ion-binding, such as activation energy and ion exchange rate.¹⁹⁸ By reproducing the kinetic properties of Mg^{2+} ion binding with water and phosphate ion, such parameterization improved the description of Mg^{2+} interacting with nucleic acid during simulations. However, these parameters underestimate the rate of water exchange and consequently they have high energetic barriers, Mg^{2+} association/dissociation too slow, binding affinity too high, and binding distance too small.¹⁸⁸ Eventually, the Li-Merz parameters¹⁹⁹ focused on divalent metal ions parameterization for MD simulations

employing the PME method. Such parameters represent a good compromise in reproducing the experimental free energies solvation, ion-oxygen distances, and coordination numbers for different divalent metal ions. Overall, among all the parameters, the more recent developed by Allnér-Villa and Li-Merz, appear to be more adequate and reliable for biochemical system simulations, even though with some limitation. Indeed both of the parameters are among the most widely used in the field.

Complementary to metal ions, fast improvements of force-fields for DNA have been achieved.^{203–205} In this regards, the main progressions can be summarized into two main paths, i.e. the work done by the Orozco group ('BSC') and the collective work done by research groups from the Czech Republic ('OL'). Currently, the most recent Amber force-fields modifications for DNA are the bsc1²⁰⁶ and the OL15²⁰⁷. Both were developed to improve the accuracy of MD simulations of double-stranded DNA. Specifically, the bsc1 includes previous bsc0²⁰⁸ modifications and additional modifications to the sugar pucker, the χ glycosidic torsion, and the ϵ and ζ dihedrals. On the other hand, the OL15 consist of the combination of bsc0 modifications and additional modifications to the χ torsion, the β , ϵ and ζ dihedrals.²⁰⁷ Given the importance of being able to simulate the structure and dynamics of nucleic acids, few studies have been conducted to compare such force-fields for DNA.^{209,210} The results of such comparisons have led to the conclusions the both parameters perform in a similar manner and they are equally reliable.

2. Enhanced Sampling Techniques

A possibility to overcome timescale issue in MD simulations and drive the system to explore the phase space of interest, is represented by the enhanced sampling techniques. These are extensively used and allow the system to sample complex (or rare) events. In general, a bias potential energy is added to the Hamiltonian of the system, thus lowering the energy barrier and increasing the sampling transition space. Such techniques include a wide variety of methods, such as umbrella sampling, replica exchange molecular dynamics, adaptive biasing force method, simulated annealing and metadynamics. This latter have been employed in this thesis to characterize rare events for the systems of interest.

1.2. Metadynamics

Metadynamics (MetaD) is an atomistic simulation method that allows acceleration of rare events and estimation of the free energy of complex molecular systems.²¹¹ In MetaD the bias is added to the potential on a selected number of degrees of freedom, denoted as Collective Variables (CVs).^{212,213} CVs are defined as a low dimensional function (S) of microscopic coordinates (R), which describe the slow motion in the process taken into account. This external bias (V_G) is history-dependent, thus discouraging previously visited states be re-sampled and it is composed of intermittently added Gaussian functions.²¹⁴ It can be written as:

$$V_G(S, t) = \int_0^t dt' \omega \exp\left(\frac{\left(S_i(R) - S_i(R(t'))\right)^2}{2\sigma_i^2}\right) \quad [7]$$

Where σ_i is the width of the Gaussian hills for the i^{th} CV, and ω is a constant energy rate defined as the ratio of the Gaussian height (W) and the deposition stride (τ_G):

$$\omega = \frac{W}{\tau_G} \quad [8]$$

The general flow of metadynamics simulations can be schematized as:

1. Initially, the walker is trapped in a minimum of the potential energy surface and the external bias is zero.
2. Gradually adding Gaussian hills on the selected CVs, the walker is forced out into the unexplored regions of the energy landscape.
3. Once the free energy surface (FES) is completely filled and the walker can freely explore all the minima in the landscape, the simulations has reached the convergence.
4. The criterion of convergence is necessary in order to reconstruct properly the FES by summing the deposited Gaussian hills.

Indeed, the free energy can be derived from a metadynamics calculation because the V_G provides an unbiased estimate of the free energy:

$$\lim_{t \rightarrow \infty} V_G(S, t) \sim -F(S) \quad [9]$$

Where $F(S)$ is expressed as:

$$F(S) = -\frac{1}{\beta} \ln \left(\int dR \delta(S - S(R)) e^{-\beta V(R)} \right) \quad [10]$$

In which $\beta = (k_B T)^{-1}$, where k_B is the Boltzmann constant, T is the temperature of the system and $V(R)$ is the potential energy function.

Two main drawbacks can arise from metadynamics simulations, the selection of the CVs and the determination of convergence. The reliability of the reconstructed free energy landscape from metadynamics crucially depends on the employed CVs. This reliance is common to all methods based on adding a bias potential that only depends on selected CVs. The latter can span from very simple variables like distances or dihedral angles to more complex variables like PCA or Path CV. In order to guarantee metadynamics to work effectively, CVs have to respect some conditions:

- For a multistable system, different metastable states should correspond to different values of the CVs.
- The CVs should be able to distinguish transition states. Indeed, metadynamics is inclined to accelerate transitions by stabilizing the transition state relative.
- The CVs employed should be limited in number, since reaching convergence of a multidimensional space becomes more computationally expensive as the dimensionality of the space grows.

Once the CVs are correctly selected, the free energy minima are filled with Gaussians and the dynamics becomes diffusive in CV space and the metadynamics gradually reaches the convergence criterion.²¹⁵ However, if transitions between different states rarely happen during the simulations, the time average of the bias potential is not guaranteed to converge to the negative of the free energy as in [9]. On the contrary, extending too much the simulations may drive the system in CV space regions that are not physically relevant. For this, determining whether to proceed or to interrupt the metadynamics is absolutely not trivial. Fortunately, a possible solution to this issue is offered by the well-tempered metadynamics method.²¹⁶

1.3. Well-Tempered Metadynamics with a Confined Procedure

Well-tempered metadynamics is a variant of metadynamics in which the heights of the Gaussian hills (W) are scaled at each step according to:

$$W = \omega \tau_G e^{-\frac{V_G(S,t)}{k_B \Delta T}} \quad [11]$$

In practice, during well-tempered metadynamics the bias deposition rate decreases over simulation time and the external bias V_G is expressed by:

$$V_G(S, t) = \Delta T \ln \left(1 + \frac{\omega N(S, t)}{\Delta T} \right) \quad [12]$$

Where ω has the dimension of an energy rate, and $N(S, t)$ is the histogram of the S variables obtained from the simulations. Consequently, using [12] the FES can be estimated as:

$$F(S, t) = -\frac{\Delta T}{T + \Delta T} V_G(S, t) \quad [13]$$

Considering the two limiting situations, i.e. $\Delta T = 0$ and $\Delta T \rightarrow \infty$, in the first case the bias is zero, thus ordinary MD is recovered, while for second case the deposition rate is constant, thus standard metadynamics is retrieved. Advantages of well-tempered method are several. For example, the exploration in the CVs space is facilitated by tuning ΔT value, which in turns may limit the exploration to physically interesting regions of the FES. Eventually, issues related to longer simulations, such as overfilling problem, are avoided.

Combined with well-tempered metadynamics, the confined procedure has been developed by La Sala et al. in order to deal with specific cases.^{216,217} Considering a system having a bistable energy profile, where basin A gathers conformations of interest while basin B includes irrelevant conformations. To optimize the use of computational resources, the goal is to sample as much as possible the states in basin A, avoiding to sample basin B. Possible solutions to achieve this goal are available, such as the use of restraining potential, however these are often related to non-trivial issues. Here, the confined approach offers a possible option. With this method, during the metadynamics simulations when a walker hits a selected CV target value, which delimits the boundary between basin A and basin B, the simulation stops. Subsequently, it is restarted from a random conformation stored in a

pool P of frames saved during the exploration of basin A. In this way, the walker continue to sample relevant conformations by improving the sampling of transversal and often rather slow degrees of freedom. In practice, the approach can be schematized as follow:

1. The metadynamics starts and the selected CV is monitored every n ps to determine which state the system is exploring.
2. If the system is in A, the current conformation is stored in the pool while the simulation can proceed.
3. If the system is found in B, the simulation is restarted from a randomly chosen conformation previously stored in the pool.

This approach have been successfully applied in this thesis in order to study the unbinding process of a leaving group from the active site of enzymes.

1.4. Path Collective Variable

In many cases regarding complex biological systems, such as large conformational changes or ligand-protein interactions, finding a limited number of CVs that take into account all the relevant degrees of freedom is not obvious. For this reason, in cases where the initial and the final states are known it is possible to define a putative path to connect them and two CVs which are functions of it. These latter allow to compute the progress along a high-dimensional path (s) and the distance from the high-dimensional path (z). The fist (s) is computed as:

$$s = \frac{\sum_{i=1}^N i \exp(-\lambda R[X - X_i])}{\sum_{i=1}^N \exp(-\lambda R[X - X_i])} \quad [14]$$

While z is measured as:

$$z = -\frac{1}{\lambda} \ln \left[\sum_{i=1}^N \exp(-\lambda R[X - X_i]) \right] \quad [15]$$

Where N is the number of high-dimensional frames (X_i) describing the path, $R[X - X_i]$ are the distances from each frames. In general, the protocol to apply Path CV approach with metadynamics simulations require the description of the path connecting the starting and ending structures.²¹⁸ This can be done using a morphing server to generates N equally spaced intermediate structures (nodes), or the nodes are defined by specific values of

secondary CVs. Thus, using the first approach, the path is formed by a series of intermediary structures with atomic coordinates, while in the second case the nodes are determined by values of CVs.

In this thesis, Path CV metadynamics has been used to investigate and characterize DNA/enzyme Michaelis-Menten complex formation.

Chapter III. Recruiting Mechanism and Functional Role of a Third Metal Ion in the Enzymatic Activity of 5' Structure-Specific Nucleases

Abstract

Enzymes of the 5' structure-specific nuclease family are crucial for DNA repair, replication, and recombination. One such enzyme is the human exonuclease 1 (hExo1) metalloenzyme, which cleaves DNA strands, acting primarily as a processive 5'-3' exonuclease and secondarily as a 5'-flap endonuclease. Recently, in crystallo reaction intermediates have elucidated how hExo1 exerts hydrolysis of DNA phosphodiester bonds. These hExo1 structures show a third metal ion intermittently bound close to the two-metal-ion active site, to which recessed ends or 5'-flap substrates bind. Evidence of this third ion has been observed in several nucleic-acid-processing metalloenzymes. However, there is still debate over what triggers the (un)binding of this transient third ion during catalysis and whether this ion has a catalytic function. Using extended molecular dynamics and enhanced sampling free-energy simulations, we observed that the carboxyl side chain of Glu89 (located along the arch motif in hExo1) flips frequently from the reactant state to the product state. The conformational flipping of Glu89 allows one metal ion to be recruited from the bulk and promptly positioned near the catalytic center. This is in line with the structural evidence. Additionally, our simulations show that the third metal ion assists the departure, through the mobile arch, of the nucleotide monophosphate product from the catalytic site. Structural comparisons of nuclease enzymes suggest that this Glu(Asp)-mediated mechanism for third ion recruitment and nucleic acid hydrolysis may be shared by other 5' structure-specific nucleases.

1. Introduction

Recent structural data have shown the recurring presence of a third metal ion close to the two-metal-ion center of nucleic-acid-processing enzymes.^{2,53,61,125} This third ion has been captured during different stages of catalysis of vital enzymatic reactions involved in DNA repair, recombination, and replication processes.²¹⁹⁻²²³ These reactions are often

related to cancer progression.^{5,224–226} Indeed, over the last few years, a third ion has been observed, or hypothesized, close to the two-metal-ion catalytic site of polymerases,² nucleases,^{79,132,144,147,227} and topoisomerases.^{134,228,229} This suggests that the third metal ion may be actively involved in catalysis.^{2,61,230} However, there is debate over how this ion is recruited from the bulk and transiently binds the enzyme, and how it could play a role in catalysis.

In this context, recent time-resolved *in crystallo* reaction intermediates¹⁴⁸ have elucidated how human exonuclease 1 (hExo1) exerts its catalytic function, with sequential structures showing how the enzyme/DNA complex evolves during catalysis. hExo1 is an essential hydrolytic enzyme for genome maintenance. Belonging to the RAD2/XPG family,^{231–237} hExo1 is a 5' structure-specific metallonuclease, which carries out a primary exonucleolytic activity on the 5' recessed-end and a secondary endonucleolytic cleavage on 5'-flap of the substrate DNA strand.^{165,238}

The structures show the enzymatic mechanism for DNA hydrolysis in hExo1, which starts in the precatalytic state with the intact double-strand DNA (dsDNA) recognized and bound (tethered) to the helix-two-turn-helix (H2TH) motif and to a monovalent (K⁺/Na⁺) ion state in hExo1 (Figure III.1). Then, catalysis begins with formation of the assembled active site, where the dsDNA bifurcates into the 5' and 3' single strands (i.e. the dsDNA 'junction'). At this point, the scissile phosphate of the processed single 5' strand is properly located at the reactive metal center at the N-terminal domain. In this state, the catalytic residues Lys85 and Arg92 interact with the scissile phosphate,¹⁶⁴ after a rotation (clamped conformation) of the mobile helical arch formed by two α -helices ($\alpha 4$ - $\alpha 5$) located near the junction. Here, the side chain of the guide residues Tyr32 and His36 are also rotated (Figure III.1). These structural motifs contribute to the "threading mechanism", whereby the 5'-flap DNA passes through the helical arch.¹⁵⁸ In this way, basic residues steer the phosphate of the 5' strand, promoting the proper location for hydrolysis of the scissile phosphate on top of the two catalytic ions, as expected for the recognized two-metal-ion mechanism.^{46,48,49,63,158,239,240}

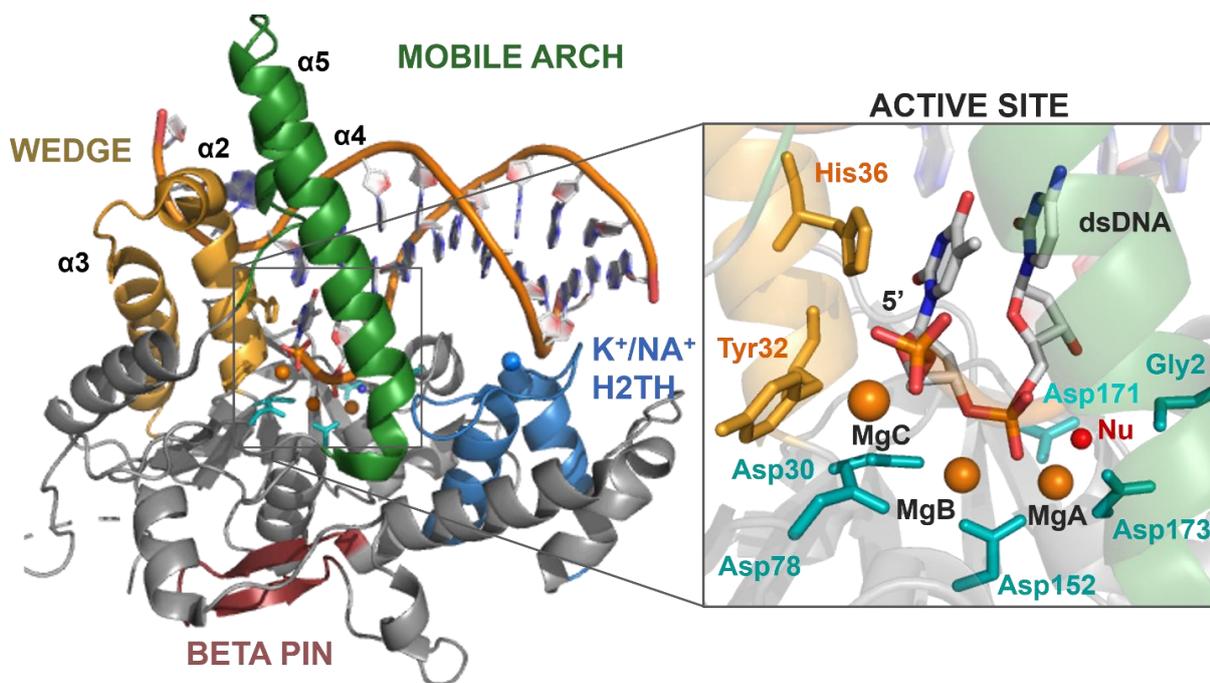


Figure III.1 Catalytic domain of hExo1 in complex with DNA substrate and the two catalytic metal ions (PDB ID 5V06). Left: hExo1 in cartoon and with colors for different structural motifs. Right: Closer view of the active site. The three metal ions (MgA, MgB, MgC) are in orange; the nucleophilic water molecule (Nu) is in red; the two guide residues (Tyr32, His36) in yellow; and the residues of the catalytic pocket (Gly2, Asp30, Asp78, Asp152, Asp171, Asp173) are in cyan. The scissile phosphate is correctly positioned for the nucleophilic attack and MgC is coordinated by the 5' terminal phosphate.

At this point, hydrolysis of the 5'-recessed end or the 5'-flap DNA substrate occurs.^{162,241,242} In hExo1, this is proposed to be favored by a structured network of interactions involving Arg95, Arg96, Arg121, and Asn124, which are located along the mobile arch. In particular, a key role in phosphate steering is proposed for the Arg96 and Asn124, which interact with the phosphate next to the scissile one (i.e. the terminal 5' phosphate in the 5' recessed-end substrate). This is similar to what has been observed in the enzyme hFEN1.¹⁵⁸ After DNA hydrolysis, the nucleotide monophosphate group can leave the active site, with the “free” enzyme that now has Tyr32 and His36 back in their initial conformation.

Remarkably, these structural data show a transient third metal ion that is intermittently located close to the catalytic site during exonuclease catalysis (Scheme III.1). This suggests that the transient third ion may play a role in substrate hydrolysis and/or leaving group departure.³³ Indeed, during hExo1 catalysis, four different structures of the assembled active site were solved in the presence of a second-shell and solvent-exposed third metal

ion, preserved close to the two-metal ion center (Figure A.1). This ion is not found in the structure of the cleaved product, demonstrating its transient nature during catalysis.^{59,61}

Here, we used force-field-based molecular dynamic (MD) simulations coupled to enhanced sampling free-energy calculations to investigate the role of the third ion during hExo1 catalysis. We compared several systems of wild-type and mutated hExo1 from the reactant state to the product state. We found that the second-shell and conserved Glu89 residue selects, recruits, and places the third ion close to the two-metal-ion catalytic site. We show that this negatively charged residue is functional and conserved among Exo1 belonging to different organisms, and that this enzymatic mechanism is likely shared by other nucleases.

2. Results

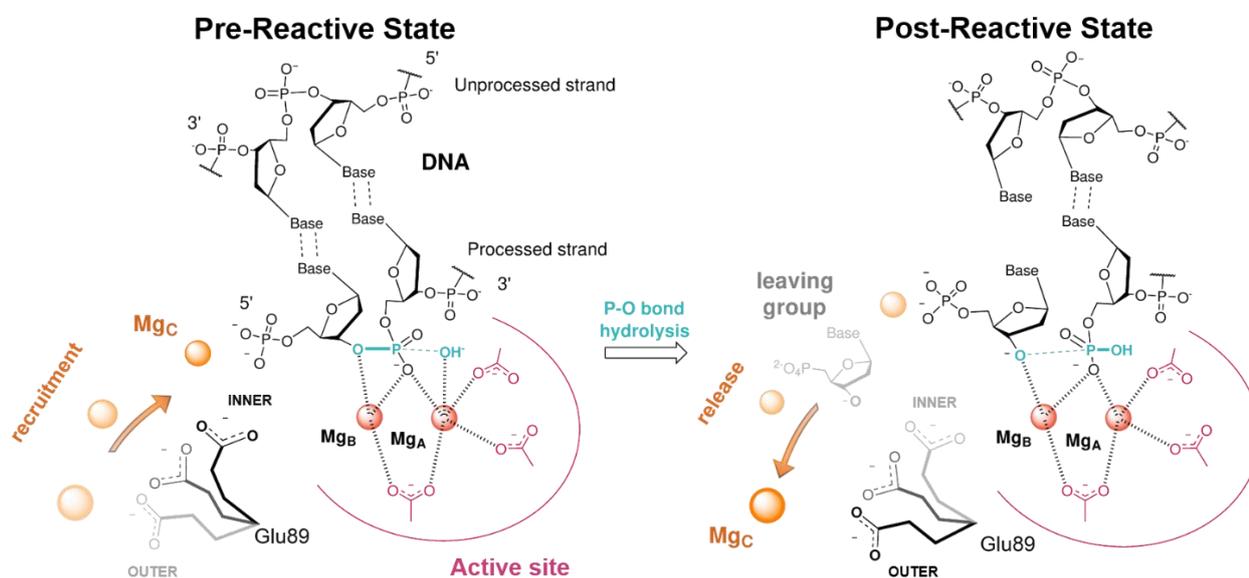
Glu89 selects, recruits, and places a third ion close to the catalytic site of hExo1.

First, we ran multiple unbiased force-field-based molecular dynamic (MD) simulations (~700 ns in total) of the wild-type (*wt*) reactant state (Figure A.2). We considered the enzyme in the reactive state, replacing the non-reactive Mn ions in the pre-reactive crystal structure (PDB ID 5V06)⁵⁹ with native Mg ions.¹⁶⁵ Notably, this structure features a third Mg ion (MgC) close to the two-metal-ion catalytic site. This ion is bound to the 5' phosphate of the processed strand.

Residues Tyr32 and His36 are thought to guide substrate binding through interactions with their flexible side chain.⁵⁹ Located above the reaction center, these residues maintain the crystallographic conformation, in which their side chain points “down” toward the two catalytic ions MgA and MgB beneath it (Figure A.3). These ions maintain an octahedral coordination²⁴³ throughout the simulated timescale (i.e. 360 ns). As a result, the nucleophilic water molecule remains optimally positioned for nucleophilic attack, sitting on top of MgA, in front of the substrate’s scissile phosphodiester bond, at 3.88 ± 0.11 Å (Figure A.3).

Despite the overall stability of the protein-DNA complex, the side chains at the base of the mobile arch (i.e. at the gateway) sample different conformations. In particular, the initial clamped conformation shows some flexibility over time, with Lys85 moving slightly further away from the scissile phosphate (~6 Å vs 4 Å in PDB ID 5V06 see Figure A.4).

Moreover, we observed an amplified mobility of MgC, reflected in its enhanced RMSD of $1.63 \pm 0.42 \text{ \AA}$, as compared to the two catalytic ions and its anchor 5' nucleotide, which are highly stable with an RMSD of $0.54 \pm 0.17 \text{ \AA}$ and $0.82 \pm 0.18 \text{ \AA}$, respectively. Indeed, MgC moves from its initial position, forming new interactions with the flexible Glu89 carboxylate along the α -helix of the gateway. As a result, the mobile MgC alternates between a bidentate and monodentate binding coordination with the 5' phosphate and Glu89 (Figure A.5), always maintaining an octahedral shell. In this regard, the motions of Glu89 side chain are described by the pseudo dihedral angle ϕ (taken along the N, C α , C δ , C γ bonds - see Figure A.5), which oscillates from $\sim 120^\circ$ to $\sim -10^\circ$ (Figure A.5), never reaching a conformation of the product state (-36° , PDB ID 5V0A)⁵⁹.



Scheme III.1 Schematic representation of the motion of the transient third metal ion, which we found to be intermittently recruited/released by Glu89 (switching its inner/outer conformations) during exonuclease catalysis. This structural evidence suggests that the transient third ion may be crucial for substrate hydrolysis and/or leaving group departure.

Concomitantly, we observed that Arg96, which is located along the mobile arch, shortens its distance from the 5' phosphate, reaching a value of $\sim 4.5 \pm 0.19 \text{ \AA}$ (compared to an initial distance of 6 \AA in PDB ID 5V06). This distance corresponds well to the value in the crystal structure of the product state (5.1 \AA , PDB ID 5V0A) (Figure A.6). Taken together, these results support the hypothesis of a gradual removal and possible departure of the third ion during catalysis, as suggested by comparing the structural data of the reactant (MgC present, PDB ID 5V06) and product states (MgC missing, PDB ID 5V0A).

To further characterize the structural impact of MgC bound close to the two-metal-ion catalytic site in the reactant state, we manually removed it and ran multiple simulations of the solvated system ($\sim 1 \mu\text{s}$, in total). This protein-DNA complex showed no major difference in the overall backbone stability (Figure A.2) compared to the three-ion reactant state (see above). However, in these replications, the catalytic residues Lys85 and Arg92 maintained their native H-bond pattern with the scissile phosphate (as in the crystallographic structure) for the whole simulation. Thus, the catalytic residue Lys85 behaved differently to the three-ion system. Also, in these simulations, at times Glu89 interacted transiently with Arg93, located along the mobile arch. This interaction was observed only when Glu89 adopted the outer conformations, in the absence of MgC. Concurrently, Arg96 maintained its starting orientation and never interacted with 5' phosphate (Figure A.6). Again, this differs from the observations in the presence of MgC (see previous para-graph).

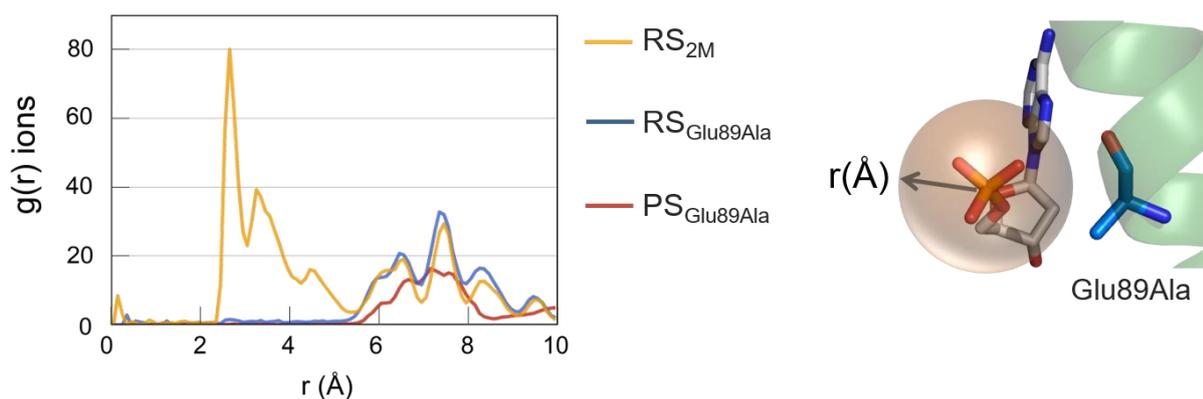


Figure III.2 Radial distribution function, $g(r)$, calculated for ions around 10 \AA from the center of mass of the 5' phosphate group. The plot shows the presence of ions $\sim 3 \text{ \AA}$ from the 5' phosphate group for the RS_{2M} system. In this system, a K^+ ion approached the negatively charged group. For the $RS_{Glu89Ala}$ and $PS_{Glu89Ala}$ systems, there are no ions within $\sim 5.5 \text{ \AA}$ of the 5' phosphate, as indicated by the $g(r)$ values of ~ 0 . In the upper right corner, the 5' phosphate group and Glu89Ala residues are shown in licorice (taken from the $PS_{Glu89Ala}$ simulations).

Intriguingly, during the initial equilibration phase ($\sim 10 \text{ ns}$), the side chain of Glu89 undergoes a marked rearrangement from the initial inner conformation to an outer conformation toward the bulk water. This rotation is captured well by the Glu89 pseudo dihedral angle ϕ , which changes from positive values $\sim +100^\circ$ (inner) to negative values $\sim -95^\circ$ (outer) (Figure A.7). Importantly, in this new solvent-exposed conformation, we observed that Glu89 carboxylate transiently recruits and binds monovalent ions from the bulk (either K^+ or Na^+ , freely diffusing in solution). This result is also confirmed by the

radial distribution function, $g(r)$, calculated as the variation of the density of the ions from the center of mass of the 5' phosphate group (Figure III.2), which displays two peaks at $\sim 2.7 \text{ \AA}$ and $\sim 3.3 \text{ \AA}$.

After metal binding ($\sim 10 \text{ ns}$), Glu89 flips back into its initial inner conformation, carrying the coordinated metal closer to the 5' phosphate, at $\sim 3.25 \text{ \AA}$. This metal ion is thus brought into a very similar location compared to MgC in the crystal structure of the pre-reactive state (PDB ID 5V06). After a few hundred ns (e.g. $\sim 200 \text{ ns}$ for K^+ ion), the third metal ion departs spontaneously from the catalytic site. Glu89 then flips again into its outer conformation toward the bulk. These unprompted metal binding and release events, synchronized with the flipping of the Glu89 side chain, were observed multiple times in our extended simulations (Figure A.7). This indicates that Glu89 may recruit a third metal ion, bringing it transiently closer to the catalytic site.²⁴⁴

To further test Glu89's role as metal ion recruiter, we ran multiple MD simulations ($\sim 1 \mu\text{s}$ in total), inserting the Glu89Ala mutation in the reactant state in the absence of MgC. The overall stability of the enzyme-DNA complex was maintained, with a low RMSD value of $1.27 \pm 0.15 \text{ \AA}$ (see Figure A.2). The pre-reactive state at the active site was also maintained throughout the simulations. That is, the two catalytic metal ions, MgA and MgB, stably maintained their internuclear distance. The nucleophilic water molecule remained properly positioned in front of the scissile phosphodiester bond and the catalytic residues. Finally, Lys85 and Arg92 maintained their initial interaction network with the scissile phosphate. Notably, in this mutated system, we did not see any ion approaching the terminal 5' phosphate from the bulk solvent. This result is supported by $g(r)$, which confirms that no ion is located within $\sim 6 \text{ \AA}$ of the center of mass of the 5' phosphate group (Figure III.2), further suggesting that Glu89 recruits a third metal ion from the bulk.

The third ion promotes leaving group departure after DNA hydrolysis. Here, we used MD simulations ($\sim 1 \mu\text{s}$ in total) of the products of the (*wt*) native state. Thus, we inserted a native aspartate at the Asp225Ala mutation and replaced Mn with native Mg ions in the post-reactive crystal structure (PDB ID 5V0A). Notably, at this catalytic stage, the DNA's processed strand is enzymatically cleaved, with the consequent generation of the leaving group, i.e. the adenosine 5'-monophosphate (AMP) nucleotide, which is now

detached from the newly formed 5' recessed-end substrate. Importantly, the third ion is not present at the catalytic active site in the crystallographic structure.

In the post-reactive crystal structure, the Glu89 side chain adopted an intermediate conformation (ϕ of Glu89 was -36° , Figure A.1) between the inner ($\sim +100^\circ$) and outer ($\sim -95^\circ$) conformations. Then, during the MD simulations, the Glu89 side chain stably adopted an outer conformation (ϕ of Glu89 becomes $\sim -100^\circ$). However, after ~ 50 ns, we observed the unprompted approach of a third Mg^{2+} ion from the bulk water (Mg_{bulk}), which came close to the Glu89 side chain. This transient ion thus reached a position close to the 5'-monophosphate of the AMP, at a distance of ~ 3.3 Å, which was equivalent to that in the pre-reactive simulations and crystal structure (PDB ID 5V06). In this position, the third metal ion interacted with the 5' phosphate and Glu89 for the remaining simulation time (Figure III.3). During this event, the two catalytic metal ions moved apart slowly, reaching a distance of ~ 5.5 Å (compared to 3.9 Å in the starting model PDB ID 5V0A). The drifting of the internuclear two-metal-ion distance was coupled to a shift in the leaving AMP. This is described well by the collective variable CV1, which measures the distance between the center of mass (COM) of the heavy atoms of AMP and the COM of the C α of the aspartates in the first coordination shell of the two-metal-ion center (i.e. Asp152, Asp171, Asp173) (Figure A.8). During our simulations, CV1 increased by ~ 2 Å, from 9 Å to 11 Å, reflecting the partial exit of the leaving AMP (Figure A.8). Moreover, His36, which was initially in the down orientation, immediately rotated into the up conformation (Figure A.9), forming a π - π interaction with AMP. This interaction helps the initial displacement of the leaving AMP. Notably, the up conformation of His36 was found in the structure of the enzyme after AMP departure (PDB ID 5V0B)⁵⁹, which further suggests the need of this rotation during leaving group release. We also noted a gradual and slight opening in the gateway region at the bottom of the mobile arch, which however maintained an ordered secondary structure (Figure A.10). This event is described well by the increase of ~ 1 Å of the two distances d1 and d2, which reflect the opening of the $\alpha 4/\alpha 5$ inter-helix passage (calculated using the C α of Glu89 and Arg92 along the $\alpha 4$ helix, and the C α of Asn124 and Ile125, located at the bottom of $\alpha 5$ helix – see Figure III.4). The probability density function of ϕ , calculated for two states ($\text{Mg}_{\text{bulk}} > 4$ Å or $\text{Mg}_{\text{bulk}} < 4$ Å from the 5' phosphate), shows the

relative peaks of the two conformations assumed by Glu89, i.e. inner and outer (Figure III.3).

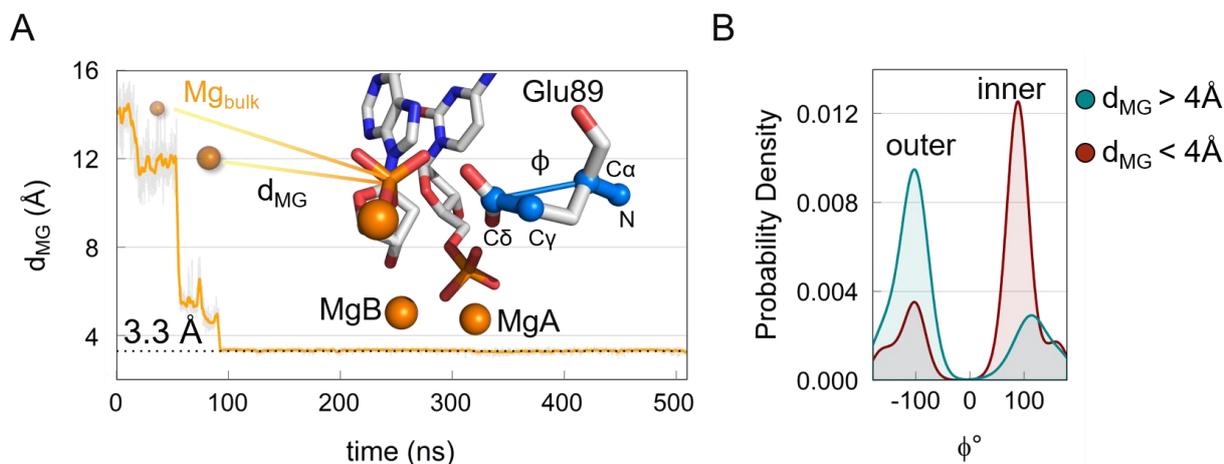


Figure III.3 (A) Distance (d_{MG} in yellow) between the third Mg^{2+} ion, from the bulk (Mg_{bulk}), and the phosphorous of 5' phosphate group of AMP. Into the graph, a representation (snapshot from the PS_{2M} simulations) of Mg_{bulk} approaching the terminal 5' phosphate, as well as the pseudo dihedral angle ϕ of Glu89 side chain (defined by the N-C α -C δ -C γ atoms) are reported. (B) Probability density of the pseudo dihedral angle ϕ in Glu89, during the simulation. In blue, the probability density, for d_{MG} values $> 4 \text{ \AA}$, shows the outer conformation as the most populated; in red the probability density, for d_{MG} values $< 4 \text{ \AA}$, shows the inner conformation is the most populated.

In the products, Arg96 invariantly interacts with the 5' phosphate in AMP (Figure A.6), as also reported for the reactant state simulations in the presence of the MgC. Moreover, after ~ 450 ns, we noted a second arginine residue located along the $\alpha 4$ helix, Arg93, which approached the same 5' phosphate of the AMP, at $\sim 5 \text{ \AA}$. These interactions thus favor the positional shift and partial departure of the AMP from the catalytic site, as indicated by CV1, with consequent destabilization of the two-metal-ion site. We compared these results with an additional post-reactive model, which initially contained a third ion at the catalytic site ($\sim 1 \mu s$ in total, see Appendix A). These simulations confirmed the enhanced instability of the prone-to-escape leaving group, with a partial shift from its starting position during the simulations (Figure A.8). Taken together, these results suggest that complete leaving group departure is eventually expected, although this would require longer simulations, and would also likely implicate the overtaking of an energetic barrier.⁶²

As with the reactant state (see above), we also ran simulations ($\sim 1 \mu s$ in total) in the product state of a system with the Glu89Ala mutation, in the absence of MgC. The overall stability of the enzyme-DNA complex was maintained (Figure A.2). Importantly, in the

absence of the Glu89 recruiter, no ion approached the 5' phosphate of the leaving group, as shown by $g(r)$ (Figure III.2). As a result, the leaving group also showed higher stability in its position, as highlighted by the low RMSD value of $1.71 \pm 0.18 \text{ \AA}$. Moreover, Arg93 almost never interacted with the 5'-phosphate of AMP, and Arg96 stably maintained its initial interaction with AMP. In addition, we did not observe any opening in the gateway region, with the distances $d1$ and $d2$ remaining unchanged during the simulations (Figure III.4). These results support the hypothesis that Glu89 recruits MgC before (or during) DNA cleavage. In return, MgC seems to promote the release of the leaving group after the chemical step for phosphodiester bond hydrolysis, acting as a shuttle for the AMP departure.

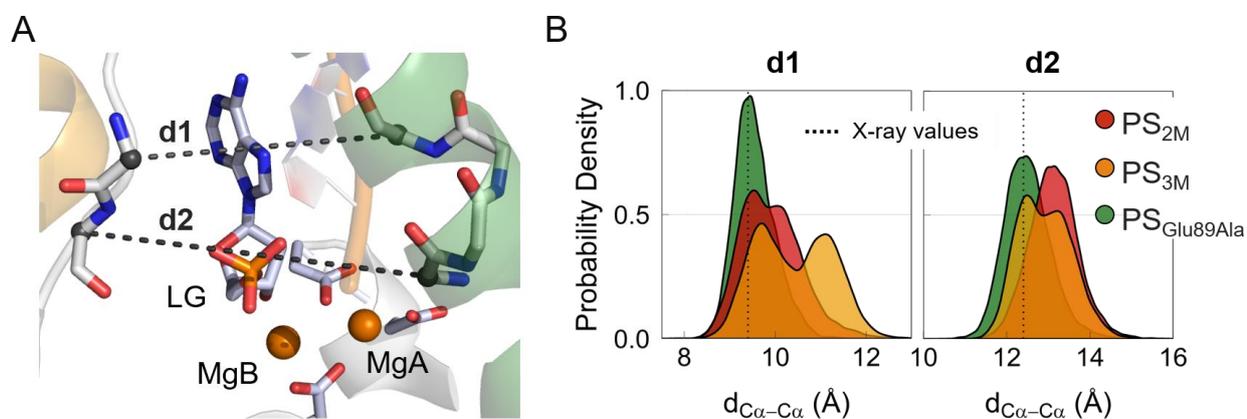


Figure III.4 (A) Graphic representation of $d1$ and $d2$ distances. PDB ID: 5V0A. (B) Probability density of the distance $d1$ and $d2$ calculated during simulations of the systems PS_{2M} (in red), PS_{3M} (in orange), and PS_{Glu89Ala} (in green).

Energetics of the Glu89 flipping and leaving group departure via metadynamics simulations. To sample and determine the semiquantitative energetics of the inner \leftrightarrow outer conformational switch of Glu89, we used the pseudo dihedral angle ϕ as the collective variable to run multiple metadynamics simulations, with and without the third metal ion at the catalytic site, in the reactant and product states (for a total of ~ 920 ns).

In the reactant state with the third ion, Glu89 tended to adopt inner conformations located in an energy minimum at $\phi \sim 70^\circ$, while outer conformations were not visited due to their high energy (Figure III.5, red profile). In the inner conformation, MgC stayed close to the reactive center. However, in the absence of the third metal ion, Glu89 could be found in two isoenergetic minima, i.e. inner and outer conformations, separated by a barrier of only $\sim 3 \text{ kcal mol}^{-1}$ (Figure III.5, blue profile). This explains the fact that both Glu89

conformations were similarly populated in our unbiased MD simulations of the system without the third metal ion.

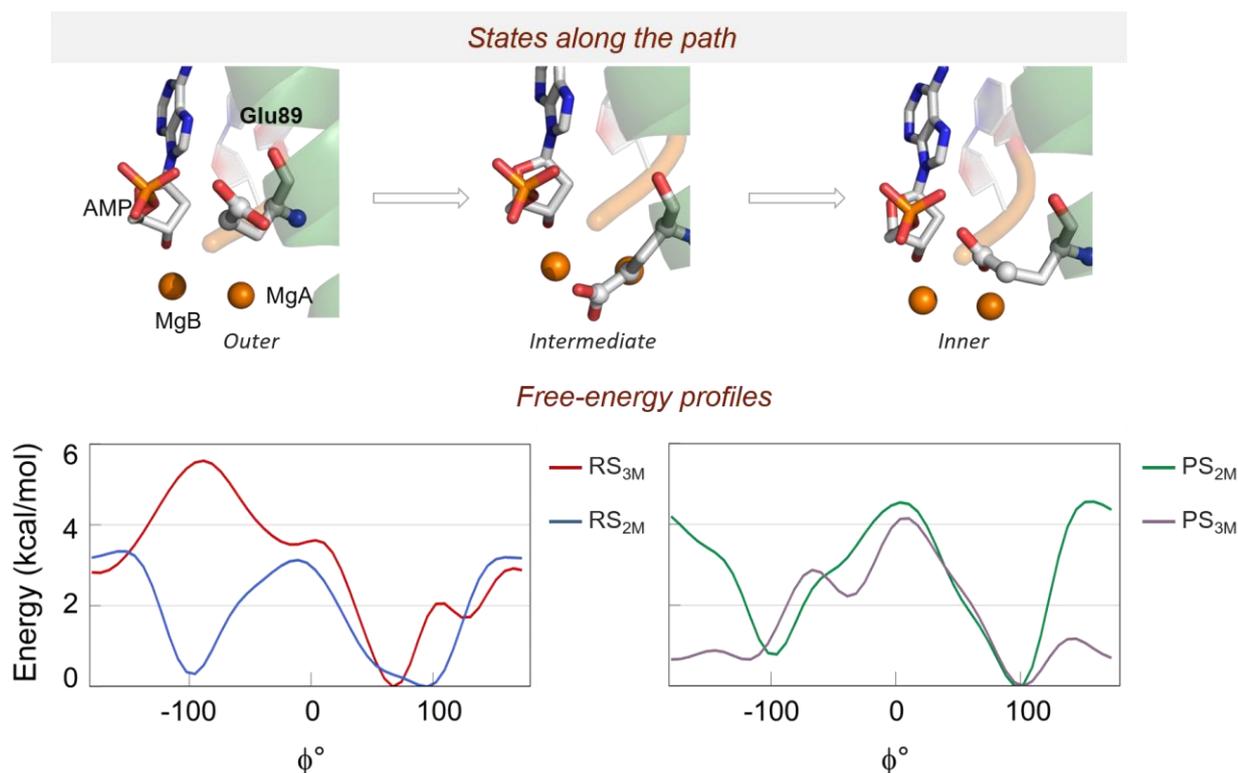


Figure III.5 (top) Graphic representations, taken from PS_{3M} simulations, of the three conformations are shown in licorice. (bottom) Free energy surface obtained through well-tempered metadynamics simulations for RS_{3M} (red), RS_{2M} (blue), PS_{2M} (green), and PS_{3M} (light purple) systems. The results show three conformations (outer, intermediate, and inner).

In the product state, regardless of the presence or absence of the third ion, Glu89 visited both the inner and outer conformations (Figure III.5, green and light purple profiles). The conformational switch showed a barrier of ~ 4.5 kcal mol⁻¹, with or without MgC. We also located a metastable conformation of Glu89 bound to MgC, at $\phi \sim -40^\circ$, in which the glutamate's side chain adopted an intermediate orientation between the two minima (inner and outer). Interestingly, this metastable state corresponds well to the crystallographic conformation ($\phi = -36^\circ$, PDB ID 5V0A) in which MgC is missing. This is likely because, at this point, Glu89 is already solvent-exposed.

Then, we evaluated possible pathways and energetics for the release of the leaving group from hExo1 in the presence and absence of MgC. We used confined metadynamics,²¹⁷ which enhances the sampling of transversal and often slow degrees of freedom of complex (rare) events, such as the exit of the adenosine monophosphate (AMP) nucleotide from the catalytic site (see Appendix A for further information). Here, the collective variable was

CV1 (Figure III.6), which captures the degree of departure of the leaving AMP from the reactive site (see definition of CV1, above).

In the presence of MgC, AMP fell into a minimum, at $CV1 = \sim 12 \text{ \AA}$, showing that the leaving group is already shifted out of the catalytic site ($CV1 = \sim 8.5 \text{ \AA}$ in the uncleaved pre-reactive state). This is in line with the plain MD simulations, where it was only in the presence of the third metal ion that the leaving group partially exited from the active site, increasing the CV1 value of $\sim 2 \text{ \AA}$, reaching a value of $\sim 12 \text{ \AA}$ (Figure A.8). At this point, to allow the full departure of the leaving group, the freed AMP stayed complexed with MgC. In this way, the AMP/MgC complex exited from the catalytic site, passing through the aperture under the mobile arch formed only when MgC is present (due to MgC-mediated AMP drifting out from the catalytic center). At this point, the Glu89 sampled the inner and outer conformations, until the leaving group overcame the gateway region. At this point, Glu89 stably adopted the inner conformation, in agreement with the crystallographic structure of the complex after the release of the leaving group (PDB ID 5V0B). The physical step for AMP/MgC unbinding occurred with a barrier of $\sim 16.5 \text{ kcal mol}^{-1}$ (Figure III.6).

In the absence of the third ion, the energy barrier for the overall unbinding process was much higher at $\sim 35 \text{ kcal mol}^{-1}$. Indeed, the system behaved quite differently. In the initial configuration, the leaving AMP fell into an energy minimum where it was poorly solvated, at $CV1 = \sim 9 \text{ \AA}$ (compared to the case where it was solvated and complexed with MgC, $CV1 = \sim 12 \text{ \AA}$, Figure III.6). From this state, the exit of the AMP alone had to overcome a first barrier of $\sim 20 \text{ kcal mol}^{-1}$, which is already higher than the barrier in the presence of MgC. Then, the exit path showed metastable states (relative energy minima at $CV1 = \sim 13.5 \text{ \AA}$ and $CV1 = \sim 17 \text{ \AA}$, Figure III.6) where the leaving AMP seemed to be transiently trapped by the formation of short-lived interactions with the enzyme. This may slow the AMP unbinding kinetics. Notably, these transient interactions were not formed for the AMP/MgC leaving complex. It is also worth noticing that in both systems PS_{3M} and PS_{2M}, we observed the exit of MgB from the catalytic pocket, which occurred concertedly with the exit of the AMP leaving. In detail, the MgB catalytic ion remained coordinated to the oxygen of the OH in C3 position of the sugar, of the AMP leaving group. Interestingly, the concerted departure of MgB, MgC and the leaving group AMP, agrees with previous

studies showing that MgB dissociates from the catalytic site ~ 200 times faster compared MgA.²⁴⁵ Thus, the concerted exit of both MgB and MgC is found here to promote AMP departure by stabilizing the newly formed negative charge on leaving group, after substrate hydrolysis.

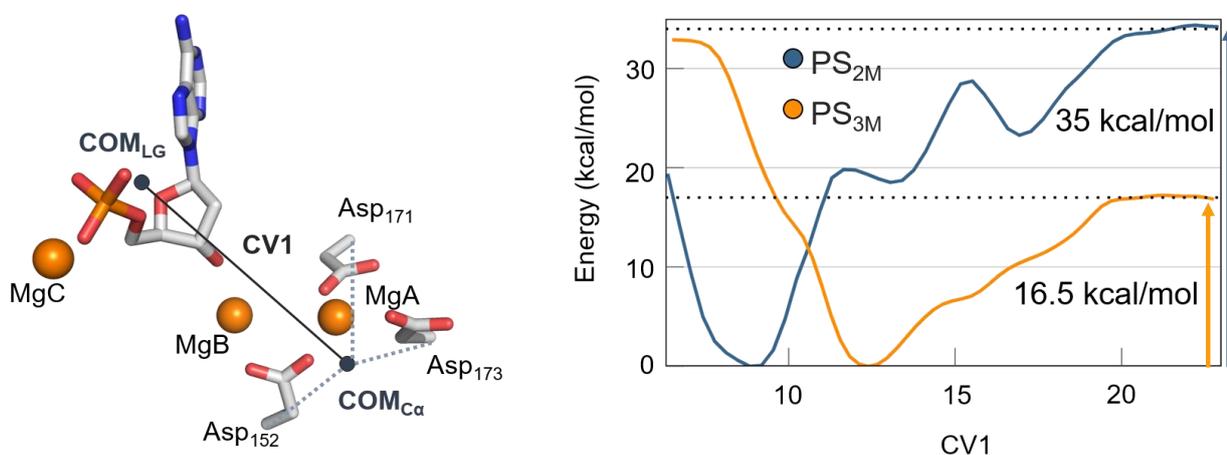


Figure III.6 Free energy surface obtained through confined well-tempered metadynamics simulations for PS_{2M} (blue) and PS_{3M} (yellow) systems. (left) Schematic representation of the CV1, exemplified using a snapshot from PS_{3M} simulations. It represents the distance between the center of mass (COM) of the heavy atoms of the nucleotide leaving group and the COM of the C α of the aspartates (Asp₁₅₂, Asp₁₇₁, Asp₁₇₃) in the first coordination shell of MgA, MgB. The two different minima, at 8.8 Å and 12.4 Å, agree with the MD results, in which a partial exit of the leaving group (CV1 \sim 12.4 Å) was seen only in the presence of MgC (Figure A.8).

3. Discussion

Recently, a time series of structural intermediates captured during human exonuclease1 (hExo1) catalysis has revealed the presence of a third metal ion (MgC) close to the active site.⁵⁹ Intriguingly, a transient metal ion was recently observed in a few DNA/RNA-processing enzymes.^{2,53,61,150} The role of this third additional metal ion at the catalytic center is still unclear.³³ Here, we used force-field-based molecular dynamics (MD) simulations and free-energy calculations to investigate the recruiting mechanism and functional role of a third metal for hExo1 catalysis. We simulated and compared several model systems, built with recent hExo1 structures of the wild-type (wt) hExo1/DNA complex, in the reactant and product states, with and without MgC.

During our multiple and extended MD simulations (~ 6 μ s in total), we first observed that Glu89 sometimes oscillated, but mostly maintained its starting conformation. In this inner conformation, the Glu89 carboxylate group points towards the 5' phosphate.

However, at times and only in the absence of MgC, the Glu89 carboxylate group switched its orientation, adopting outer conformations that pointed towards the bulk solvent. Free-energy calculations confirmed that, in the reactant state and in the presence of MgC, Glu89 tended to adopt inner conformations, which are located in an energy minimum at $\phi \sim 70^\circ$. Outer conformations are not visited due to their high energy. However, in the absence of MgC, the inner and outer conformations become isoenergetic, with a barrier of only ~ 3 kcal mol⁻¹ in between.

Importantly, following this conformational switch in equilibrium MD simulations in the absence of MgC, we observed that transient monovalent ions were freely recruited from the bulk, by the outer conformation of Glu89. This conformation therefore seems to act as an anchor point for (third)metal-enzyme complexation. Then, Glu89 could switch back, and adopt the inner conformation, bringing the bound metal ion (either K⁺ or Na⁺, from these simulations) close to the terminal 5' phosphate (~ 3.5 Å) (Figure A.7). This third metal was spontaneously recruited and located in the same position as the third ion captured in the pre-reactive crystal structure (PDB ID 5V06). This Glu89-mediated mechanism for metal recruitment was further validated by simulations of mutated Glu89Ala systems. These simulations confirmed that, in the absence of Glu89, no metal ion from the bulk was spontaneously recruited close to the catalytic center. Interestingly, the role of Glu89 in hExo1 is similar to the role previously proposed for Glu188 in *Bacillus halodurans* ribonuclease H (*BhRNase H*), where MD simulations suggested that this residue attracted a transient third ion.²⁴⁴ Intriguingly, a transient third solvent-exposed cation was found close to the two-metal-ion active site of *D. mobilis* homing endonuclease, I-DmoI.⁵⁸

In the unbiased MD simulations of the product state, we observed the unprompted entry of the third metal ion MgC from the bulk, reconstituting the three-metal-ion system (Figure III.3). This happened concomitantly to the rotation of the Glu89 side chain from outer to inner, thus destabilizing the geometry of the catalytic active site. The system thus evolved toward the final catalytic step, i.e. the exit of the leaving group from the catalytic site. In this system, Glu89 was free to populate the inner and outer conformations, overcoming an energy barrier of ~ 4.5 kcal mol⁻¹, calculated from metadynamics simulations. In addition, we computed the energetics for the full release of the leaving group in the presence or absence of MgC, using confined well-tempered metadynamic simulations.²¹⁷ The energetic

barrier for AMP departure was ~ 16.5 kcal mol⁻¹ in the presence of MgC, and ~ 35 kcal mol⁻¹ in its absence.

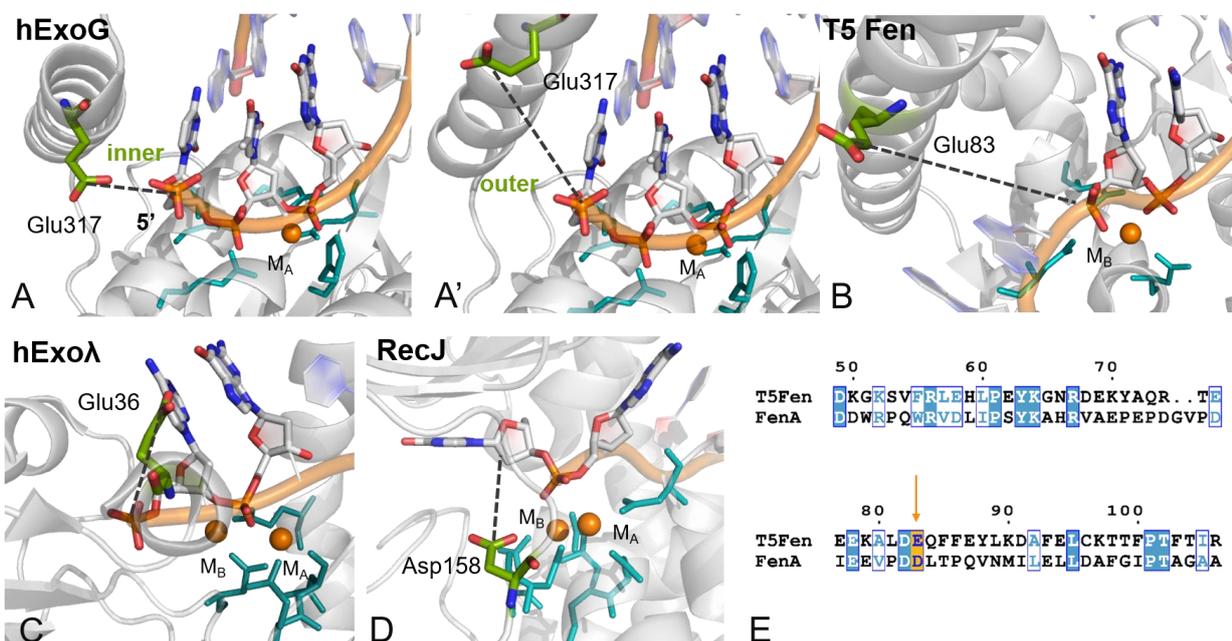


Figure III.7 Close views of the active site of 5' metallonuclease members that possess an analogous acid residue (light green) close to the two-metal-ion center (MA, MB, in orange), the active site (in cyan) and the leaving group (indicated by a dashed line). (A) Human ExoG, in which Glu317 is pointing in the inner (PDB ID 5T5C) and (A') outer conformations (PDB ID 5T40, merged with the DNA substrate from PDB ID 5T5C). (B) Escherichia phage T5Fen (PDB ID 5HNK). (C) Human λ -Exonuclease, (PDB ID 3SM4). (D) *D. radiodurans* RecJ (PDB ID 5F55). (E) Sequence alignment of *M. smegmatis* FenA and *E. phage* T5Fen. The conserved acid residue (Glu/Asp) is indicated in orange.

The Gibbs free energy (ΔG^\ddagger) for the overall catalytic process of hExo1 is 19.6 kcal mol⁻¹, computed using the experimental k_{cat} for hExo1 (see Appendix A for further information).¹⁶⁵ This energy value corresponds fairly well to our estimation of the free-energy barrier for the unbinding process of the leaving group in presence of the third ion, i.e. ~ 16.5 kcal mol⁻¹. The leaving group departure may therefore be rate-limiting for the exonuclease catalytic process in hExo1, as already proposed for other metallonucleases (e.g. FENs, APE1, PvuII, MunI, NaeI, SfiI, EcoRI, EcoRV).^{246–253}

These results suggest a mechanism where Glu89 recruits a third metal ion in the reactant state. Clearly, quantum calculations are needed to evaluate the mechanistic implications of this additional ion for the chemical step of DNA hydrolysis.^{65,254–256} However, from these classical MD simulations, it emerges that the third ion promotes leaving group departure, acting as a shuttle for the exit of the nucleotide monophosphate product from the catalytic

site. Notably, this result is in line with evidence of a third-ion-mediated leaving mechanism for pyrophosphate departure in polymerase enzymes.⁶²

To further test this mechanistic hypothesis and investigate whether this enzymatic strategy is shared by other nucleases, we performed sequence alignments via the Needleman-Wunsch algorithm,²⁵⁷ using 10 different eukaryotic species of Exo1 (see Appendix A for more information). We found that the Glu89 is fully conserved among these enzymes, as for those residues forming the reaction center, and second-shell residues like Lys85, Arg92 and the guide residues Tyr32 and His36 (Figure A.11). This suggests that Glu89 is an integral part of the enzymatic machinery for efficient catalysis in Exo1.

We also performed structural comparisons using recent crystallographic structures of additional nucleases and identified a shared spatial localization in these enzymes of an acidic residue (Glu/Asp), in analogy to Glu89 in hExo1. In such enzymes, in fact, we always identified the presence of a Glu/Asp residue located in a second shell sphere centered on the two-metal-ion active site. This acidic residue is always situated in a solvent accessible position (thus able to recruit ions from the bulk), being strategically located on the side of the expected exit path for leaving group departure, in respect to the catalytic center. For example, *human* ExoG,^{258,259} which is 5' metallo-exonuclease enzyme co-crystallized in complex with the DNA substrate, has a glutamate (Glu317) residue located near the terminal 5' phosphate. Here, Glu317 resides in a solvent-exposed area. Notably, Glu317 can assume different orientations in the available crystals (see PDB ID 5T5C vs PDB ID 5T40),²⁵⁸ which suggests that this glutamate may act as a recruiter of metal ions in the same way as for Glu89 in hExo1 (Figure III.7). Another case is the *human* λ -Exonuclease,^{260,261} where Glu36 is located close to the 5' phosphate of the DNA substrate (Figure III.7). Here, too, it has been hypothesized that a third metal ion may transiently bind close to the two-metal-ion site, likely aiding the leaving group departure.²⁴⁵ A further example is RecJ nuclease,²⁶² where Asp158 is solvent-exposed and close to the active site, in a similar position as Glu89 in hExo1 (Figure III.7).

Then, we looked at the hExo1 family member bacteriophage T5 flap endonuclease (T5 Fen, PDB ID 5HNK).¹⁴⁷ We identified Glu83, which is located along the mobile arch, in analogy to Glu89 in hExo1. Finally, we considered the recent high-resolution X-ray structure of *M. smegmatis* FenA,²²⁷ which is a 5' structure-specific nuclease and close

homologue of phage T5Fen, and its sequence alignment with T5Fen. It suggests that Glu83 in T5Fen may correspond to Asp85 in FenA (Figure III.7). Notably, holo forms of T5Fen and FenA were crystallized in complex with three metal ions in the active site.^{147,227} This further corroborates the idea that multi-metal-ion catalytic sites may be necessary for nucleic-acid processing in these enzymes. Indeed, while the presence of a third metal ion, located in the vicinity of the two-metal-ion active site, is a novel aspect in polymerases and nucleases, the exact position of such additional metal ion in respect to the catalytic center can vary.² For example, T5FEN and FenA enzymes were solved with a third ion in a different relative position, although always in the close proximity of the reactive two-metal-ion centre. It is thus plausible that the third transient ion may play different roles during catalysis, according to its specific location at the catalytic centre.

Another intriguing aspect is the recurring presence of second-shell positively charged residues that surround the metal-aided catalytic site in nucleic-acid-processing enzyme.¹²⁵ During our extended MD simulations, we observed the interaction of Arg93, along the helical arch, with the terminal phosphate of the substrate. We structurally aligned hExo1 (PDB ID 5V0E)⁵⁹ with hFEN1 (PDB ID 5KSE)¹⁵⁸ and we noted that the terminal guanidine, amide, and amine groups of Arg93, Asn124 (in hExo1), and Lys132 (in hFEN1) residues were all within a sphere of ~ 3 Å, and close to the 5' phosphate in the leaving group (Figure A.12). This result further suggests that Arg93, together with Arg96, may have a crucial role in 5' phosphate steering.¹⁵⁸ In this respect, Arg93 is an important point of control in Exo1 poly (ADP-ribose) binding, as proved by in vitro and in vivo assays on natural Arg93Gly mutation.²⁶³ Moreover, a common feature of different nucleic-acid-processing enzymes is a solvent-exposed and positively charged residue that interacts with the negatively charged moiety in the leaving group (e.g. Arg96-5' phosphate).⁶⁶ Interestingly, this Arg96-5' phosphate interaction comes in addition to an already complex architecture characterized by a number of positively charged residues surrounding the active site. Indeed, hExo1 is one of a large set of nucleic-acid-processing enzymes characterized by the recurring presence of positively charged elements in the vicinity of the reactive site.¹²⁵ These elements in hExo1 comprise the catalytic Lys85 and Arg92. These positively charged residues are thought to play a role in the phosphate steering during the threading mechanism. These residues therefore seem crucial for substrate

recognition, binding, catalysis, translocation, and initial product release. However, the Glu89-mediated recruitment and binding of a third ion from the bulk also seems necessary for the full departure of the leaving group from the enzyme's catalytic site in hExo1, and likely in several other nuclease enzymes.

4. Conclusions

Our results provide new insights into the functional role of a third metal ion, which was recently found transiently located at the catalytic site of nuclease enzymes during catalysis. Using molecular dynamics and free-energy simulations applied to multiple systems, we considered the conformational switch of the side chain of a specific residue, Glu89, which is located near the active site in hExo1. We noted that this conformational switch favors the recruitment of a third metal ion from the bulk. The third metal ion is thus promptly positioned near the catalytic center, in accordance with the structural evidence. Our simulations also indicate that this ion serves as an exit shuttle for the leaving group departure from the catalytic site after DNA hydrolysis. The exit mechanism is also favored by the initial involvement of positively charged residues, which are located in an extended and highly structured second-shell area at the two-metal-ion active site.^{125,264,265} Finally, our structural analyses of nuclease enzymes show that such a negatively charged residue (Glu/Asp) is persistently found in a similar, structurally conserved, and strategic position in several other 5' structure-specific nucleases, which seem to share this enzymatic mechanism to promote DNA hydrolysis. These findings may have an implication for de novo enzyme engineering and structure-based drug design.^{266,267}

5. Material and Methods

Structural models. We used six different model systems: i) the wild-type (*wt*) reactant state (RS_{3M}), based on the recent time-resolved X-ray structure of the complex of the hExo1 (PDB ID 5V06), which includes the 5'-recessed end DNA substrate. In this structure, the third metal ion is located close to the catalytic site; ii) the same *wt* reactant state with the third metal ion manually removed (RS_{2M}); iii) the mutated Glu89Ala reactant state (RS_{Glu89Ala}), modelled on the same X-ray structure (PDB ID 5V06); iv) the product state (PS_{2M}), based on the recent time-resolved X-ray structure of the ternary complex of

hExo1(PDB ID 5V0A), characterized by the newly formed 5'-recessed end DNA substrate and the leaving adenosine monophosphate (AMP); v) the product state with the third metal ion close to the 5' phosphate of the AMP (PS_{3M}). This system is modelled on the X-ray structure of the pre-reactive complex, in which we manually cleaved the scissile phosphate; vi) the mutated Glu89Ala product state (PS_{Glu89Ala}), modelled on the same X-ray structure (PDB ID 5V0A).⁵⁹

Classical Molecular Dynamics simulations. To investigate the functional dynamics of the hExo1/DNA complex, we used extensive force-field-based MD simulations, which are highly informative for complex enzyme/nucleic acid assemblies.^{268–272} Here, the AMBER/ff14SB²⁷³ and OL15^{274,275} force fields were used to treat the hExo1 enzyme and the DNA respectively. The terminal 5' thymine monophosphate, the 5' adenosine monophosphate, and the terminal 5' guanidine monophosphate were treated with the general amber force field (GAFF).¹⁸³ The atomic charges were derived by fitting the electrostatic potential according to the Merz-Singh-Kollman scheme,²⁷⁶ the RESP fitting procedure.²⁷⁷ The length of all bonds involving hydrogen atoms was constrained to the equilibrium using the P-LINCS algorithm,²⁷⁸ and a time integration step of 2 fs was used. All simulations were performed using GROMACS 5.1 code.²⁷⁹ Long-range electrostatic interactions were calculated with the particle mesh Ewald method with a Fourier grid spacing 1.6 Å.^{280,281} Periodic boundary conditions in the three directions of Cartesian space were applied. The magnesium ions were treated with a non-bonded approach based on the “atoms in molecules” theory partitioning scheme.^{282,283} The systems were solvated with TIP3P water molecules²⁰⁰ and neutralized adding Mg²⁺, Na⁺, K⁺, and Cl⁻ ions, as indicated in the crystallization procedure.⁵⁹ The total number of atoms was ~60,000 for each system (see Appendix A for more information). We followed a two-step procedure for the MD simulations: first, the equilibration phase in which we followed different procedures depending on the system (see Appendix A). Then, the production phase was carried out in an NPT ensemble, a constant temperature of 310K imposed using the velocity-rescaling thermostat,²⁸⁴ and a constant pressure of 1 bar maintained with a Parrinello-Rahman barostat.²⁸⁵ We collected MD simulations of ~0.7 μs for RS_{3M} and ~1 μs for each of RS_{2M}, RS_{Glu89Ala}, PS_{2M}, PS_{3M}, PS_{Glu89Ala}, for a total of ~6 μs of MD.

Free-energy calculations. We used well-tempered metadynamics²¹⁶ to characterize and estimate the free-energy landscape of Glu89 conformational flexibility. We selected a collective variable (CV) that distinguished between the inner, outer, and intermediate conformations adopted by Glu89 during the MD simulations. Thus, the selected CV was the pseudo dihedral angle ϕ defined by the N, C α , C δ , C γ atoms on Glu89 (see Appendix A). In particular, based on our MD simulations, inner conformations are adopted at $\sim 70^\circ < \phi < \sim 100^\circ$, intermediate conformations at $\sim -40^\circ < \phi < \sim -10^\circ$, and outer conformations at $\sim -150^\circ < \phi < \sim -100^\circ$. We performed well-tempered metadynamics by biasing the CV using an initial hill height of 0.02 kcal mol⁻¹, a hill width of 20°, a fictitious CV temperature of 1550 K, and a deposition rate of 1ps⁻¹. The simulations were conducted until convergence (see Appendix A for more information).

Well-tempered metadynamics was also used to evaluate possible pathways and the semiquantitative energetics for the release of the leaving group (i.e. adeno-sine 5'-monophosphate, AMP) from the active site. We used a confined metadynamics approach,²¹⁷ which excludes regions of the conformational space that are not relevant to the chemical event under investigation. The selected CV was the distance between the center of mass (COM) of the heavy atoms of AMP and the COM of the C α of the aspartates (Asp152, Asp171, Asp173) in the first coordination shell of the two catalytic metal ions (see Appendix A). This CV indicates the degree of departure of AMP from the active site in our MD simulations. We used an initial hill height of 0.29 kcal mol⁻¹, a hill width of 0.6 Å, a fictitious CV temperature of 3720 K, and a deposition rate of 1ps⁻¹. The simulations were conducted until convergence (see Appendix A for more information).

Chapter IV. Molecular mechanism of phosphate steering for DNA binding, cleavage localization, and substrate release in nucleases

Abstract

Structure-specific endonucleases (SSEs) cleave the DNA substrate in a precise position based on the specific DNA 3-dimensional structure. The human flap endonuclease 1 (hFEN1) is a 5' SSE that prevents DNA instability by processing Okazaki fragment 5'-flaps with remarkable efficiency and selectivity using two-metal-ion catalysis. Recent structural and mutagenesis data of hFEN1 suggest that phosphate steering favors specificity and catalysis. Here, we investigate the phosphate steering mechanism at the atomistic level using microsecond-long molecular dynamics and well-tempered metadynamics simulations of wild-type and mutant systems of hFEN1. We show how positively charged second and third-shell residues operate the phosphate steering mechanism to promote catalysis through i) substrate recruitment; ii) precise cleavage localization; and iii) substrate release, thus actively preventing off-target incision of the substrate. Importantly, structural comparisons of hFEN1 and other nuclease enzymes suggest that phosphate steering may also serve the structure-based selection of the specific DNA substrate by other 5' structure-specific nucleases.

1. Introduction

Structure-specific endonucleases (SSEs) are found in all branches of life and are central to processing DNA secondary structures during DNA replication, repair, recombination, and transcription.^{79,82,114,155,220,286,287} SSEs hydrolyze the DNA substrate at a precise position in the strand, according to the specific DNA 3-dimensional structure, rather than its sequence.^{114,286,288–291} This is pivotal to avoid genomic instability and the onset of aberrant enzymatic activity.^{155,231} Indeed, mutations in SSE-encoding genes are involved in many human diseases, suggesting that SSEs are a potential target for drug discovery.^{84,177,292–294}

The human flap endonuclease 1 (hFEN1) is the most experimentally characterized SSE. hFEN1 is a 5' structure-specific two-metal-aided endonuclease that prevents DNA instability.^{153,155,224,287,295,296} This is achieved by processing, with extraordinary efficiency and selectivity, Okazaki fragment 5'-flaps.^{287,297} This allows the completion of lagging strand DNA synthesis during cell proliferation.^{298,299} However, it is unclear how hFEN1 achieves this enzymatic precision.

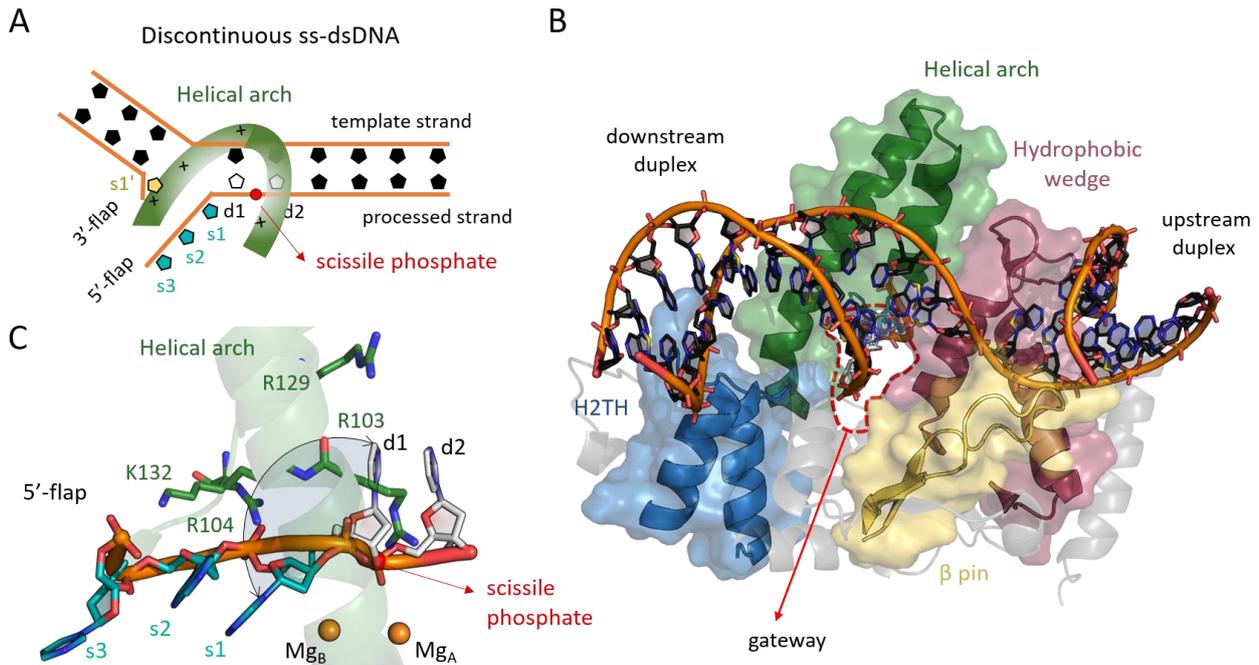


Figure IV.1 A) Schematic representation of the optimal double-flap DNA substrate for hFEN1 catalysis incising the bond between nucleotides d1 and d2 of dsDNA. In addition, a schematic representation of the helical arch (in green) with the positive charges given by the steering residues side chain. B) Overall representation of DNA substrate (in cartoon and licorice) and hFEN1 enzyme (in cartoon and surface, colored by the four main DNA binding sites). The red dashed line highlights the gateway. C) Representation of the inverted orientation of 5'-flap region (nucleotides s1 to s3, in cyan), the two catalytic metal ions (Mg_A and Mg_B , in orange spheres), and the phosphate steering residues of the helical arch (R103, R104, R129, K132 - in licorice, colored in green).

A wealth of structural and biochemical data on hFEN1 has clarified how this enzyme preferentially binds and cleaves double-flap substrates with one nucleotide 3'-flap and a 5'-flap of any length (Figure IV.1).^{153,157,300,301} hFEN1 thus catalyzes a single hydrolytic incision between the first and second nucleotides in the dsDNA (see Figure IV.1).¹⁶⁶ hFEN1 catalytic activity operates through the threading mechanism,^{147,153} which has been reported in other SSE enzymes (e.g. the Exo1 and Exo λ enzymes^{59,162,238,302}). This mechanism involves the 5'-flap region of the substrate being threaded through the

‘gateway’ formed by the helical arch (Figure IV.1). The gateway is wide enough to allow through only a single strand of DNA,¹⁶⁶ thus discriminating a 5’-flap single strand from a 5’ double strand. However, once the single-strand DNA is threaded through the ‘gateway’, it is unclear how hFEN1 selects and precisely cleaves, via the two-metal-ion mechanism, the targeted phosphodiester bond along the substrate strand.^{46,48,240}

Based on recent crystallographic structures and mutagenesis data, researchers have proposed a phosphate steering mechanism for the specificity and catalysis in hFEN1.¹⁵⁸ The proposed mechanism consists in an electrostatically driven steering of the 5’-flap phosphates, which would properly position the scissile bond at the catalytic site. Intriguingly, hFEN1/DNA complex features an inverted orientation of the phosphates within the 5’-flap in precatalytic states. The single-strand flap is rotated (i.e. steered) along the backbone axis, with the phosphates facing away from the active site, when the scissile bond is ~6 Å away from the two-metal-ion reaction center. Four specific positively charged residues positioned along the helical arch at the gateway (i.e. Arg103, Arg104, Arg129, and Lys132 - hereinafter referred to as phosphate steering residues) are proposed to electrostatically control the positioning of the DNA backbone for catalysis (see Figure IV.1). Mutagenesis data demonstrate that these second-shell and third-shell phosphate steering residues affect the catalytic rate, although they are distant from the reaction center.¹⁵⁸ Notably, all the phosphate steering residues are (semi)conserved in the FEN superfamily, further supporting their relevance for nuclease activity.¹⁵⁸ But researchers have not yet clarified the mechanistic action of the steering residues for the exact selection of the targeted phosphate and its correct placement at the catalytic site for hydrolysis. In this regard, the available structural and kinetics data are an excellent basis for investigating the mechanistic steps involved in steering the DNA substrate in hFEN1.

Here, we report microsecond-long force-field-based molecular dynamics (MD), and well-tempered metadynamics simulations to compare the wild-type (*wt*) and multiple mutated systems of hFEN1, considering both the endo- and exonucleolytic substrates in the threaded and post-reactive states. We show that phosphate steering residues, in concert with Arg100, act to favor the formation of a competent Michaelis-Menten complex at the reaction center. We also show that the steering interactions formed with +1 phosphate are critical for the release of leaving group. This explains the experimental kinetic data, which

report a drop of 300-fold in hFEN1 enzymatic activity for the exonucleolytic substrate lacking the 5' phosphate at the +1 position. Finally, based on structural analyses, we expand the proposed mechanistic action of substrate steering to other 5' structure-specific nucleases, underlining the relevance of this mechanism for specific substrate recognition, incision, and final release.

2. Results and Discussion

Phosphate steering residues and Arg100 promote positioning of the inverted 5'-flap orientation and hinder off-target incision. First, we ran extended equilibrium force-field-based molecular dynamics (MD) simulations of the wild-type (*wt*) threaded state (PDB ID 5KSE¹⁵⁸, ~1 μ s of simulation time). This allowed us to investigate the binding process of the double flap DNA substrate to hFEN1 and determine the exact mechanistic action of the four phosphate steering residues (Arg103, Arg104, Arg129, and Lys132). Notably, the threaded state is an intermediate where the scissile phosphate of the double-flap substrate DNA is 6.2 Å away from two-metal-ion active site. This intermediate's *wt* native structure was reconstructed by replacing the non-reactive Sm³⁺ ions at the catalytic site with the native Mg²⁺ ions. Importantly, we also restored the catalytic arginine residue at the Arg100Ala mutation (PDB ID 5KSE¹⁵⁸, see Methods for details) in order to investigate the dynamics of the native enzyme.

The DNA-enzyme complex is stable in the threaded state for the entire simulation, with an average RMSD value for the heavy atoms of 2.6 ± 0.3 Å. At this stage, the single-strand 5'-flap remains in its inverted orientation while inserted into the gateway. The inversion of the single-strand 5'-flap is described well by the α 1d1 angle, formed by the planes of the two nucleobases before and after the +1 phosphate (where the +1 phosphate is the one after the scissile phosphate, along the 5' flap, see Figure IV.2). This angle is $\sim 160^\circ \pm 21^\circ$ during the MD simulations, thus conserving its initial value (X-ray 154°). In this state, the scissile phosphate remains at $\sim 6.5 \pm 0.5$ Å, properly oriented to bind the two catalytic metal ions. However, the 5'-flap phosphate backbone points away from these ions (Figure IV.2). After 230ns, this state is further stabilized by Arg100 through stable interactions of its side chain with both the scissile and +1 phosphates of the substrate. Due to these interactions, the +1 phosphate is pulled ~ 11 Å away from the two metal-ion center (Figure IV.2).

Concomitantly, Arg103 breaks its interaction with the +1 phosphate in the threaded state and moves into its reactant crystallographic position (PDB ID 5UM9). Here, Arg103 stabilizes the nucleobases d1 and d2 that border the scissile phosphate (Figure B.1).

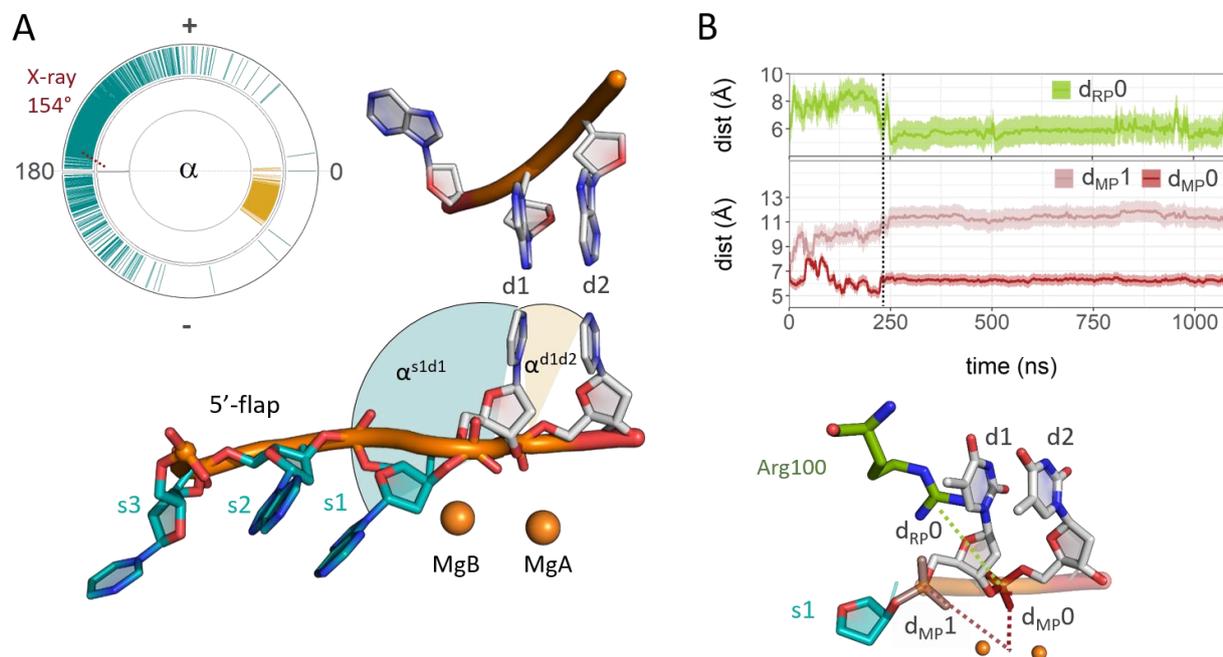


Figure IV.2 A) Distribution of the α^{s1d1} and α^{d1d2} dihedral angles during the MD simulation of the wt system. α^{s1d1} is the angle between the planes of the nucleobases of nucleotides (nt) d1 and s1. α^{d1d2} is the angle between the nucleobases of d1 nt and d2 nt within the 5' strand. The molecular scheme shows nucleotides d2, d1, s1, s2, and s3, the catalytic Mg ions (as orange spheres), and a representation of the dihedral angles α^{s1d1} and α^{d1d2} . B) Time series of distances d_{RP0} , d_{MP0} , and d_{MP1} during the MD simulation of the wt system. The molecular scheme shows nucleotides d2, d1, and s1 together with the +1 and scissile phosphates (in pink and red, respectively), the Arg100 residue (in green licorice), the catalytic Mg ions (as orange spheres), and a representation of the distances d_{RP0} , d_{MP0} , and d_{MP1} .

Interestingly, this dynamic reorganization of Arg100, Arg103, and +1 phosphate in the threaded state corresponds well to the reactant state (PDB ID 5UM9¹⁵⁸, RMSD $\sim 2 \pm 0.3$ Å Figure B.1). This indicates that, during our MD simulations, the 5'-flap phosphate motif of the threaded state freely shifts into a position that closely resembles that of the reactants. Notably, this structural change is only observed when the Arg100 is present and stably interacting with the nearby phosphates. In fact, simulations of the Arg100Ala mutated system maintained well the overall conformation of the threaded state, including those of the 5'-flap phosphate motif (Figure B.2).

Interestingly, the specific interactions of Arg100 with the scissile and +1 phosphates of the substrate are flanked by those of two additional positively charged residues, Arg104

and Lys132, which are located near the 5'-flap region. During our MD simulations, these residues mutually interact with the inverse orientation of the nucleotides of the flap. Thus, through electrostatic and cation- π interactions, Arg100, Arg104, and Lys132 contribute to maintaining the substrate in the correct orientation to bind the catalytic center (Figure B.3). In addition, we observed that the catalytic Tyr40 forms a hydrogen bond with the +1 phosphate. This H-bond is maintained for $\sim 60\%$ of the whole simulation time (see Figure B.4). Notably, this H-bond is present only in the reactant structure (PDB ID 5UM9¹⁵⁸) and not in the threaded structure (PDB ID 5KSE¹⁵⁸). Similarly, Arg129 interacts mainly with nucleotides of the dsDNA region of the substrate, as in the reactant state. This shows that, during our simulations of the threaded state of the native enzyme, the system is prone to spontaneously evolve toward the reactant state. Together, these residues operate as anchor points to precisely shift the scissile phosphate from the threaded state toward the reaction center, thus decreasing the possibility of an off-target incision of the substrate by hFEN1.

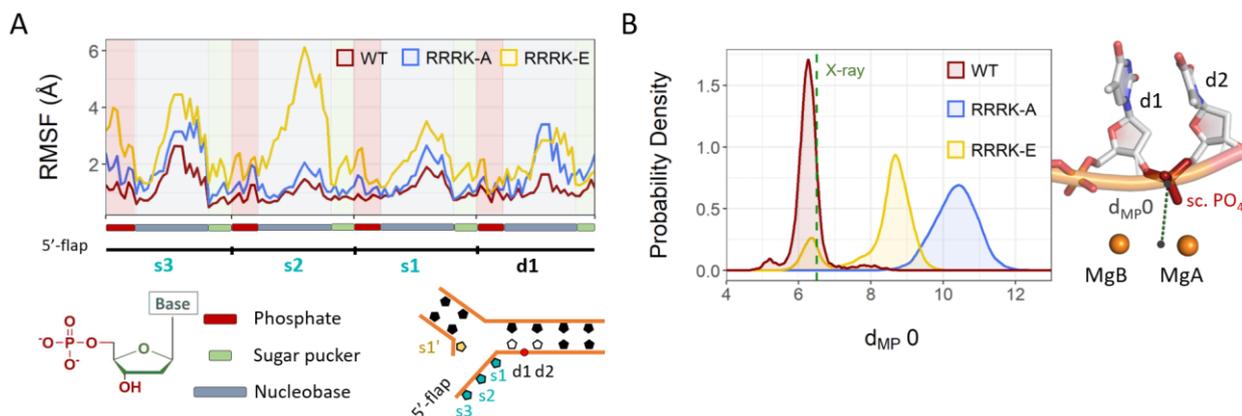


Figure IV.3 A) Root mean squared fluctuations (RMSF) of nucleotides s3 to s1 and d1 within the 5'-strand of DNA (see scheme below the graph) in the wt (dark red), RRRK-A (blue), and RRRK-E (yellow) systems. Atoms belonging to phosphate, sugar, and nucleobase groups are in red, light green, and light blue, respectively (see legend below the graph). B) Left: Probability density of the distance d_{MP0} between the center of mass (C.O.M.) of the scissile phosphate (COM_{PO_4}) and the C.O.M. of the catalytic Mg ions (COM_{Mg}) calculated during MD simulations of the wt (dark red line), RRRK-A (blue), and RRRK-E systems (yellow). Right: schematic representation (snapshot from MD simulations of wt system) of the distance d_{MP0} , the catalytic MgA, MgB ions and d1 and d2 nucleotides.

To further validate the mechanistic action of the four phosphate steering residues, we ran additional MD simulations ($\sim 2 \mu s$ in total) starting from the threaded state (PDB ID 5KSE¹⁵⁸, see Methods for details), where the four phosphate steering residues were

mutated in either alanine or glutamate (i.e. RRRK-A and RRRK-E systems, respectively). Notably, kinetic data show that the reaction rate is 18,000-fold lower for RRRK-E than for *wt* hFEN1.¹⁵⁸

In both RRRK-A/E mutated systems, the overall DNA-enzyme complex stably maintains conformations consistent with the threaded state, with an RMSD mean value for the heavy atoms of $2.7 \pm 0.3 \text{ \AA}$ and $2.6 \pm 0.3 \text{ \AA}$, respectively. Notably, this result reflects the lower effect of these mutations on the DNA-enzyme complexation (vs. the experimental effect on the reaction rate, see above), in line with the experimental K_d values for DNA-enzyme binding (e.g. RRRK-E K_d was 17-fold higher than *wt* hFEN1¹⁵⁸). However, in both RRRK-A/E systems, the 5'-flap region shows increased flexibility compared to the *wt* system (RMSF in Figure IV.3). This is primarily due to the missing interactions with the mutated Arg104 and Lys132 residues, whose side chains directly contact the substrate in the native system (see above). This greater flexibility of the 5'-flap region leads to its spontaneous rotation along the phosphodiester backbone axis to restore a non-inverted orientation of the nucleotides of the 5'-flap (Figure IV.3 and Figure B.5). Arg100 alone is thus not enough to stably maintain the inverted orientation of the 5'-flap. The spontaneous reorientation of the 5'-flap into its non-inverted conformation is even more evident in the RRRK-E system, where each positively charged phosphate steering residue (RRRK) is mutated into a negatively charged glutamate (E). This inversion of the electrostatics of the steering residues boosts the repositioning of the 5'-flap into its non-inverted orientation. Consequently, the distance of the scissile phosphate from the active site in the RRRK-A/E systems becomes $\sim 10.5 \text{ \AA}$ and $\sim 9 \text{ \AA}$, respectively (compared to 6.5 \AA in the *wt* system, Figure IV.3). Notably, the +1 phosphate comes closer to the catalytic center ($\sim 7.5 \text{ \AA}$) than the scissile phosphate. This positional shift of the +1 phosphate, which comes closer to the catalytic center, likely generates the possibility of an off-target incision (Figure B.6).

These data further show the crucial action of the four phosphate steering residues in keeping the 5'-flap region rotated to guide the scissile bond into the correct place for catalysis. Indeed, mutations of the steering residues highlight their necessary action in stabilizing and maintaining the inverted 5'-flap orientation. In this regard, we calculated the electrostatic interactions between the DNA substrate and hFEN1 using a Debye-Huckel

approximation.³⁰³⁻³⁰⁵ The resulting energy values show a destabilization for both the RRRK-A/E mutant systems ($\sim +8$ kcal mol⁻¹ and $\sim +13$ kcal mol⁻¹ for RRRK-A/E, respectively, as compared to the *wt* system - see Figure B.7). Overall, these results support the functional contribution of the steering residues in substrate positioning and stability at the catalytic site.

Kinetic data have shown decreased catalytic activity of hFEN1 when the substrate lacks the 5'-flap. Remarkably, the activity is 300-fold lower when the substrate ends with a 5'-OH (i.e. lacking the 5'-phosphate at the +1 position), but only 7-fold lower when the substrate ends with the 5'-monophosphate group.¹⁵⁸ To clarify the mechanistic reasons for this difference, we ran additional MD simulations (~ 2 μ s in total) of two systems starting from the threaded state (PDB ID 5KSE¹⁵⁸) in which the last nucleotide at the 5'-end of the DNA substrate was either a 5'-monophosphate or a 5'-OH (hereafter referred to as 5PO and 5OH systems, respectively; see Figure B.8).

In the 5PO and 5OH systems, the overall DNA-enzyme complex is stable with an average RMSD value for the heavy atoms of 3.1 ± 0.4 Å. However, during simulations of the 5OH system, the 5' terminal nucleotide alone shows considerable flexibility over time, as indicated by its average RMSD value for the heavy atoms of 2.4 ± 0.8 Å. However, in the 5PO system, this nucleotide is stable for the whole simulation, with an average RMSD value of 1.4 ± 0.4 Å. The higher stability of the 5' terminal nucleotide in the 5PO system is due to specific interactions with the Arg100, Arg104, and Lys132 side chains, which interact with the substrate even in the absence of the 5'-flap. Indeed, it is only in the 5PO system that Arg104 and Lys132 can form specific salt bridge interactions with the terminal phosphate (i.e. the +1 phosphate), contributing to stabilizing the terminal 5' nucleotide (Figure B.8). Because of the lack of the phosphate group, these interactions are impossible in the 5OH system. Thus, these results suggest that the 5'-phosphate at the +1 position is crucial to keeping the substrate stably oriented toward the catalytic ions, which would explain the kinetic data.¹⁵⁸

Taken together, these equilibrium MD simulations of multiple *wt* and mutated forms of hFEN1, in complex with different substrates, show that the *wt* threaded state can spontaneously evolve toward the reactant state. That is, the insertion of Arg100 in the initial mutated crystal (PDB ID 5KSE¹⁵⁸) triggers a partial structural evolution of the starting

threaded system, which shifts toward the reactant state. In consideration of the limited configurational sampling of plain MD simulations, this structural evolution is only partial, yet marked. The system is clearly dynamically oriented to position the scissile phosphate precisely at the two-metal-ion center for hydrolysis, avoiding off-target incision. Thus, our simulations show that hFEN1 catalysis operates through a cooperative mechanism involving second-shell (i.e. Arg100, Arg103, Arg104) and third-shell (i.e. Arg128 and Lys132) positively charged residues, which are crucial for substrate recognition, precise incision site selection, and ultimately for the pre-organization of the Michaelis-Menten complex.

Energetics of Michaelis-Menten complex formation and leaving group departure via metadynamics simulations. To evaluate the energetics related to the formation of the Michaelis-Menten complex for hFEN1 catalysis, we used multiple enhanced sampling path metadynamics (pMtD) simulations of the *wt* in complex with different DNA substrates (i.e. the 5'-flap, 5PO, and 5OH), as well as pMtD of the mutated RRRK-A/-E systems. The reference path connects linearly the threaded (T) and the reactant (R) states (PDB ID 5KSE and PDB ID 5UM9, respectively¹⁵⁸) and consists by nine equally distanced points defining the formation of a competent reactant state. Then, two collective variables (CVs) were used in pMtD to measure: i) the progress of the system along the reference path (S) and ii) the distance of the sampled configurations from the path (Z) - see Methods for details.

In the free energy surface (FES) of the *wt* system, the initial T state is a relative minimum that is connected to an intermediate (I) through a barrier of ~ 6 kcal mol⁻¹. However, the I state is at -10 kcal mol⁻¹ compared to T, which explains the evidence from equilibrium MD of the *wt* system. That is, the sole reinsertion into T of the catalytic Arg100 (mutated into an alanine in the original structure PDB ID 5KSE¹⁵⁸) promoted the partial and spontaneous T→I transition (see above). In this case, enhanced sampling pMtD simulations show the full T→I transition with the 5'-flap in the inverted orientation, and with Arg104 and Lys132 side chains interacting with the phosphodiester backbone. Additionally, the +1 phosphate interacts through H-bond with the Tyr40 side chain. This residue in turn spontaneously rotates its side chain, as shown by the dihedral angle δ (defined by the C- α -C β -C γ atoms of Tyr40, see Figure B.9). This dihedral changes from $\sim 62^\circ \pm 12^\circ$ in T to

$\sim 150^\circ \pm 20^\circ$ in I. The latter is also the value of δ in R (see Figure B.9), where it allows a π - π stacking interaction formed by the phenyl ring of Tyr40 and the d1 nucleotide.

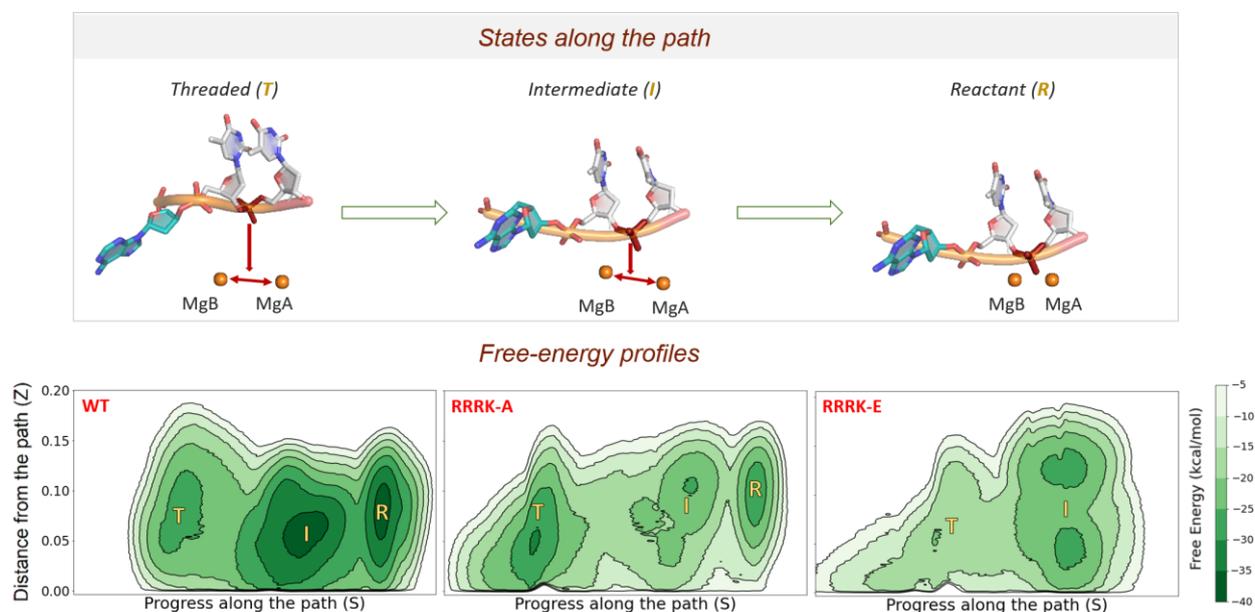


Figure IV.4 On top, structural representation of the active site for the threaded (T), intermediate (I), and reactant (R) states identified by the pMtD simulations of the wt system. On bottom, pMtD free energy landscape of the wt, RKKK-A, and RKKK-E systems, where the threaded (T), intermediate (I), and reactant (R) states are indicated.

However, in I, the scissile phosphate is still too far (i.e. $\sim 5 \text{ \AA}$, see Figure B.9) from the active site, which in turn is not (yet) properly assembled for catalysis. The two catalytic metal ions are $\sim 4.6 \pm 0.4 \text{ \AA}$ from each other (versus an optimal MgA-MgB internuclear distance for catalysis of $\sim 3.6 \text{ \AA}$). Indeed, from I, the system must overcome a barrier of $\sim 13 \text{ kcal mol}^{-1}$ to reach R, where the overall catalytic system is correctly assembled for catalysis. Notably, this computed activation barrier is compatible with the experimental kinetics ($k_{\text{cat}} = 14.8 \pm 2.1 \text{ min}^{-1} \rightarrow \Delta G^\ddagger = 16.6 \text{ kcal mol}^{-1}$ calculated using the Eyring–Polanyi equation).¹⁶⁵ This barrier also explains why our equilibrium MD simulations were unable to sample this energetically costly I \rightarrow R transition. Indeed, during I \rightarrow R transition, the first and second coordination shell of MgA and MgB (Asp34, Asp86, Glu160, Asp179, Asp181, and 5 water molecules) undergo a significant structural rearrangement to allow the proper coordination of the oxygen atoms of the scissile phosphate to the catalytic metal ions. Ultimately, the I \rightarrow R transition brings the system to the absolute minimum of the FES, at about -3 kcal mol^{-1} relative to the I state (Figure IV.4). Here, the internuclear distance

of the catalytic ions is $\sim 3.6 \text{ \AA}$, which is optimal to accommodate the substrate at the reaction center for hydrolysis.

The FES of both RRRK-A/-E systems show significant differences compared to the *wt* FES. In the RRRK-A system, the T state is the absolute minimum of the FES. From T, the system must first overcome an energy barrier of $\sim 16 \text{ kcal mol}^{-1}$ to reach the I state, which is $+5 \text{ kcal mol}^{-1}$ higher in energy. Pushing the system through this costly T \rightarrow I transition, the 5'-flap rotates to assume the non-inverted orientation, with the +1 phosphate remaining close to the catalytic active site. Due to this, the Tyr40 side chain cannot form any interaction with the +1 phosphate, as is found in the *wt* system (see above). Once in I, the system must overcome a second energy barrier of $\sim 10 \text{ kcal mol}^{-1}$ to reach the R state, which is $+2 \text{ kcal mol}^{-1}$ higher in energy than T (Figure IV.4). Thus, the complete T \rightarrow R transition of the RRRK-A system leads to a state in which the catalytic site is neither properly formed nor energetically favored. In the RRRK-E system, we found a broad minimum I that traps the system. In this minimum, the 5'-flap dynamically assumes a non-inverted orientation, disfavoring phosphate steering. The scissile phosphate is $\sim 4 \pm 0.4 \text{ \AA}$ from the reaction center, and the overall active site architecture is not properly assembled for catalysis. In particular, the two catalytic metal ions have an internuclear distances of $\sim 4.8 \pm 1 \text{ \AA}$. These results indicate the inability of the RRRK-E system to form the Michaelis-Menten complex, which would explain the substantially lower (18,000-fold) activity compared to the *wt* hFEN1.

A further intriguing question is related to the large difference in hFEN1 activity for substrates that differ only in the presence/absence of the +1 phosphate. hFEN1 activity is 300-fold lower without the +1 phosphate in the substrate, but only 7-fold lower with the +1 phosphate in the substrate. Thus, we performed additional pMtD simulations of the *wt* in complex with the substrate, in the presence or absence of the +1 phosphate, namely the 5PO or 5OH substrate. These simulations aimed to explore the mechanistic reasons for this huge difference in hFEN1 activity. As for the *wt* and mutant systems, the path connects T and R states. Mechanistically, we found that, during the T \rightarrow R transition in 5PO, the residues Arg104 and Lys132 and Tyr40 stably interact with the +1 phosphate, guiding the substrate towards the proper position in the active site. These specific interactions are missing in the dynamics of the 5OH system, in which Arg104 and Lys132 are more

solvent-exposed and flexible. Nonetheless, these interactions do not affect the overall energetics for the T→R transition (Figure B.10). In fact, the FES of both the 5PO and 5OH systems are similar. They show a deep minimum at the initial T state that is connected to R through a barrier of ~ 20 kcal mol⁻¹, thus completing the T→R transition. Hence, in our simulations, the presence/absence of the +1 phosphate in the substrate does not severely affect the formation of R, i.e. the binding of the substrate in 5PO and 5OH is equally plausible, which is in line with the experimental K_d (5OH is bound ~ 20 -fold more weakly than 5PO, while its turnover rate is ~ 300 -fold more slowly than 5PO¹⁵⁸).

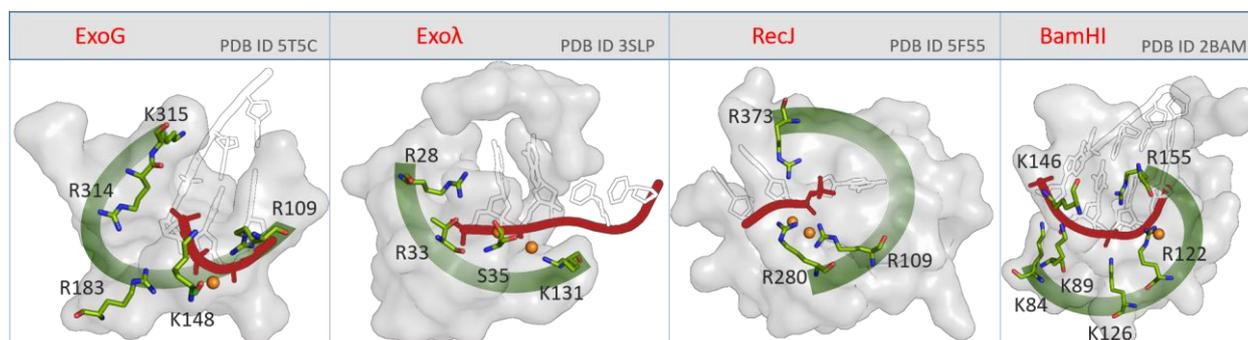


Figure IV.5 Close views of the active site of 5' nucleases members that possess an analogous second/third shell of positively charged cleft (light green) close to the 5'-end (red) and the two-metal-ion center (MgA, MgB, in orange). Human ExoG, with second-shell and third-shell positively charged residues (Arg109, Lys148, Arg183, Arg314, and Lys315) in licorice (PDB ID 5T5C). Human λ-Exonuclease, with residues suggested to act for the electrostatic ratchet of the substrate (Arg28, Thr33, and Ser35) and the catalytic Lys131 in licorice (PDB ID 3SM4, 3SLP for Lys131). *D. radiodurans* RecJ, with the second-shell and third-shell arginine residues (i.e. Arg109, Arg280 and Arg373) in licorice (PDB ID 5F55). BamHI, with the second-shell and third-shell residues in licorice (PDB ID 2BAM).

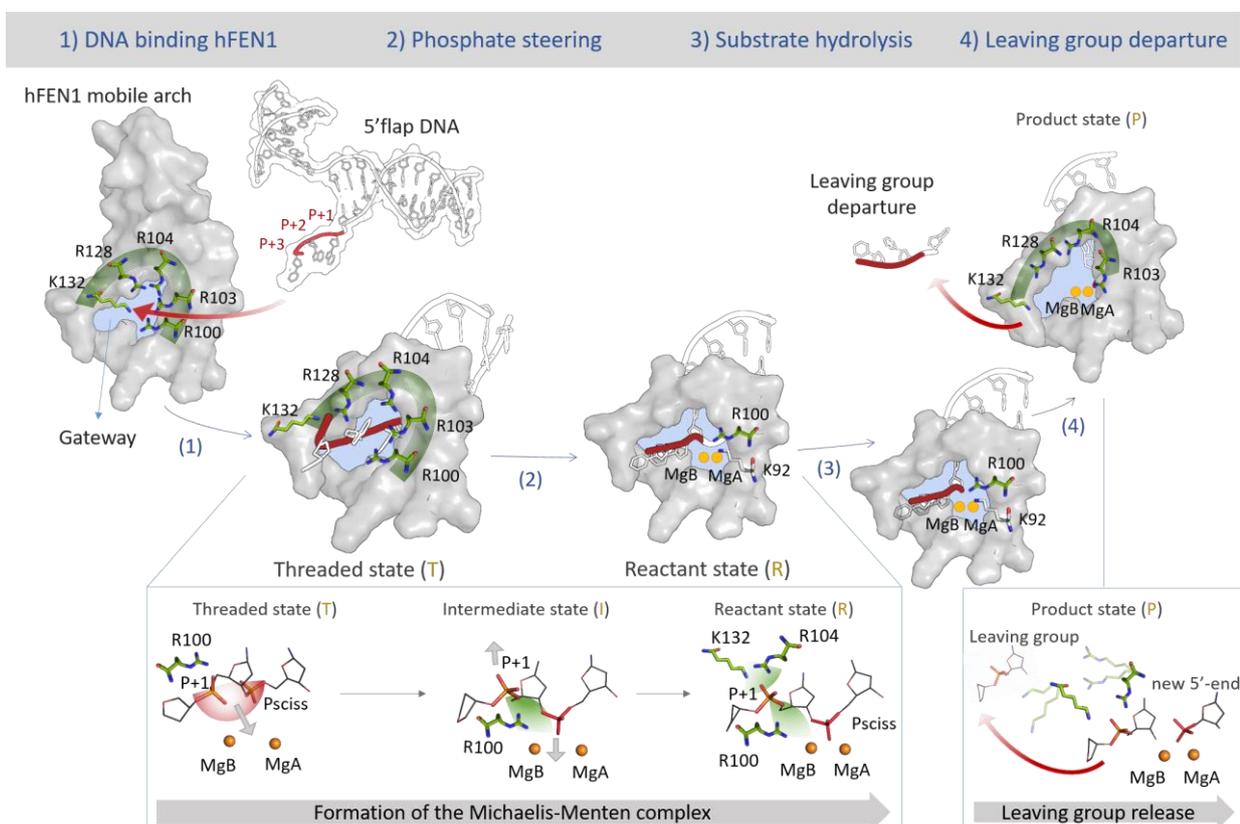
Based on this evidence, we turned our attention to the effect of the +1 phosphate on the release of the leaving group from hFEN1 active site. We ran confined well-tempered metadynamics (see Methods section) and compared three different leaving groups, namely: i) the thymine monophosphate (TMP) nucleotide as the reaction product for the 5PO system; ii) the 5'OH thymine nucleotide (TOH), as the reaction product for the 5OH system; and iii) the 5'-flap (see Appendix B for further information). Two CVs described the departure of the leaving TMP/TOH/5'-flap group from the active site: CV1, defined as the distance between the C.O.M. of the C α of Glu160, Asp179, Asp181 in the first coordination shell of the two-metal-ion center and the C.O.M. of the heavy atoms of TMP/TOH/nucleotide along the flap coordinating the catalytic MgB; CV2, defined as the

distance between the C.O.M. of the heavy atoms of the two nucleobases involved in the base pair of the leaving group (Figure B.11).

As expected, we found that the leaving group departure is favored with the 5'-flap, which overcomes a barrier of ~ 15 kcal mol⁻¹. The exit path is characterized by 2 transitory metastable states (Figure B.11) that lead the leaving flap motif to the bulk solution. Interestingly, Arg104 and Lys132 steering residues maintained their interactions with the flap product throughout the simulations, assisting the full release of the leaving group. This exit pathway and the overall mechanism are conserved with the TMP, where the terminal phosphate interacts with the surrounding steering residues Arg104 and Lys132. In this case, the energy for leaving group departure is ~ 20 kcal mol⁻¹, thus a bit higher than that computed with the 5'-flap. Through the Eyring–Polanyi equation, this computed barrier is in line with the experimental k_{cat} value for this particular system ($k_{\text{cat}}=0.024 \pm 0.003$ min⁻¹ $\rightarrow \Delta G=21.6$ kcal mol⁻¹).¹⁶⁵ Intriguingly, this data shows that the flap region is thus beneficial for leaving group departure. Notably, this is even more evident with the substrate that misses the last phosphate, which has to overcome an even higher barrier of ~ 30 kcal mol⁻¹. That is, the flap region is readily expelled despite its large size, compared to the shorter leaving groups examined here. Together, these free energy calculations, although semiquantitative, provide a possible explanation for the mechanism adopted by hFEN1 to specifically select the correct substrate. In particular, the phosphate steering residues play a crucial role in the Michaelis-Menten complex formation in the presence of the natural substrate, which preferentially contains a long 5'-flap accommodated through the narrow gateway. However, for the shorter substrate without the 5'-flap, the steering residues anchor the +1 phosphate to assist the release of the leaving group through the gateway.

Extension to other 5' nucleases. Based on our findings, we investigated whether other nucleases exploit this enzymatic strategy for phosphate steering. A sequence alignment among hFEN1 family members (i.e. hExo1, hGEN1, hXPG) already showed that those specific basic residues are (semi)conserved.¹⁵⁸ To further analyze this strategic location of basic residues (Arg/Lys) in other metalloenzymes,^{66,125,306} we performed structural comparisons of hFEN1 and additional nucleases. This allowed us to identify basic residues similarly localized in the vicinity of the catalytic site, as for the steering residues in hFEN1 (Figure IV.5). For example, the human ExoG45, which is 5' metallo-exonuclease enzyme

co-crystallized in complex with the DNA substrate, has two lysine and two arginine residues (i.e. Lys148, Arg183, Arg314, Lys315) located near the terminal 5' phosphate (Figure IV.5). Notably, an additional arginine (Arg109) is positioned close to the catalytic active site, as is Arg100 in hFEN1. Likewise, an electrostatic-driven binding force was hypothesized for Exo- λ ,^{260,307} where the substrate is threaded towards the active site through the attraction of the terminal 5' phosphate of the DNA substrate to a positively charged pocket. This pocket is formed by Arg28, Thr33, and Ser35 residues, which interact through salt bridge and H-bond with the terminal 5' phosphate (Figure IV.5). Similarly, three arginine residues (i.e. Arg109, Arg280, Arg373) are found around the phosphodiester backbone in RecJ in complex with the DNA substrate (Figure IV.5).^{262,308} Notably, the mutation of both Arg280 and Arg373 leads to the inactivation of nuclease activity.^{309,310}



Scheme IV.1 Schematic representation of the steps during hFEN1 catalysis. Bottom left box: the molecular mechanism of phosphate steering residues Arg104 and Lys132 acting concertedly with Arg100 to promote the Michaelis-Menten complex formation. Bottom right box: the molecular mechanism for leaving group departure assisted by phosphate steering residues.

In addition, in BamHI³¹¹ three lysine residues (i.e. Lys84, Lys89, Lys146 – see Figure IV.5), together with Arg155, are located close to the 5' end of the substrate. Moreover, similarly to the hFEN1 catalytic residues Lys92 and Arg100, in BamHI we note that

Arg122 and Lys126 are close to the scissile phosphate. Altogether, these structural observations further corroborate the hypothesis that these numerous second and third-shell positively charged residues enable phosphate steering, a mechanism that promotes the specific nuclease enzymatic activity.

Interestingly, this structural evidence from other nucleases nicely complements recent MD simulations of hExo1 in complex with a 5PO substrate, where we detailed the crucial interaction of two basic residues (Arg93, Arg96) with the +1 phosphate.³¹² Those interactions promoted the stabilization of the DNA substrate in the precatalytic state and the partial release of the leaving group in the products. According to sequence alignment,¹⁵⁸ Arg96 in hExo1 was suggested as the counterpart of Arg104 in hFEN1. The putative steering residues in hExo1 are Arg95, Arg96, Arg121, and Asn124 (corresponding to Arg103, Arg104, Arg128, and Lys132 in hFEN1, respectively). In addition, our MD simulations highlighted the role of Arg93, which acts with Arg96 for 5' phosphate steering in hExo1.³¹² Notably, hExo1 is primarily a 5' exo-nuclease³¹³; thus, it preferentially binds and processes substrates without the 5' flap. In this regard, the exit of the short 5PO substrate in hExo1 is further promoted by a specific and transient third magnesium ion,^{151,312} which we did not observe in hFEN1. This may be due to the presence of the long 5'-flap being the leaving group in hFEN1. Overall, these results suggest that these 5' nucleases also select and bind the optimal substrate thanks to strategically localized second-shell and third-shell positively-charged residues,¹²⁵ which enable the functional phosphate steering of the substrate.

3. Conclusions

In summary, we have elucidated the molecular mechanism of action of phosphate steering residues during hFEN1 catalysis (Scheme IV.1 and Scheme B.1). We defined how the phosphate steering residues act dynamically during the entire catalytic process of hFEN1. During the binding of 5'-flap DNA substrate, the phosphate steering residues favor the threading of the 5'-flap through the narrow gateway towards the solvent-exposed side. The strategically located positively charged side chains of these residues establish contacts with the phosphodiester backbone at the beginning of the binding process. Consequently, the substrate is specifically placed with the scissile phosphate oriented to bind the two

catalytic metal ions. Once there, Arg100 establishes salt bridge interactions with both +1 and the scissile phosphates, thus correctly positioning the substrate for catalysis. Importantly, this may reduce the possibility of off-target incisions. Intriguingly, phosphate steering may play a role for the mechanical stress of the substrate to favor precise and efficient bond breaking at the catalytic site, somehow mimicking what has been demonstrated for polymer hydrolysis in solution.^{314,315} This mechanistic hypothesis, however, remains to be explored.

When the Michaelis-Menten complex is formed, the reaction for DNA cleavage can take place. Modelling this chemical step will clearly require a quantum chemical approach, such as hybrid quantum mechanics/molecular mechanics (QM/MM) calculations to quantify the energetics of bond breaking and forming, and evaluate the mechanistic implications for catalysis.^{267,316} However, from our enhanced MD simulations, we realize that after catalysis, the release of the leaving group from the active site is favored by the solvent-exposition of the 5'-flap leaving group, combined with the steering residues' dragging action (Scheme IV.1). In addition, when hFEN1 hydrolyzes a shorter substrate lacking the 5' flap, after substrate hydrolysis, the steering residues are equally critical, promoting the exit of the leaving group, which has to cross the gateway for departure from the catalytic site. This occurs thanks to key interactions with the +1 phosphate (Scheme B.1), which also explains hFEN1's dramatically lower catalytic activity for substrates lacking the +1 phosphate¹⁵⁸ (and so incapable of forming this crucial interaction for leaving group departure).

In hFEN1, the mechanistic action of the phosphate steering residues is therefore critical for substrate binding and correct site incision, eventually prompting leaving group departure. Finally, our structural analyses suggest that this catalysis-enabling mechanism is conserved and used by other metallonucleases, with implications for enzyme design and drug discovery.²⁶⁶

4. Material and Methods

Structural models. We used eight different model systems for the MD simulations: (i) the wild-type (*wt*) threaded state, based on the recent X-ray structure of the DNA-hFEN1 complex (PDB ID 5KSE¹⁵⁸), which includes the 5'-flap DNA substrate; (ii) the mutated

RRRK-A and (iii) RRRK-E threaded state (RRRK-A/-E, respectively), modeled on the wt X-ray structure (PDB ID 5KSE¹⁵⁸); the *wt* hFEN1 bound to the exonucleolytic DNA substrate in the threaded state in which the last nucleotide on 5'-end of the DNA was either (iv) a 5'-monophosphate (5PO) or (v) a 5'-OH (5OH) modeled on the X-ray structure (PDB ID 5KSE¹⁵⁸); the *wt* product state of either (vi) the 5'-flap substrate, (vii) the 5'-monophosphate or (viii) the 5'-OH (the product systems are modeled on the X-ray structure of the pre-reactive complex - PDB ID 5UM9¹⁵⁸, in which we manually cleaved the scissile phosphate). The most probable protonation state of titratable residues in *wt* and mutated enzymes was established through pKa calculations based on the Poisson-Boltzmann equation (using the DelPhiPka web server³¹⁷). Each system was solvated with a 12-Å layer of TIP3P water molecules,^{200,318} and the ion concentration was set to the concentration used for crystallization. The final models include a total of ~70,000 atoms.

MD simulation set-up. We used extensive force-field-based MD simulations, which are highly informative for complex enzyme/nucleic acid assemblies.^{268,269,271,319,320} Here, the AMBER-ff14SB²⁷³ and OL15^{274,275} force fields were used to treat the hFEN1 enzyme and the DNA, respectively. The terminal 5'-monophosphate group of the nucleotide was parametrized with the general Amber force field (GAFF)¹⁸³, and the atomic charges were derived using the RESP procedure, according to the Merz–Singh–Kollman scheme.^{276,277} Monovalent metal ions (i.e. Na⁺, K⁺ and Cl⁻) were treated using the Joung–Cheatham parameters.³²¹ Divalent metal ions (i.e. Mg²⁺) were parametrized according to Li et al.³²² A time integration step of 2 fs was used and the lengths of all bonds involving hydrogen atoms were constrained using the P-LINCS algorithm.²⁷⁸ A velocity-rescaling thermostat was used to set a system temperature of 310K,²⁸⁴ while the Parrinello–Rahman barostat maintained a constant pressure of 1bar.²⁸⁵ Long-range electrostatic interactions were calculated with the particle mesh Ewald (PME) method using a Fourier grid spacing of 1.6 Å. Periodic boundary conditions in the three directions of Cartesian space were applied. All MD simulations were performed with Gromacs2020.2.²⁷⁹ The systems were all subject to the same MD simulations procedure. First, we carried out energy minimization to relax the water molecule and the ions. Here, the catalytic ions (i.e. MgA, MgB) together with both hFEN1 and DNA backbones were kept fixed with harmonic positional restraints of 1000 kcal mol⁻¹ Å². Then, the systems were heated up from 0 to 310 K with an NVT

simulation for a total of ~ 1.5 ns with the same positional restraints used in the energy minimization. Additionally, ~ 1 ns of simulation in NPT ensemble was performed with $100 \text{ kcal mol}^{-1} \text{ \AA}^2$ restraints on the same atoms. Finally, a production run were performed in the NPT ensemble for each system. We collected overall $\sim 5.5 \mu\text{s}$ of MD trajectories, specifically: (i) $\sim 1.1 \mu\text{s}$ for the *wt* threaded system; (ii) $\sim 1.1 \mu\text{s}$ for each of the RRRK-A/-E mutated systems; (iii) $\sim 1 \mu\text{s}$ for each of the exonucleolytic DNA substrate-hFEN1 complexes (i.e. 5PO and 5OH); (iv) ~ 100 ns for each of the cleaved state.

Free energy calculations. Five metadynamics simulations were performed to characterize the formation of the Michaelis-Menten complex in the *wt*, RRRK-A/-E, 5PO and 5OH systems. We used the PATH³²³ method to compute the progress of the process. The reference path connecting the threaded state and the reactant state is defined by a total of nine equally distanced points in the space defined by two key distances: i) the distance between the C.O.M. of the scissile phosphate and the C.O.M. of the catalytic Mg ions (i.e. d_{MP0}) and ii) the distance between the catalytic MgA, MgB (i.e. d_2). Each step of the path is equally spaced with $\Delta d_{\text{MP0}} \sim 0.2 \text{ \AA}$ and $\Delta d_2 \sim 0.1 \text{ \AA}$. The threaded structure is identified by $d_{\text{MP0}} \sim 6 \text{ \AA}$ and $d_2 \sim 5 \text{ \AA}$, and the reactants structure is identified by $d_{\text{MP0}} \sim 2.8 \text{ \AA}$ and $d_2 \sim 3.8 \text{ \AA}$. As per Branduardi et al.,²¹⁸ two-path collective variables (pCVs) are used (S and Z), which define the progress along the reference path and the distance from the reference path, respectively. To sample the free energy landscape, we used well-tempered metadynamics simulations.²¹⁶ An upper limit for d_2 distance was set to 4.5 \AA , together with upper and lower limits for d_{MP0} distance set to 6 \AA and 2 \AA , respectively. The height of the Gaussian was set to 1.2 kJ mol^{-1} , while the width of the Gaussian was set to 0.1 on S and 0.001 on Z. A fictitious CV temperature of 3720 K and a deposition rate of 1 ps^{-1} was used. We collected ~ 500 ns for the *wt* system and ~ 230 ns for all other systems (i.e. RRRK-A/-E, 5PO, 5OH).

To characterize the release of the leaving group from the active site, we then performed well-tempered metadynamics using a confined procedure,²¹⁷ which excludes regions of the conformational space that are not relevant to the chemical event under investigation.³¹² Here, two collective variables were used: i) the distance between the C.O.M. of the heavy atoms of the leaving group (i.e. TMP/TOH/3'-end nucleotide along the flap coordinating the catalytic MgB) and the C.O.M. of the $\text{C}\alpha$ of Glu160, Asp179, Asp181 in the first

coordination shell of the two catalytic metal ions; ii) the distance between the C.O.M. of the heavy atoms of the two nucleobases involved in the base pair of the leaving group. We used an initial hill height of 1.2 kJ mol^{-1} , a hill width of 0.5 \AA , a fictitious CV temperature of 3720 K , and a deposition rate of 1 ps^{-1} . The simulations were conducted until convergence.

Chapter V. Quercetin and luteolin are single-digit micromolar inhibitors of the SARS-CoV-2 RNA-dependent RNA polymerase

Abstract

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has rapidly become a global health pandemic. Among the viral proteins, RNA-dependent RNA polymerase (RdRp) is responsible for viral genome replication and has emerged as one of the most promising targets for pharmacological intervention against SARS-CoV-2. To this end, we experimentally tested luteolin and quercetin for their ability to inhibit the RdRp enzyme. These two compounds are ancestors of flavonoid natural compounds known for a variety of basal pharmacological activities. Luteolin and quercetin returned a single-digit IC_{50} of 4.6 μ M and 6.9 μ M, respectively. Then, through dynamic docking simulations, we identified possible binding modes of these compounds to a recently published cryo-EM structure of RdRp. Collectively, these data indicate that these two compounds are a valid starting point for further optimization and development of a new class of RdRp inhibitors to treat SARS-CoV-2 and potentially other viral infections.

1. Introduction

The COVID-19 pandemic, caused by the emerging new severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), is having a tragic impact on humans and also affecting our economy. Thanks to an unprecedented and extensive collaboration between academia, biotech companies, and governments, vaccines have been discovered to combat and contain this pandemic. Despite the vaccination programs, SARS-CoV-2 continues to be a human threat worldwide. In addition, the emergence of virus variants is an additional threat in relation to the spread of COVID-19. It is likely that COVID-19 will remain an endemic disease.³²⁴ Therefore, small molecule drugs to treat SARS-CoV-2 infections are an additional weapon to fight SARS-CoV-2.

The publication of the viral genome sequence revealed that the SARS-CoV-2 genome is closely related to the earlier SARS-CoV (more than 80% sequence identity) and, to a lesser

extent, to MERS-CoV viruses.³²⁵ This information has triggered the identification of druggable targets based on what was already known for SARS-CoV and MERS-CoV. In particular, the spike protein, 3-chymotrypsin-like protease (M^{pro}), papain-like cysteine protease (PL^{pro}), and the RNA-dependent RNA polymerase (RdRp) have emerged as potential targets for drug discovery campaigns owing to their crucial role in viral entry and host-cell invasion.^{326–328} Specifically, the spike protein recognizes the host receptor, facilitating fusion between the viral envelope and the host cell membrane.³²⁹ The protease M^{pro} catalyzes the proteolysis of polyproteins translated from the viral genome. The RdRp enzyme is responsible for the replication of RNA from an RNA template.^{2,3} Therefore, RdRp is a nonstructural protein that plays a crucial role in the virus life cycle, acting during the viral replication and transcription processes.^{326,327,330,331} Additionally, the absence of a human RdRp counterpart and the high similarity of RdRp within different RNA viruses make this enzyme an attractive target for drug repurposing and development of drugs for COVID-19 and potentially other viral infections.^{327,331,332}

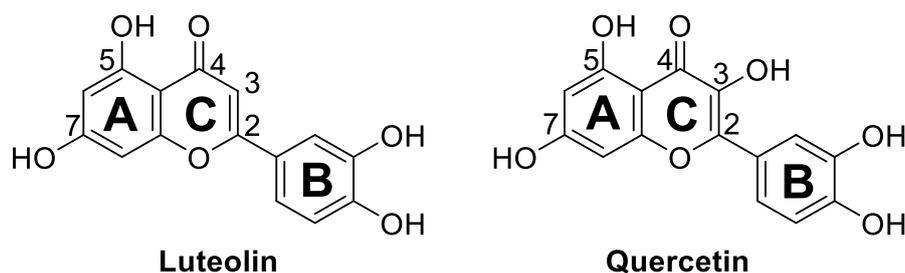


Figure V.1 Chemical structures of luteolin and quercetin.

Given RdRp's essential role, a wide array of approved nucleoside and nucleotide analogs have been considered for repurposing as candidates to block RdRp of SARS-CoV-2.^{333–336} Among them, remdesivir and favipiravir have reached clinical trials. But despite the promising inhibitory effects of remdesivir and favipiravir, with EC_{50} values of 0.77 μ M and 61.88 μ M, respectively,³³⁷ clinical trials showed adverse effects and no statistically significant benefits for hospitalized patients.³³⁸ More recently, another nucleoside analog, molnupiravir, has entered clinical trials. Molnupiravir is an orally available and efficacious ribonucleoside analog inhibitor of influenza viruses and, similarly to remdesivir, it has been repurposed against SARS-CoV-2.³³⁹ However, the RdRp complex of coronavirus can excise erroneous mutagenic nucleotides incorporated into viral RNA, thus creating

resistance to nucleotide analog drugs.^{340,341} Consequently, non-nucleoside inhibitors could hamper the development of resistance.

In this context, natural products are another source of active compounds with promising antiviral activity. These compounds may serve as a starting point for the development of newer molecular entities with greater efficacy and affinity, and with fewer side effects.³⁴² Among them, luteolin and quercetin, which are two ancestors of flavonoid compounds, are known for having a range of basal pharmacological activities, including antiviral properties against picornavirus (RNA virus) and DNA viruses, such as hepatitis B virus, herpes simplex, and adenovirus.^{343–345}

As depicted in Figure V.1, luteolin and quercetin are based on a 15-carbon skeleton with a chromone core comprising bicyclic 1,4-benzopyrone (A- and C-rings) substituted on carbon 2 with a catechol moiety (B-ring). Ring A features a phloroglucinol substitution pattern with two free hydroxyl groups in position 5 and 7. Notably, quercetin differs from luteolin by only one additional hydroxyl group in 3 position.

Luteolin and quercetin have already been the subject of *in silico* and *in vitro* studies focused on the SARS-CoV-2 M^{pro} and spike proteins.^{346,347} Docking calculations followed by *in vitro* testing showed that luteolin and quercetin inhibit the viral protease 3CL^{pro} with IC₅₀ values in the micromolar range (20 μ M and 24 μ M, respectively), and with K_i \sim 7 μ M in the case of quercetin.^{348–350} In addition, quercetin was found to be active against two crucial targets of SARS-CoV, namely M^{pro} and NTPase/helicase.^{351,352}

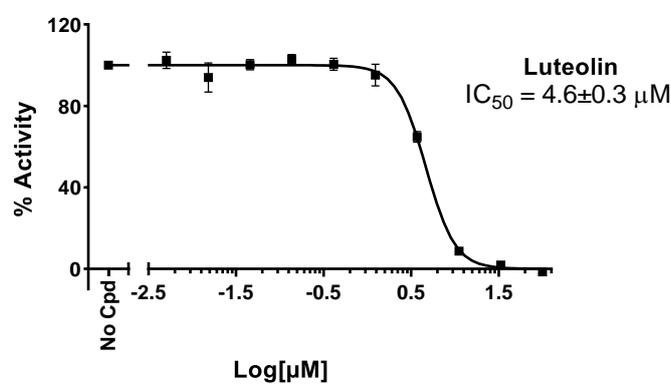
Furthermore, molecular docking analysis of natural compounds in the active site of RdRp of SARS-CoV-2 suggest luteolin and quercetin as potential inhibitors of this crucial viral enzyme.³⁵³ Nevertheless, to the best of our knowledge, they have never been experimentally tested on SARS-CoV-2 RdRp. Thus, we decided to evaluate their activity against this specific target. Here, we report their potency and computed binding mode at the viral RdRp target.

2. Results and Discussion

First, luteolin and quercetin were tested at two fixed concentrations of 25 μ M and 100 μ M for their ability to block the viral RdRp target. Both compounds completely inhibited the enzyme at 100 μ M, with an inhibition of more than 80% at 25 μ M (Table 1). Prompted by

these preliminary data, we measured dose-response curves to calculate the IC_{50} values (Table 1) by determining each compound's inhibition activity at 10 different concentrations, ranging from 0.005 μM to 100 μM (Figure V.2). Luteolin returned an IC_{50} of $4.6 \pm 0.3 \mu\text{M}$ and quercetin an IC_{50} of $6.9 \pm 1.0 \mu\text{M}$. Thus, both compounds displayed a greater potency against RdRp polymerase than those reported against SARS-CoV-2 M^{pro} and spike proteins (see above). Encouraged by such promising single-digit IC_{50} values for these compounds, we decided to evaluate *in vitro* their drug-like properties, namely aqueous kinetic solubility, together with metabolic and plasma stabilities (Table V.1).

A



B

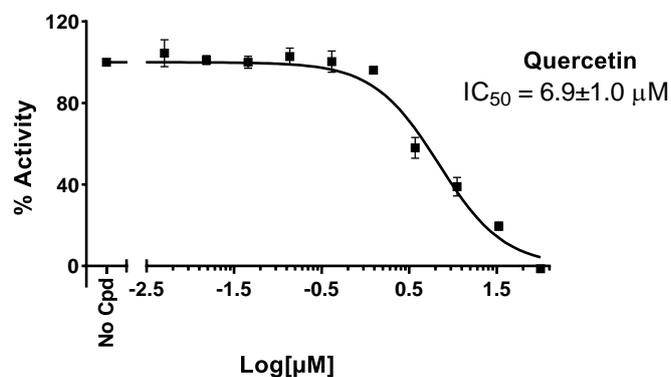


Figure V.2 Dose-response curves with IC_{50} values for luteolin (A) and quercetin (B).

Luteolin and quercetin have a kinetic solubility of $21 \pm 4 \mu\text{M}$ and $16 \pm 5 \mu\text{M}$ in PBS neutral buffer (pH 7.4), respectively. In terms of metabolic stability, both compounds showed an optimal microsomal stability ($t_{1/2} > 60 \text{ min}$). In blood plasma, luteolin was stable in the measured time span (120 min), while quercetin was poorly stable ($t_{1/2} = 7 \pm 2 \text{ min}$), probably due to the additional hydroxyl group in 3 position. A recently published

cryoelectron microscopy (cryo-EM) structure of SARS-CoV-2 RdRp in complex with two molecules of suramin (PDB ID 7D4F),³⁵⁴ a century-old non-nucleotide analog drug, has revealed two new druggable pockets at the protein target. The binding of suramin to one pocket, B_{RNA} (Figure V.3), prevents the binding of the RNA template strand, while the binding of suramin to the other pocket, B_{NTP} (Figure V.3), prevents both the entry of the nucleotide triphosphate into the catalytic site and the binding of the RNA processed strand (Figure V.3). However, despite a promising IC₅₀ value of 0.26 μ M,³³ suramin is associated with a high risk of off-target effects on other enzymes in the cell, together with its highly negative charge, which may hinder its penetration into cells.³⁵⁴ Nevertheless, these two newly identified binding sites at RdRp are suitable pockets to target with non-nucleotide analog drug hits.

Table V.1 In vitro inhibitory activity against SARS-CoV-2 RdRp, kinetic solubility in neutral water, microsomal stability in mouse, and plasma stability in mouse of luteolin and quercetin.

Compound	% inhib. 100 μ M	% inhib. 25 μ M	IC ₅₀ (μ M)	Sk (μ M)	t _{1/2} micr. (min)	t _{1/2} pl. (min)
Luteolin	100	89	4.6 \pm 0.3	21 \pm 4	>60	>120
Quercetin	100	81	6.9 \pm 1.0	16 \pm 5	>60	7 \pm 2

Based on these structural findings, we investigated the possible binding modes and the protein-ligand interactions for luteolin and quercetin at the B_{RNA} and B_{NTP} pockets in RdRp. Specifically, the B_{RNA} cavity is formed by the conserved G motif and the N terminus of B motif of the enzyme, and the key residues interacting with suramin are Asn497, Lys500, Arg569, Gln573 and Lys577. In contrast, the B_{NTP} cavity is located near the catalytic active site, which is formed by conserved A, C, E, and F motifs. Here, the key interactions are formed between suramin and Lys551, Arg553, Arg555, Arg836, and Asp865 residues.³⁵⁴ To explore the binding modes for luteolin and quercetin, we performed molecular docking of both molecules on the two binding sites (i.e. B_{RNA} and B_{NTP}) after removing the suramin (Figure 4). The Schrödinger's Protein Preparation Wizard tool was used to prepare the protein, with the addition of hydrogens and the prediction of pK_a values for ionizable residues. Subsequently, an extensive visual inspection was carried to check the overall quality of the final structures. Then, for the molecular docking, luteolin and quercetin were processed with the LigPrep tool to properly prepare the ligands (e.g. assigning atom charges, converting 2D to 3D structures, and generating tautomeric and ionization states –

at $\text{pH}=7.0 \pm 0.4$). LigPrep generated four structures (i.e. two for luteolin and two for quercetin), which differ in the protonation state of the OH group at position 7 (Figure C. 1), namely LutOH, LutO⁻, QueOH, and QueO⁻. These structures were used for protein–ligand docking with Glide. The docking grid was centered on the suramin’s center of mass, either bound to the B_{RNA} or the B_{NTP} pocket. Extra-precision Glide (XP)³⁵⁵ was used and a maximum of 20 poses for each molecule were generated (for a total of 24 and 39 poses for B_{NTP} and B_{RNA} pockets, respectively). The resulting docking scores for luteolin and quercetin are shown in the Supporting Information (Table C. 1). For the B_{NTP} binding pocket, the best docking scores (in kcal mol⁻¹) correspond to -7.62 and -5.23 , for QueO⁻ and LutOH molecules, respectively. At the B_{RNA} pocket, the main interactions are formed between QueO⁻/LutOH molecules and His439, Ser549, Lys551, Arg555, Ser814, His816, and Arg836 residues (Figure V.4). In contrast, at the B_{RNA} binding pocket, the predicted higher docking scores corresponded to -7.69 and -6.18 , for QueO⁻ and LutO⁻ molecules, respectively. The main interactions identified by the molecular docking are between the ligands and Asn496, Asn497, Lys500, Arg 569, Gln573, Lys577, and Tyr689 (Figure V.4).

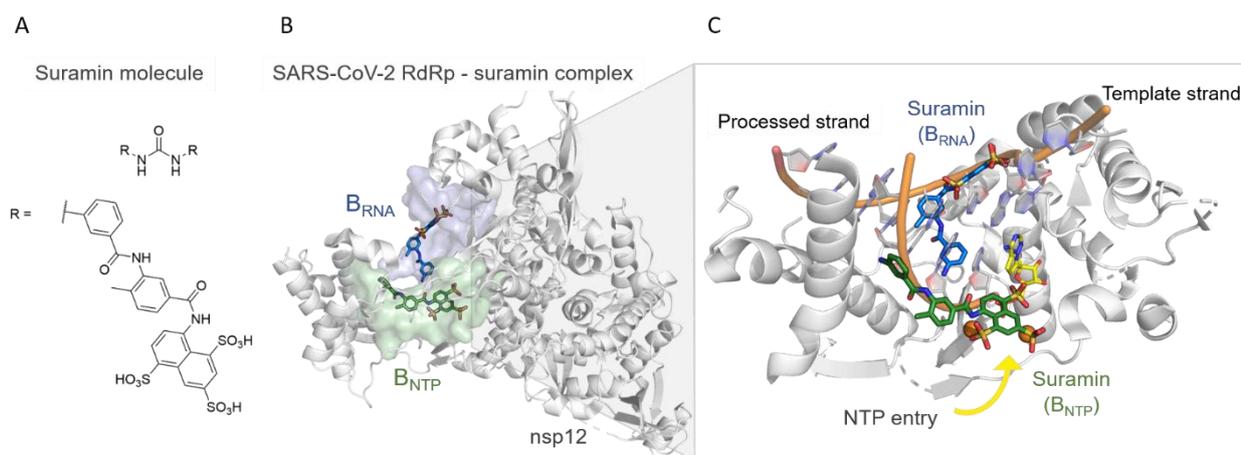


Figure V.3 (A) 2D structure of suramin. (B) Crystallographic structure (PDB ID 7D4F)[33] of the RdRp-suramin complex. Only the catalytic nonstructural protein 12 (i.e. nsp12) is depicted. On the left, the two binding pockets with suramin molecules bound, i.e. B_{RNA} and B_{NTP}, are depicted as blue and green surfaces, respectively. (C) Close view of the two binding sites, the suramin molecules (as blue and green licorice), the superimposed double-strand RNA (as cartoon), and the incoming nucleotide (as yellow licorice).

To further check the stability of the docked structures, we ran equilibrium force-field-based MD simulations (~480 ns in total) of the four selected XP poses for luteolin and quercetin at the B_{RNA} and B_{NTP} binding pockets. The integration of experimental results with

molecular docking and MD simulations provided a detailed molecular understanding of the inhibitory action of luteolin and quercetin on RdRp.

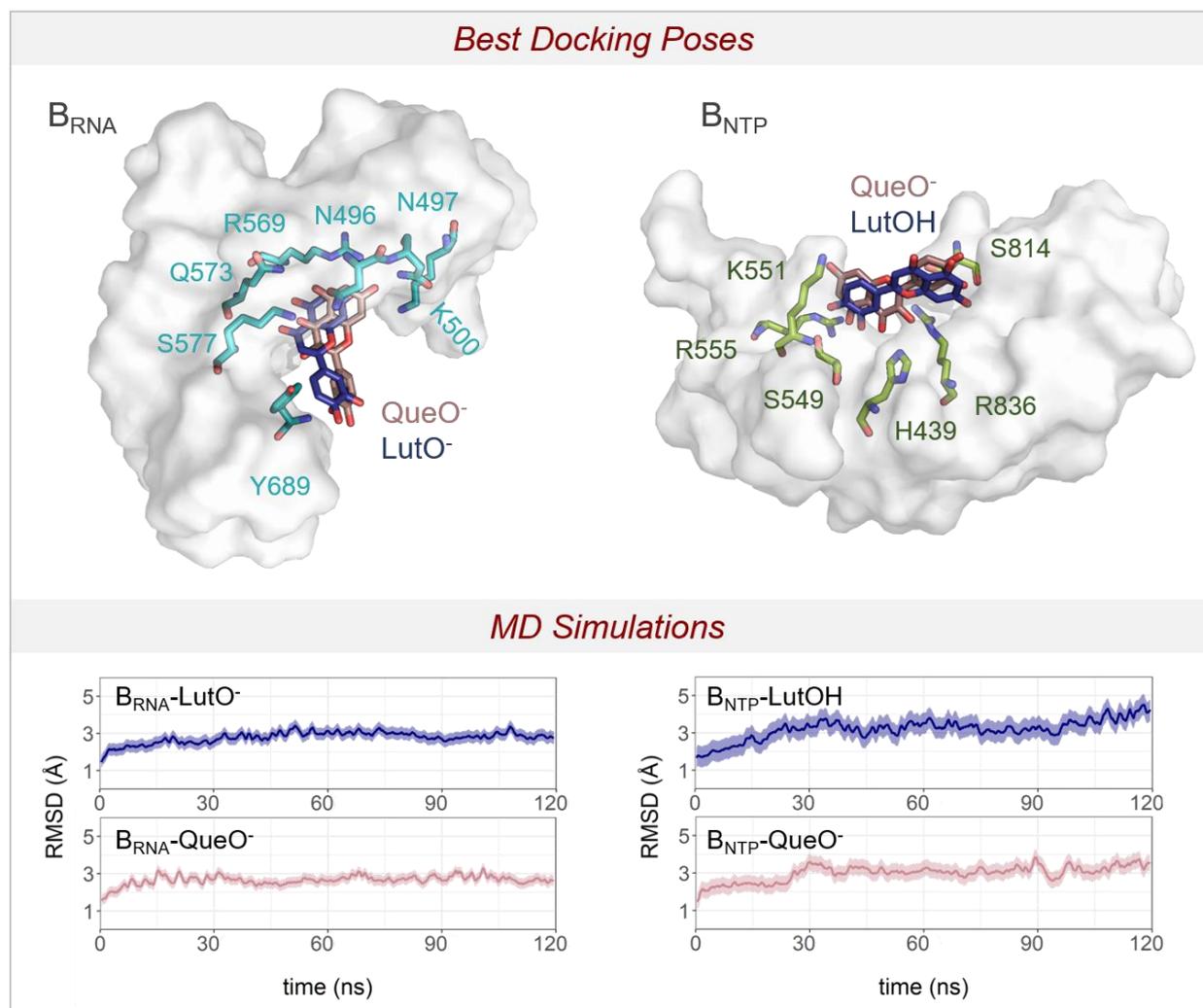


Figure V.4 Top: XP Glide docking poses for luteolin and quercetin in dark blue and pink licorice, respectively. The interacting residues are in green licorice for B_{NTP} pocket and light blue for B_{RNA} pocket. Bottom: the root mean squared deviations (RMSD) of the MD simulations for the four systems: i.e. i) B_{RNA} -LutO⁻, ii) B_{RNA} -QueO⁻, iii) B_{NTP} -LutOH, and iv) B_{NTP} -QueO⁻.

This strategy had already been successfully applied in several other cases.^{356–360} A total of four MD simulations of 120 ns each were performed, i.e. i) B_{RNA} -LutO⁻, ii) B_{RNA} -QueO⁻, iii) B_{NTP} -LutOH, and iv) B_{NTP} -QueO⁻ (see Supporting Information for details). In both B_{NTP} -LutOH/QueO⁻ systems, the overall ligand-enzyme complex stably maintains the interactions of the starting docking structure. The RMSD mean values for the heavy atoms of the complexes are 3.2 ± 0.5 Å and 3.0 ± 0.4 Å, respectively (Figure V.4). During the MD simulations of B_{NTP} -LutOH system, the ligand showed some flexibility over time, with the RMSD mean value for the heavy atoms of 7.5 ± 1.3 Å (Figure C. 2). This reflects the

reorientation and reorganization of the interactions established between LutOH and the enzyme. Specifically, after the first ~50ns, the ligand moved closer to the side chain of Ser814, Arg836, and Asp865, forming a stable network of interactions that was maintained during the simulations (Figure C. 2). Interestingly, these residues also interact through H-bond with suramin in the crystal structure. Additionally, other crystal structures of the SARS-CoV-2 RdRp-RNA complex showed that Ser814 and Arg836 interact with the RNA primer strand,^{326,361–367} further supporting the relevance of these residues for ligand recognition and binding. In contrast, B_{NTP}-QueO⁻ system showed a slightly more stable conformation of the ligand within the pocket, with an RMSD value for the heavy atoms of 6.3 ± 0.9 Å (Figure C. 3). Here, the interactions formed by QueO⁻ with the RdRp enzyme involved His439, Ser549, Ser814, Arg836 side chains (Figure C. 3) and the Ile548 backbone. Although His439, Ile548, and Ser549 do not directly participate in RNA binding, they are positioned within ~10 Å from the double-strand RNA and from the entry path of the incoming nucleotide.

For the B_{RNA}-LutO⁻/QueO⁻ systems, the overall ligand/enzyme complex showed no major differences in the overall stability, as reported by an RMSD mean value for the heavy atoms of 2.8 ± 0.3 Å and 2.6 ± 0.3 Å, respectively (Figure V.4). Here, both ligands showed a reduced flexibility compared to the B_{NTP}-LutOH/QueO⁻ systems, with a remarkable stability of the Que-O⁻ ligand at the pocket. Indeed, the RMSD mean values for the heavy atoms of LutO⁻/QueO⁻ are 6.4 ± 2.1 Å and 2.0 ± 0.8 Å, respectively (Figure C. 4 and Figure C. 5). This also reflects the stable H-bond interactions formed by QueO⁻ and RdRp enzyme. In detail, interactions are established between the side chains of Arg569, Gln573, and both LutO⁻ and QueO⁻ ligands (Figure C. 4 and Figure C. 5), while only the latter interacts with the side chain of Asn497, Tyr689, and Ser759 (Figure C. 5). Notably, considerable structural evidence (e.g. PDB 6XEZ, 7B3B, 7B3C, 7B3D)^{366,367} shows that these three residues interact with the template strand, thus stabilizing the RNA substrate binding.^{326,361–367} Moreover, Ser759 side chain is located close to the incoming nucleotide binding side.³⁶⁸ These residues are therefore an optimal anchor point for inhibitors of the catalytic activity of the RdRp enzyme.

Overall, the results of our MD simulations indicate that both binding pockets may properly bind and stably host luteolin and quercetin. Nevertheless, the increased stability and higher

number of contacts between the B_{RNA} pocket and these ligands suggests that this binding site may be more suitable for ligand binding and structure-based drug design.

These computational insights will serve to start future campaigns for hit-to-lead design, as witnessed recently in computational studies used to guide experiments for drug design targeting viral proteins.^{369,370} Notably, Jorgensen *et al.*³⁷¹ recently performed a virtual screening of ~2000 approved drugs with a consensus virtual screening protocol used together with MD simulations and biochemical assay. This indicated 14 known drugs active in the micromolar range against 3CL^{pro}. Starting from this evidence, the group subsequently applied free-energy perturbation (FEP) calculations to fine-tune the drug-target interaction of the initial hit, perampanel. This led to the design of a new set of compounds with IC₅₀ values in the low nanomolar range, whose binding poses have been corroborated by co-crystal structures.⁵² With this successful example in mind, our results now form the basis for a hit-to-lead campaign targeting the SARS-CoV-2 RdRp enzyme.

3. Conclusions

In summary, starting from the pharmacological properties of flavonoids, we experimentally tested luteolin and quercetin against SARS-CoV-2 RdRp, a crucial target of the virus responsible for the COVID-19 pandemic. The IC₅₀ value is $4.6 \pm 0.3 \mu\text{M}$ for luteolin and $6.9 \pm 1.0 \mu\text{M}$ for quercetin. To the best of our knowledge, this is the first study that quantifies the inhibitory potency of luteolin and quercetin against RdRp, with the evidence of a one-digit micromolar range. Notably, this inhibitory activity is better than previous IC₅₀ values reported for these two compounds against other viral proteins of SARS-CoV-2. We also investigated and proposed potential binding modes of these compounds to the target protein. Thus, our experimental and computational results complete previous computational investigations that proposed these two known natural products against COVID, providing experimental values for activity and new mechanistic insights.³⁵³ Taken together, our results endorse a further exploration of a new chromone class of RdRp polymerase inhibitors to treat Sars-CoV-2 and potentially other viral infections.

4. Material and Methods

Biochemical assay. The natural flavonoids luteolin and quercetin were tested against SARS-CoV-2 RdRp with an in vitro enzymatic inhibition assay in collaboration with BPS Bioscience. The assay was performed with compounds obtained from commercial sources (luteolin from Fluorochem, quercetin from Sigma-Aldrich). Compounds purity is >99% based on our HPLC analysis (see SI). The RdRp reactions were conducted in duplicate at 37°C for 60 minutes in a 10 µl mixture containing assay buffer (20 mM Tris pH8.0 and 0.01% Triton X100), RNA duplex, ATP substrate and enzyme, and the test compound. The enzyme was produced by BPS Bioscience, and was formulated as 45 mM Tris-HCl pH 8.0, 124 mM NaCl, 2.4 mM KCl, 4 mM MgCl₂, 1 mM TCEP, 10% glycerol. Typical purity was 95-97%, and typical concentration was 1 mg/ml. These 10 µl reactions were carried out in wells of 384-well Optiplate (PerkinElmer). A 10 mM stock solution of test compound in DMSO was prepared. Dilutions of this stock solution were prepared in assay buffer (5% DMSO concentration) and 2 µl of the dilution was added to a 6 µl of RdRp (final concentration 0.08 mg/mL) containing RNase inhibitor for preincubation (30 minutes at room temperature with slow shaking). Reaction was started by addition of 2 µl of the substrate mix containing RNA duplex (40 nM) and ATP substrate (3 µM). Final concentration of DMSO was 1% in all reactions (reference compound – 0% DMSO). After enzymatic reactions, 10 µl of anti-Dig Acceptor beads (PerkinElmer, diluted 1:500 with 1x detection buffer) were added to the reaction mix. After brief shaking, plate was incubated for 30 minutes. Finally, 10 µl of AlphaScreen Streptavidin-conjugated donor beads (Perkin, diluted 1:125 with 1x detection buffer) were added. In 30 minutes, the samples were measured in AlphaScreen microplate reader (EnSpire Alpha 2390 Multilabel Reader, PerkinElmer). In the absence of the compound, the intensity (C_e) in each data set was defined as 100% of activity. In the absence of the enzyme, the intensity (C_0) in each data set was defined as 0% of activity. The percent activity in the presence of each compound was calculated according to the following equation: % activity = $(C - C_0) / (C_e - C_0)$, where C is the intensity in the presence of the compound. As a positive control, the reference compound 6-chloropurine-ribose TP was tested at three different concentrations (0.02 µM, 0.2 µM, and 2 µM).

***In vitro* microsomal stability.** 10 mM DMSO stock solution of test compound was pre-incubated at 37 °C for 15 min with mouse liver microsomes added 0.1M Tris-HCl buffer (pH 7.4). The final concentration was 4.6 µM. After pre-incubation, the co-factors (NADPH, G6P, G6PDH and MgCl₂ pre-dissolved in 0.1M Tris-HCl) were added to the incubation mixture and the incubation was continued at 37°C for 1h. At each time point (0, 5, 15, 30, 60min), 30 µL of incubation mixture was diluted with 200 µL cold CH₃CN spiked with 200 nM of internal standard, followed by centrifugation at 3500g for 15 min. The supernatant was further diluted with H₂O (1:1) for analysis. The concentration of test compound was quantified by LC/MS-MS on a Waters ACQUITY UPLC/MS TQD system consisting of a TQD (Triple Quadrupole Detector) Mass Spectrometer equipped with an Electrospray Ionization interface. The analyses were run on an ACQUITY UPLC BEH C18 (50x2.1mmID, particle size 1.7µm) with a VanGuard BEH C18 pre-column (5x2.1mmID, particle size 1.7µm) at 40 °C, using 0.1% HCOOH in H₂O (A) and 0.1% HCOOH in CH₃CN (B) as mobile phase. Electrospray ionization (ESI) was applied in positive mode. The percentage of test compound remaining at each time point relative to t = 0 was calculated. The half-lives (t_{1/2}) were determined by a one-phase decay equation using a non-linear regression of compound concentration versus time.

***In vitro* Plasma Stability.** 10 mM DMSO stock solution of test compound was diluted 50-fold with DMSO-H₂O (1:1) and incubated at 37°C for 2 h with mouse plasma added 5% DMSO (pre-heated at 37°C for 10 min). The final concentration was 2 µM. At each time point (0, 5, 15, 30, 60, 120min), 50 µL of incubation mixture was diluted with 200 µL cold CH₃CN spiked with 200 nM of internal standard, followed by centrifugation at 3500 g for 20 min. The supernatant was further diluted with H₂O (1:1) for analysis. The concentration of test compound was quantified by LC/MS-MS on a Waters ACQUITY UPLC/MS TQD system consisting of a TQD (Triple Quadrupole Detector) Mass Spectrometer equipped with an Electrospray Ionization interface. The analyses were run on an ACQUITY UPLC BEH C18 (50x2.1mmID, particle size 1.7µm) with a VanGuard BEH C18 precolumn (5x2.1mmID, particle size 1.7µm) at 40 °C, using 0.1% HCOOH in H₂O (A) and 0.1% HCOOH in CH₃CN (B) as mobile phase. Electrospray ionization (ESI) was applied in positive mode. The response factors, calculated on the basis of the internal standard peak

area, were plotted over time. When possible, response vs. time profiles were fitted with Prism (GraphPad Software, Inc., USA) to estimate compounds half-life in plasma.

Aqueous kinetic solubility. The aqueous kinetic solubility was determined from a 10 mM DMSO stock solution of test compound in Phosphate Buffered Saline (PBS) at pH 7.4. The study was performed by incubation of an aliquot of 10 mM DMSO stock solution in PBS (pH 7.4) at a target concentration of 250 μ M resulting in a final concentration of 2.5% DMSO. The incubation was carried out under shaking at 25°C for 24h followed by centrifugation at 21.100g for 30 min. The supernatant was analyzed by UPLC/MS for the quantification of dissolved compound by UV at a specific wavelength (215 nm). The analyses were performed on a Waters ACQUITY UPLC/MS SQD system consisting of a SQD (Single Quadrupole Detector) Mass Spectrometer equipped with Electrospray Ionization interface. The analyses were run on an ACQUITY UPLC BEH C18 column (50x2.1mmID, particle size 1.7 μ m) with a VanGuard BEH C18 pre-column (5x2.1mmID, particle size 1.7 μ m), using 10 mM NH₄OAc in H₂O at pH 5 adjusted with AcOH (A) and 10mM NH₄OAc in CH₃CN-H₂O (95:5) at pH 5 (B) as mobile phase.

Molecular docking of luteolin and quercetin with SARS-CoV-2 RdRp. First, the SARS-CoV-2 RdRp was retrieved from PDB database (PDB ID 7D4F³⁵⁴) and subsequently prepared for docking. The preparation was carried by Schrödinger's Protein Preparation Wizard tool and included: i) addition of hydrogen atoms, ii) elimination of water molecules not involved in ligand-binding interaction, iii) assignment of atomic charges. Subsequently, energy minimized 3D molecular structures of luteolin and quercetin were generated and prepared for docking using LigPrep tool. Additionally, possible ionization states were generated using LigPrep tool, thus resulting in two possible states for each molecules (see Figure C. 1). Eventually, the SARS-CoV-2 RdRp structure (PDB ID 7D4F³⁵⁴) was used for docking luteolin and quercetin. The grid was centered on the suramin's center of mass, and the docking was performed using Glide XP methodology^{355,372}.

Structural Models for molecular dynamics simulations. We used four different systems for the MD simulations: i) B_{RNA}-LutO⁻, ii) B_{RNA}-QueO⁻, iii) B_{NTP}-LutOH and iv) B_{NTP}-QueO⁻. Each system was solvated with a 12-Å layer of TIP3P water molecules, and Na⁺

ions were added to neutralize the net charge of the systems. The final models include a total of ~183,000 atoms.

Molecular dynamics simulations set-up. We used force-field-based MD simulations to check the stability of the docked structures. Here, the AMBER-ff14SB³⁷³ force field was used to treat the RdRp enzyme. All four ligands were parametrized with the general Amber force field (GAFF)³⁷⁴, and the atomic charges were derived using the RESP procedure, according to the Merz–Singh–Kollman scheme.³⁷⁵ Na⁺ metal ions were treated using the Joung–Cheatham parameters.³⁷⁶ A time integration step of 2 fs was used and the lengths of all bonds involving hydrogen atoms were constrained using the P-LINCS algorithm.³⁷⁷ A velocity-rescaling thermostat was used to set a system temperature of 310K,³⁷⁸ while the Parrinello–Rahman barostat maintained a constant pressure of 1bar.³⁷⁹ Long-range electrostatic interactions were calculated with the particle mesh Ewald (PME) method using a Fourier grid spacing of 1.6 Å. Periodic boundary conditions in the three directions of Cartesian space were applied. All MD simulations were performed with Amber2020. The systems were all subject to the same MD simulations procedure. First, we carried out energy minimization to relax the water molecule and the ions. Here, both the ligand and the RdRp backbone were kept fixed with harmonic positional restraints of 300 kcal mol⁻¹ Å². Then, the systems were heated up from 0 to 310 K with NVT simulations for a total of 1 ns with 300 kcal mol⁻¹ Å² restraints on the ligand. Additionally, 1 ns of simulations in NPT ensemble was performed with the same positional restraints used in the NVT simulations. Three additional NPT simulations of 1 ns each were performed gradually removing the restraints on the ligand. Finally, a production run were performed in the NPT ensemble for each system. We collected overall ~480 ns of MD trajectories, specifically ~120 ns for each system.

Chapter VI. Final Remarks and Perspective

Nucleic acids processing enzymes are responsible for vital processes during cell life, like repair, recombination and replication mechanisms. Among them, metallo-nucleases employ catalytic metal ions to catalyze the phosphodiester bond hydrolysis, in particular the two-metal-ion mechanism is the most commonly used. Other metalloenzymes share the same strategy to efficiently perform the catalytic reaction, such as DNA and RNA polymerases, for this they share common catalytic core. Over the years, a lot of attention has been dedicated to investigate the two-metal-ion catalysis, thus allowing a deep comprehension at atomistic level of such mechanism. Moreover, further studies have suggested additional structural elements as catalytically relevant to ensure fidelity, specificity, and efficiency for nucleic acids processing. Such elements locates in close proximity of the two-metal-ion centre. However, whether these elements have a catalytic function and what is their role during catalysis is still unclear.

In this context, recent *in crystallo* reaction intermediates of hExo1 have captured a third divalent metal ion intermittently bound close to the two-metal-ion active site. Evidences of this third ion has been observed in several polymerases and nucleases. This work presents a computational investigation at atomistic level on the recruiting mechanism and functional role of such third metal ion during hExo1 catalysis (0). Second-shell acidic residue (Glu), is found to act as recruiter for the third metal ion, which is turn serves as an exit shuttle for the leaving group departure after hydrolysis. Further structural analysis show that such second-shell negatively charged element is persistently found in a similar, structurally conserved, and strategic position in several other nucleases, which seem to share this enzymatic mechanism to promote DNA hydrolysis.

In addition, crystallographic structures together with mutagenesis data of hFEN1, suggest phosphate steering as molecular mechanism to favor specificity and catalysis. Intriguingly, four basic residues located in the second- and third-shell of the two-metal-ion active site are proposed to electrostatically govern and control such molecular mechanism. This work offers a molecular understanding of how these conserved positively charged residues operate the phosphate steering mechanism to promote catalysis (Chapter IV). The

mechanistic action of such residues is found to be critical in hFEN1 catalysis for substrate binding and correct site incision, eventually prompting leaving group departure. Additionally, structural analysis of other metallonucleases identified basic residues similarly localized in the vicinity of the catalytic site, thus suggesting that this catalysis-enabling mechanism is conserved and used by other metallonucleases.

In parallel to the main research area, a drug discovery project has been carried on. To date, the new SARS-CoV-2, is having a ruinous impact on human health care as well as global economy. Although the massive vaccination campaign carried out in several countries, the development of effective drugs to treat SARS-CoV-2 infection remain an urgent unmet need. Among the possible targets, the RdRp has been identified as potential target for a drug discovery campaign. Given the pharmacological properties of flavonoids, luteolin and quercetin have been experimentally tested against SARS-CoV-2 RdRp. Notably, the resulting IC_{50} values of $4.6 \pm 0.3 \mu\text{M}$ for luteolin and $6.9 \pm 1.0 \mu\text{M}$ for quercetin turned out to be better than previous IC_{50} values reported for these two compounds against other viral proteins of SARS-CoV-2. We also investigated and proposed potential binding modes of these compounds to the target protein. Taken together, our results endorse a further exploration of a new chromone class of RdRp polymerase inhibitors to treat SARS-CoV-2 and potentially other viral infections.

To conclude, metalloenzymes offer intriguing applications such as target therapy, engineering enzymes and biotechnology. Even though extensive efforts have contributed to elucidate key mechanistic aspects of catalysis in DNA and RNA processing, there is still much to clarify. The work presented here comprises molecular mechanism insights that increase our understanding of structure-function properties of metallonucleases with implications for enzyme design and drug discovery.

Appendix A

Structural Models. To model the reactant state, we employed the X-ray structure of the wild-type pre-reactive system (PDB ID 5V06)⁵⁹ with a resolution of 2.75 Å. Notably, we removed the last three residues (residues 355, 356 and 357) due to the fact the previous 8 residues, i.e. residues between 347 and 354, were missing from the X-ray structure. We removed also the residue 346, in order to make the systems consistent with the product system, where the residue 346 is missing. Moreover, also the first residue (Met1) was missing. To reproduce the wild-type structure of hExo1 in a competent state for catalysis, the manganese ions were replaced with magnesium ions. The final structure of the wild-type system includes residues from 2 to 346, and 4 metal ions, the two catalytic MgA and MgB, the third ion MgC and the K⁺ bound to the DNA and H2TH motif. In order to model the reactant system without the presence of the third MgC in the vicinity of the terminal 5' phosphate, we manually removed MgC.

From the reactant structure without the third metal ion in the vicinity of the active site, we modelled the Glu89Ala mutant system, where the native Glu89 was replaced by an alanine. To model the product state, we employed the X-ray structure of the cleaved system (PDB ID 5V0A)⁵⁹, with a resolution of 2.38 Å. We removed the last two residues, i.e. residues 355 and 356. Also, in this structure the first residue (Met1) was missing. To reproduce the wild-type structure of hExo1 in the product state, the Ala225 mutation and the manganese ions were replaced with the native aspartate and magnesium ions. As for the reactant mutant system, we built the Glu89Ala mutant for the product system. From the wild-type product structure, the Glu89 was replaced by the alanine.

Then, we modelled an additional product state, which was built using the pre-reactive crystal structure as a template, which also contains the third ion. In this model, we manually cleaved the scissile phosphate in order to free the leaving group, and we inserted a bond between the electrophilic phosphorous and the nucleophilic oxygen. Finally, we also adjusted the geometry of the cleaved newly formed 5' phosphate, in order to have the correct geometry for a S_N2-like reaction, as expected for this two-metal-ion catalysis. As for the reactant system, the manganese ions were replaced with magnesium ions.

Hydrogen atoms were added to each system. Subsequently, each system was immersed in a water box where the distance between the solute and the edge of the box was set to 12 Å.

To neutralize the overall charge of the systems Cl^- , K^+ , Na^+ and Mg^{2+} ions have been added. For all these three procedures, we used the tLEAP program of the AMBER 17 package.

Classical Molecular Dynamics Simulations. The protocol we adopted to run classical MD simulations was formed by two steps: i) equilibration phase; ii) production phase. For the equilibration phase, each system for the reactant state was minimized using a steepest-descent minimization algorithm keeping everything but the water fixed in the crystallographic position with a harmonic force constant of $5000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$. Then a second minimization with the same algorithm without any restraint. This procedure was adopted for the reactant state because one water molecule was missing in the first coordination shell of one of the two catalytic metal ions. For the product state the system was minimized using a steepest-descent minimization algorithm. Then, each minimized system was heated from 0 to 310 K in 450 ps in NVT ensemble, keeping the backbone atoms and the two catalytic magnesium atoms fixed in their original positions, using a harmonic force constant of $5000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$. After this, 5 ns NPT simulation was carried out at 310 K and 1 bar, maintaining backbone and magnesium atoms constrained. Eventually, for the production phase, MD was carried out in the NPT ensemble without any restraint. The first 50 ns of the production phase was considered as part of the equilibration and for this reason it was discarded from subsequent analyses. A short-range neighbor list cut-off of 12 \AA was used.³⁸⁰ A Parrinello-Rahman barostat²⁸⁵ and a velocity-rescaling thermostat²⁸⁴ were employed with a relaxation time τ of 2 ps and 0.1 ps, respectively. All the snapshots were saved every 50 ps.

Confined Well-tempered Metadynamics. In the present work, we were interested to calculate the free-energy surface (FES) for the release of the leaving group (AMP) in presence/absence of the third metal ion. In this case, regions of the solvent box very far from the active site were of no interest. Thus, our goal was to sample as much as possible the region where the distance between the centre of mass (COM) of the heavy atoms of AMP and the COM of the $\text{C}\alpha$ of the aspartates in the first coordination shell of the two catalytic metal ions, was no longer than 26 \AA . In order to avoid sampling conformational space of no interest, we used a the metadynamics approach, with a confined procedure²¹⁷

already applied by our group.²¹⁷ This method is based on the confinement of the single walker metadynamics run in a specific region of the conformational space. During the well-tempered metadynamics simulations, the value of the Collective Variable (CV) is checked. If the CV value is lower than a specified CV boundary value, determining the region of interest, then that conformation is stored in a pool of frames. When the CV value is higher, the simulation restarts from a randomly chosen conformation among the stored (see more details in Ref.²¹⁷).

Transition State Theory. The Eyring-Polanyi equation is used to derive the rate of a chemical reaction, from the state A to the state B ($k_{A \rightarrow B}$), within the temperature T.³⁸¹ The equation derives from the Transition State Theory and has the following formula:

$$k_{A \rightarrow B} = \kappa \frac{k_B T}{h} e^{-\frac{\Delta^\ddagger G_{A \rightarrow B}}{RT}} \quad [16]$$

where k_B is the Boltzmann's constant, h is the Planck's constant, R is the gas constant, $\Delta^\ddagger G_{A \rightarrow B}$ is the Gibbs energy of activation and κ is the transmission coefficient that is often assumed equal to one in order to refer to the fundamental no-recrossing assumption of Transition State Theory. The Transition State Theory provides a simplistic view of the kinetic of the enzyme and here it is used to compare the order of magnitude of the enzymatic turnover rate. From the experimental value of the catalyst rate constant, k_{cat} ,¹⁶⁵ we derived the theoretical value for the Gibbs energy of activation for the rate-determining step of the overall catalytic process using the Eyring equation. The result obtained agrees with our estimation of the free energy barrier for the leaving group departure process, which is therefore compatible with the overall catalytic reaction.

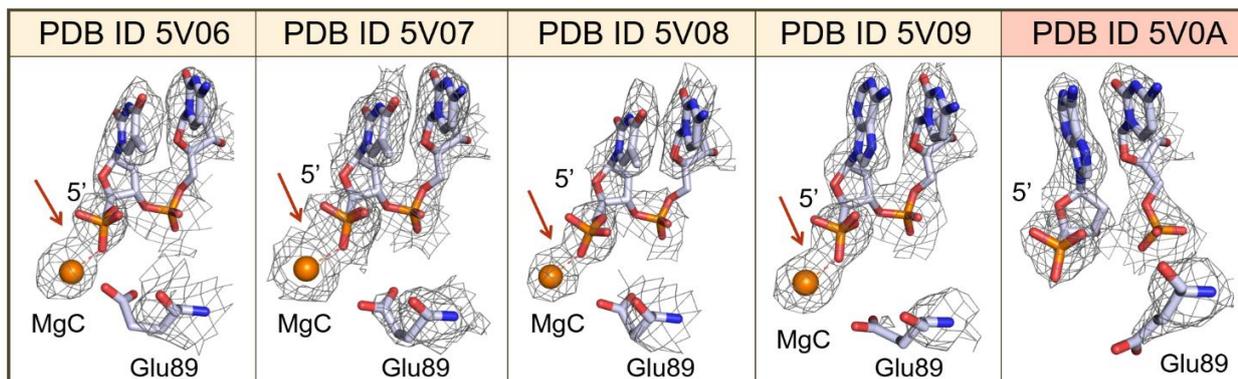


Figure A.1 Simulated annealing omit map (1σ contour) identifies the presence of MC coordinated by the terminal 5' phosphate (PDB ID 5V06, 5V07, 5V08, 5V09)⁵⁹ and the different orientations of Glu89 in the pre-reactive and post-reactive states. Diffraction data were collected at 100 K at the Advanced Photon Source or the Advanced Light Source, as reported in Shi, Y. et al⁵⁹.

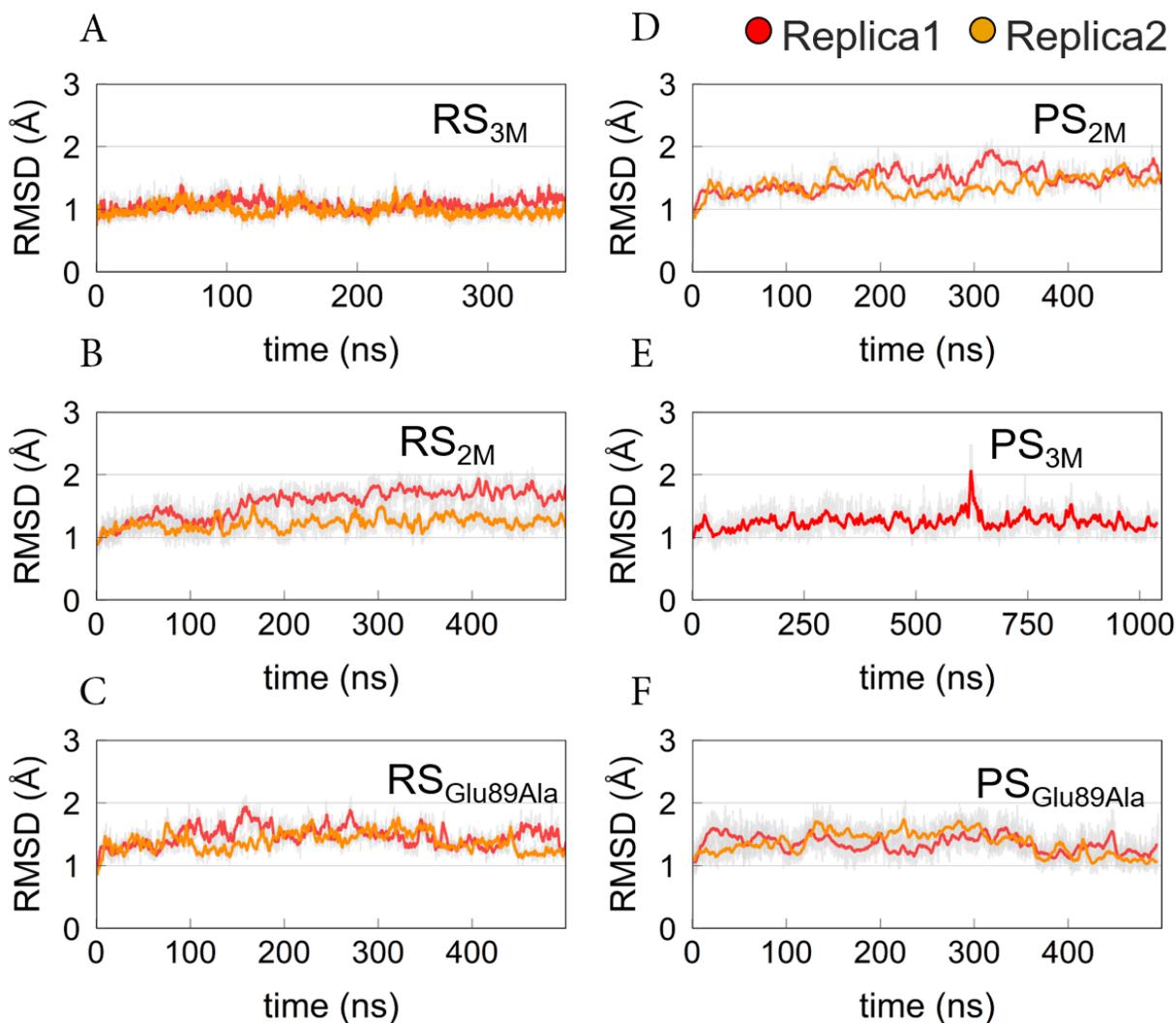
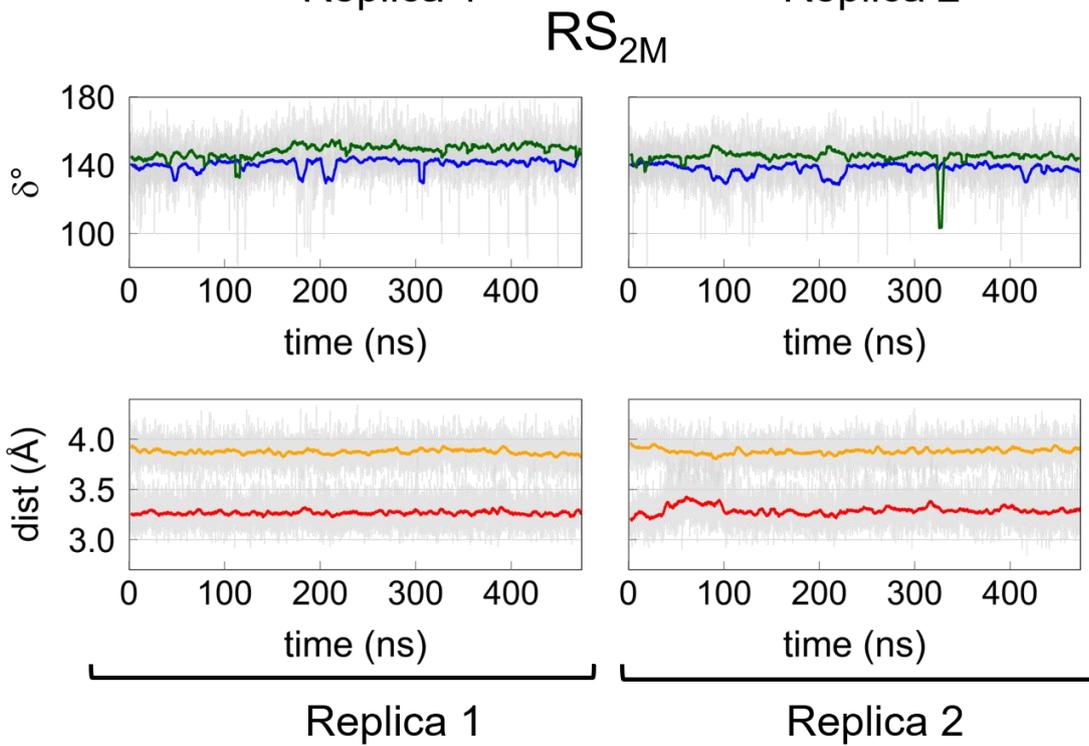
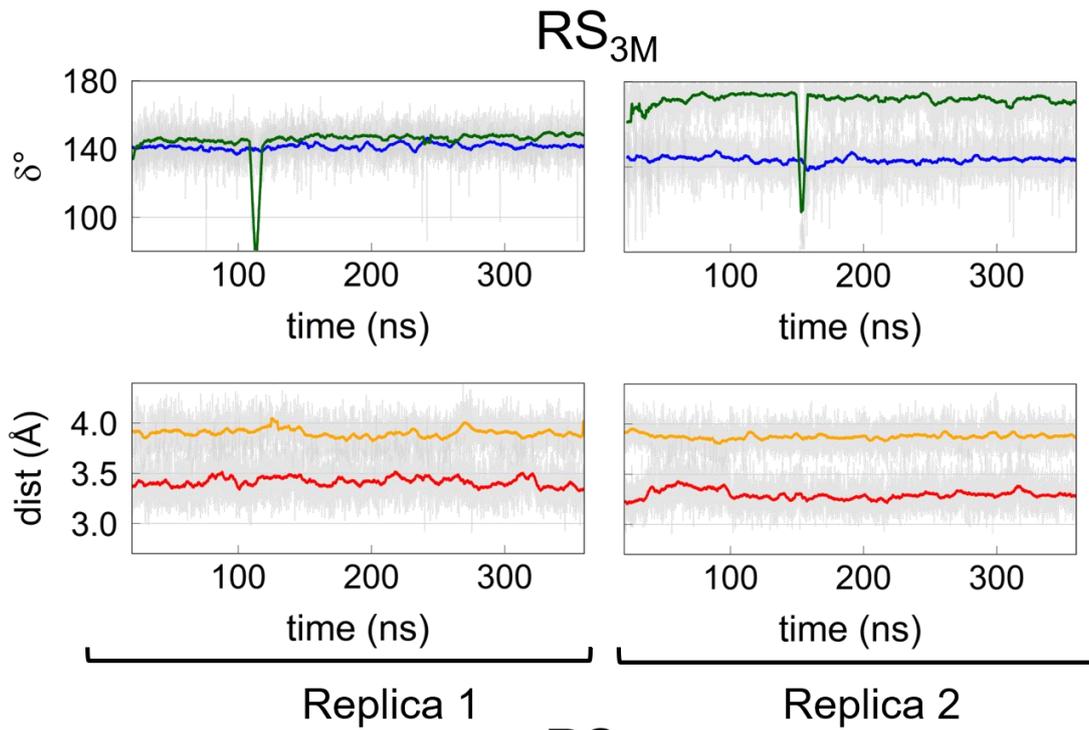


Figure A.2 Time evolution of root mean square deviation (RMSD) for the backbone atoms of hExo1. (A) Wild-type Reactant State (RS_{3M}). The average RMSD value for the Replica1 is 1.02 ± 0.11 Å and for Replica2 is 0.99 ± 0.12 Å. (B) Wild-type Reactant State without the third metal ion (RS_{2M}). The average RMSD value for the Replica1 is 1.56 ± 0.23 Å and for Replica2 is 1.23 ± 0.14 Å. (C) Mutated Glu89Ala Reactant State ($RS_{Glu89Ala}$). The average RMSD value for the Replica1 is 1.27 ± 0.15 Å and for Replica2 is 1.17 ± 0.15 Å. (D) Wild-type Product State (PS_{2M}). The average RMSD value for the Replica1 is 1.46 ± 0.22 Å and for Replica2 is 1.46 ± 0.17 Å. (E) The Product State with the third metal ion (PS_{3M}). The average RMSD value is 1.25 ± 0.15 Å. (F) Mutated Glu89Ala Product State ($PS_{Glu89Ala}$). The average RMSD value for the Replica1 is 1.35 ± 0.17 Å and for Replica2 is 1.35 ± 0.21 Å.



Continue on the next page.

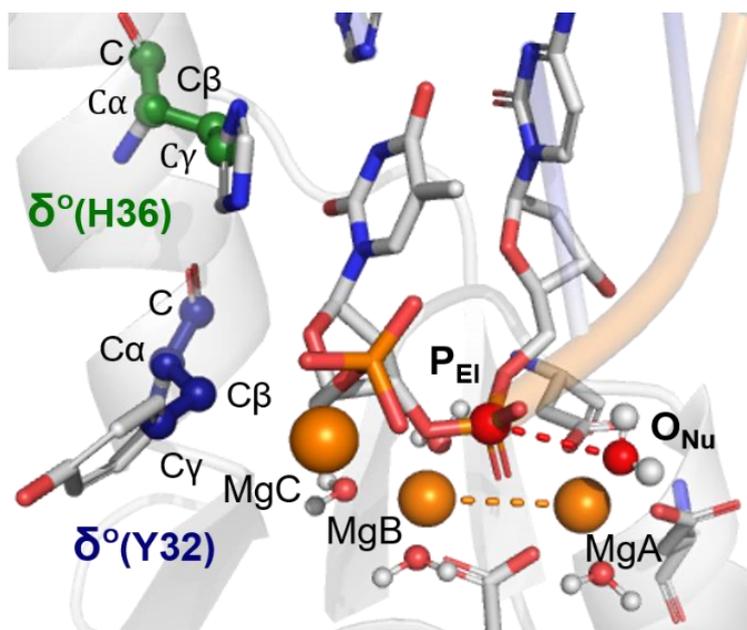


Figure A.3 (A) Representation of time evolution distances between: i) the nucleophile oxygen of the water (O_{Nu}) and the phosphorous (P_{EI}), in red and between ii) the two catalytic magnesium ions (MgA , MgB), in orange. Representation of time evolution of the dihedral angles (δ), taken along C, $C\alpha$, $C\beta$, $C\gamma$ bonds, for the guide residues Tyr32, in blue, and His36, in green. Data are taken from the reactant systems RS_{3M} and RS_{2M} . (B) Schematic representation of the distances and dihedral angles reported in (A). The structure is taken from PDB ID 5V06.

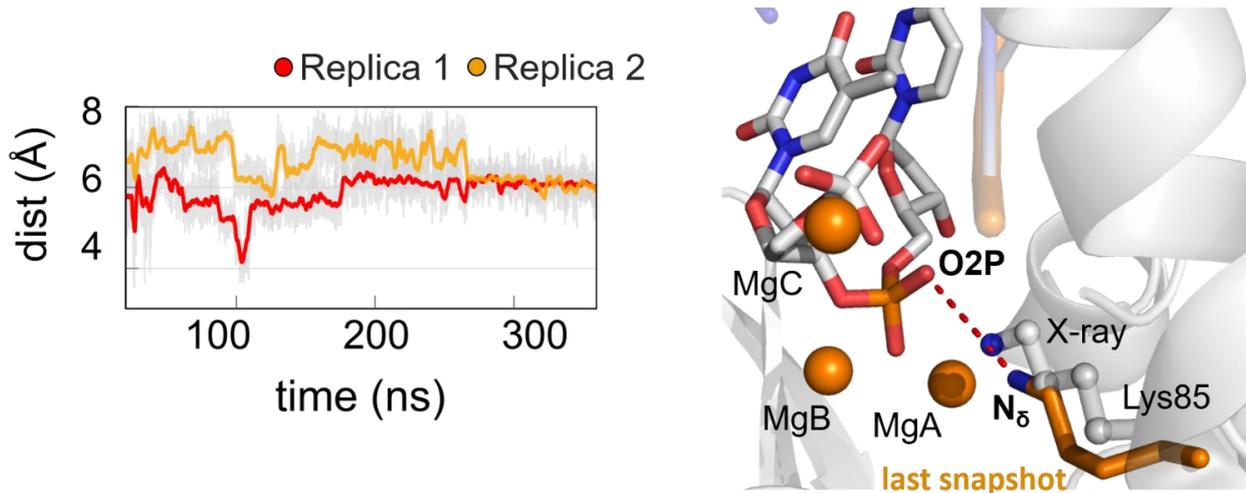


Figure A.4 (left) Representation of time evolution distance between nitrogen atom of the terminal amine group of Lys85 (N_δ) and oxygen atom of the scissile phosphate (O2P). (right) Schematic representation of the distance analysed. Lys85 in the last snapshot of the simulation is represented in licorice (orange); the magnesium ions, the DNA nucleotides and Lys85, in the crystallographic position (PDB ID 5V06) are represented in licorice and ball and stick, respectively (white).

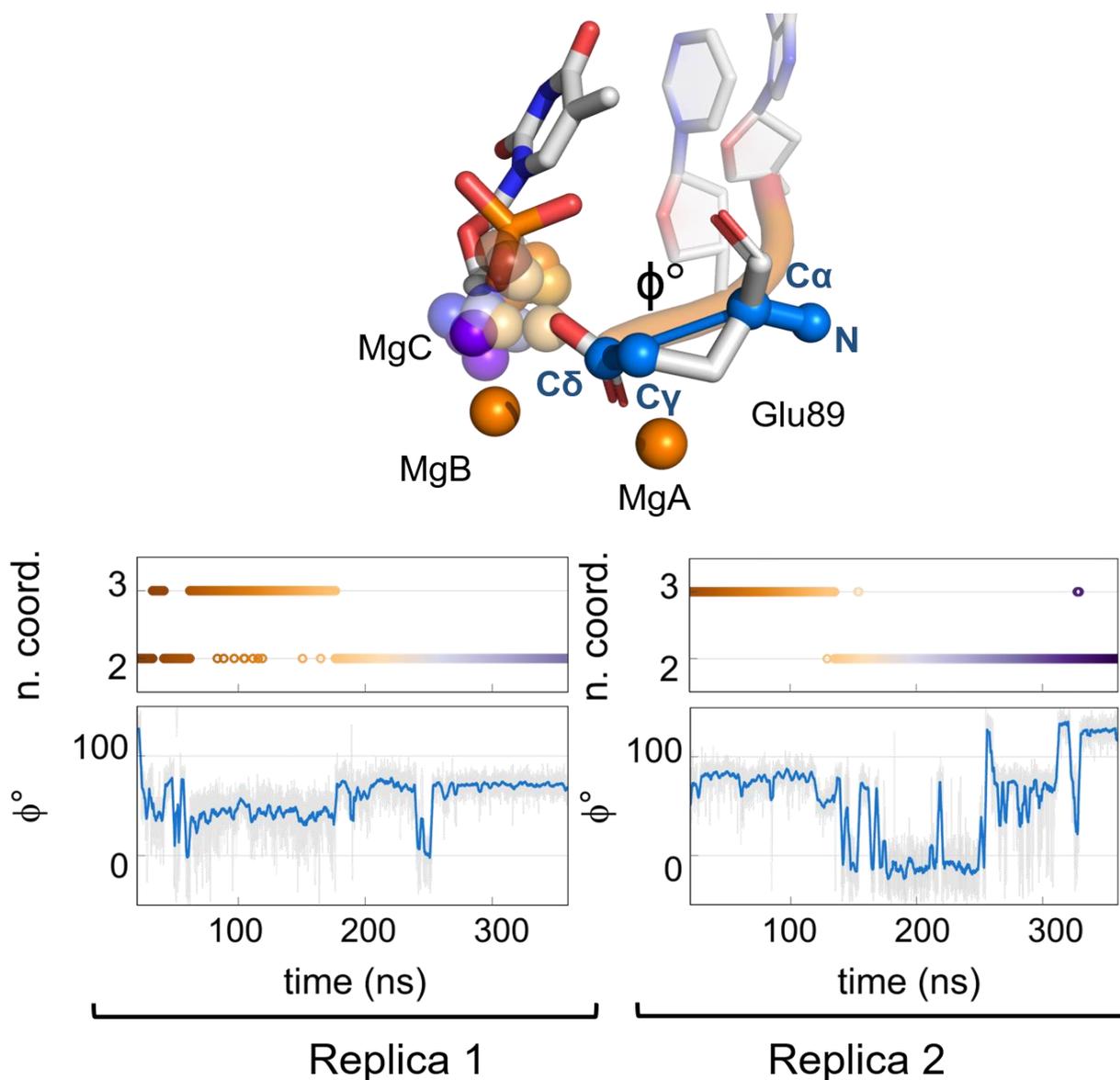


Figure A.5 (Bottom) Graphs of time evolution of coordination number of the MgC ion with non-water oxygen atoms (coloured from brown to purple, as a function of time) and the pseudo dihedral angle (ϕ) taken along N, C α , C δ , C γ bonds of Glu89. The octahedral geometry of MgC is maintained for the entire simulation time and it is completed by water molecules. (Top) Schematic representation of the pseudo dihedral angle (ϕ) and the different positions of MgC during time, exemplified using a snapshot taken from RS_{3M} simulations (the colour scheme for the MgC spheres is the same used for the coordination number in the plots in the bottom).

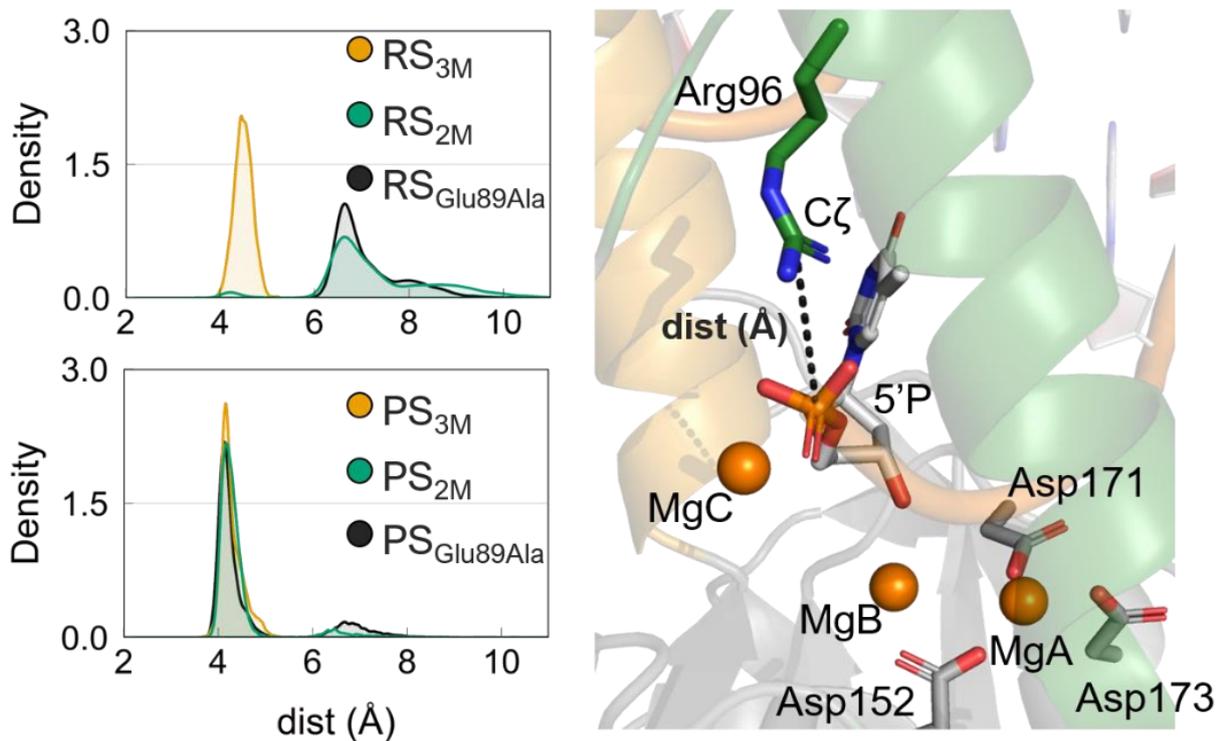


Figure A.6 (left) Frequency distribution of the distance between C ζ in Arg96 and phosphorous in the terminal 5' phosphate group. Top, the frequency distribution related to the Reactant State; Bottom, the frequency distribution related to the Product State. (right) Schematic representation of the terminal 5' nucleotide, the Arg96 and the catalytic site, taken from RS_{3M} trajectory.

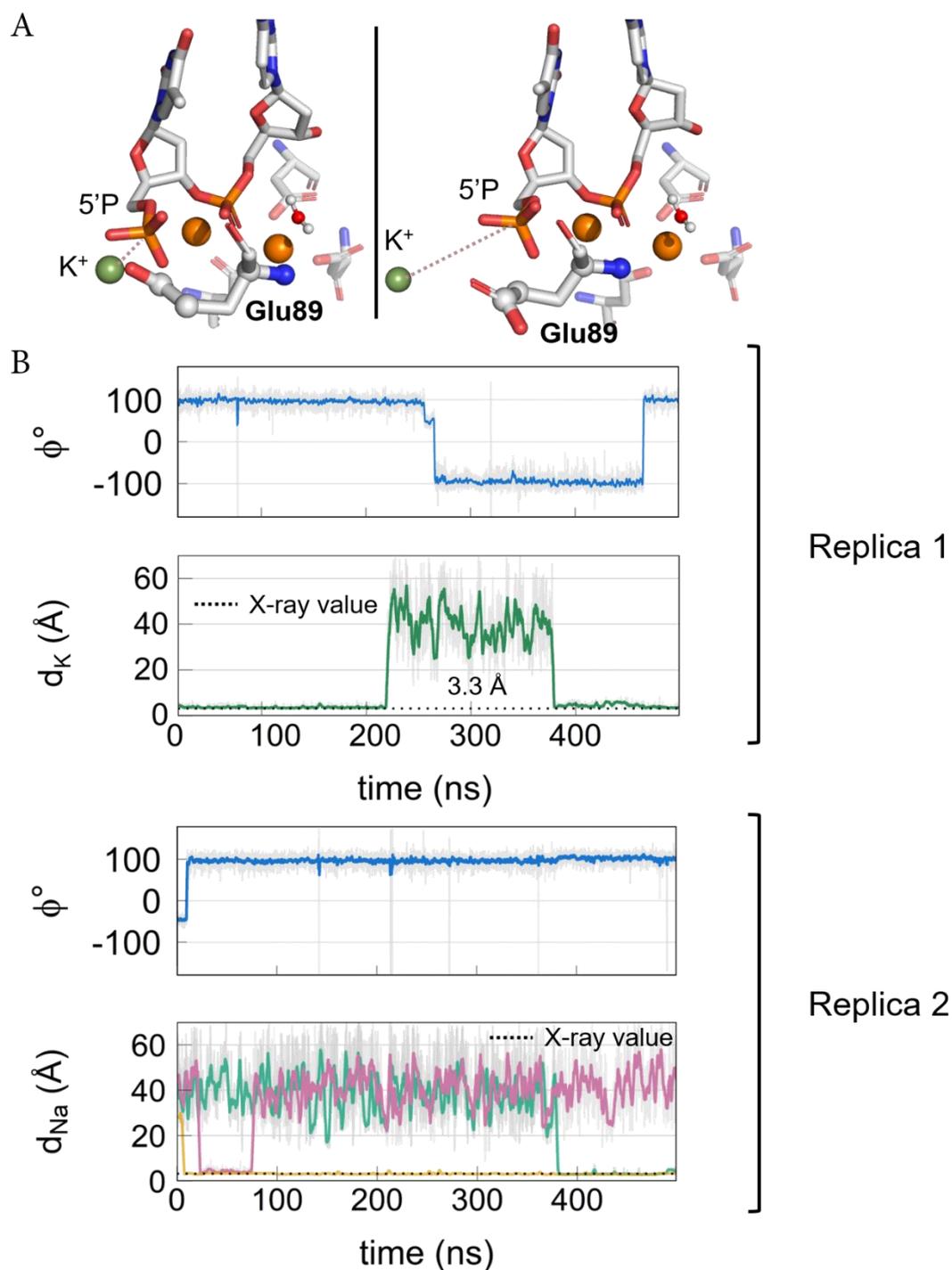


Figure A.7(A) Schematic representation of the distance between K^+ ion from the solvent bulk and the terminal 5' phosphorous concomitant with the inner/outer flipping observed during the RS_{2M} (Replica1) trajectory. (B) Graphs of time evolution of the pseudo dihedral angle (ϕ) taken along N, $C\alpha$, $C\delta$, $C\gamma$ bonds of Glu89, which describes the inner and the outer conformations and the representation of time evolution of the distance between the monovalent ion from the solvent bulk and the terminal 5' phosphorous. Data are taken from the RS_{2M} system.

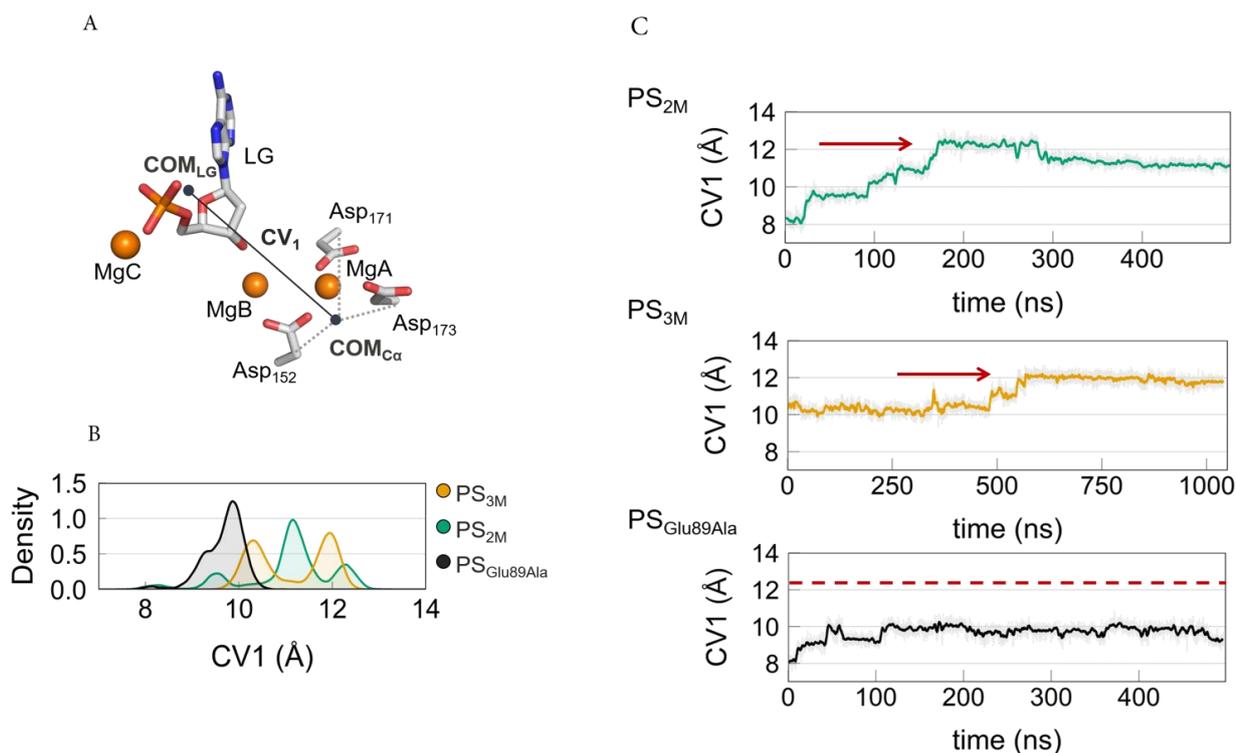


Figure A.8 CV1 measures the distance between the centre of mass (COM) of the heavy atoms of AMP and the COM of the $C\alpha$ of the aspartates in the first coordination shell of the two catalytic metal ions, i.e. Asp152, Asp171 and Asp173. (A) Schematic representation of CV1, exemplified using a snapshot taken from PS_{3M} simulations. (B) Frequency distribution of CV1 for PS_{2M} (green), PS_{3M} (orange), $PS_{Glu89Ala}$ (black). (C) Time evolution of CV1 for the Product State systems, PS_{2M} (green), PS_{3M} (orange), $PS_{Glu89Ala}$ (black).

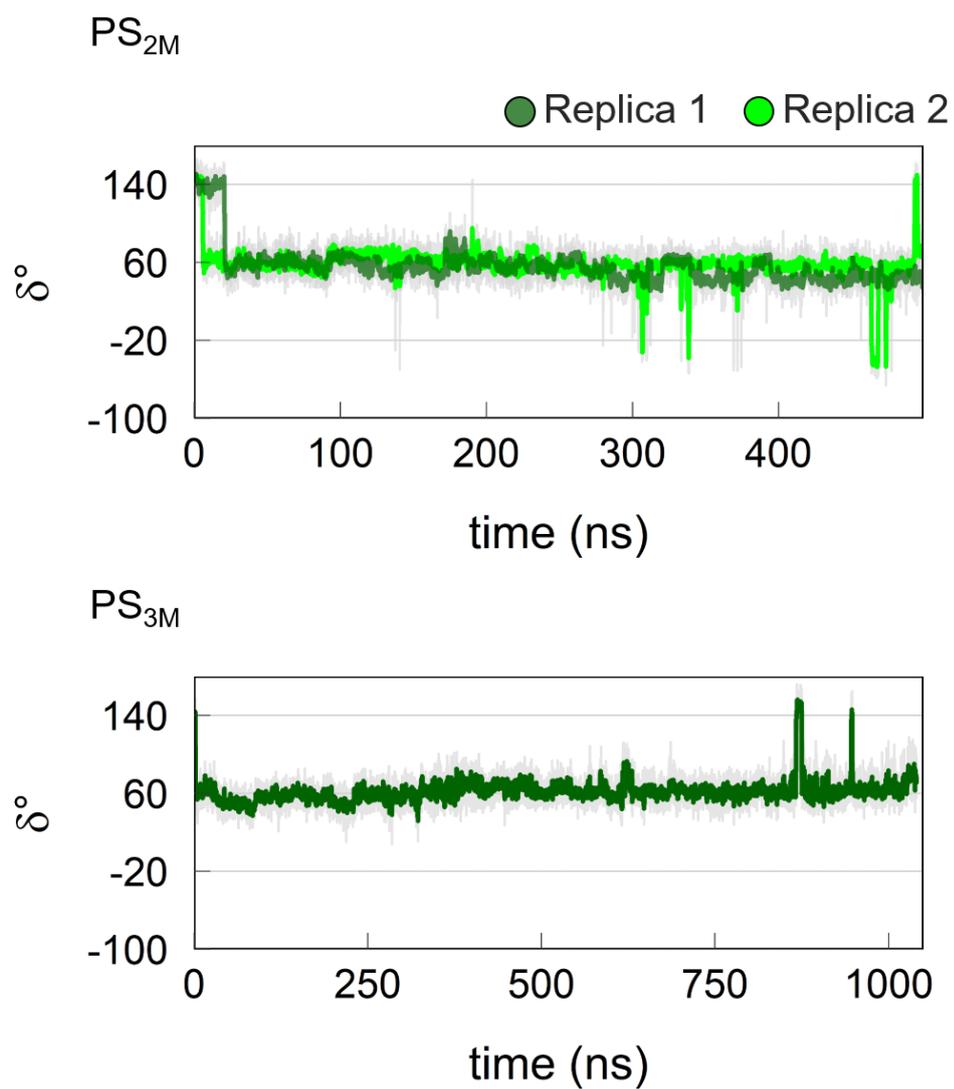


Figure A.9 Time evolution of dihedral angle (δ), taken along C, C α , C β , C γ bonds, for the guide residue His36 for the PS2M (top) and PS3M (bottom) systems.

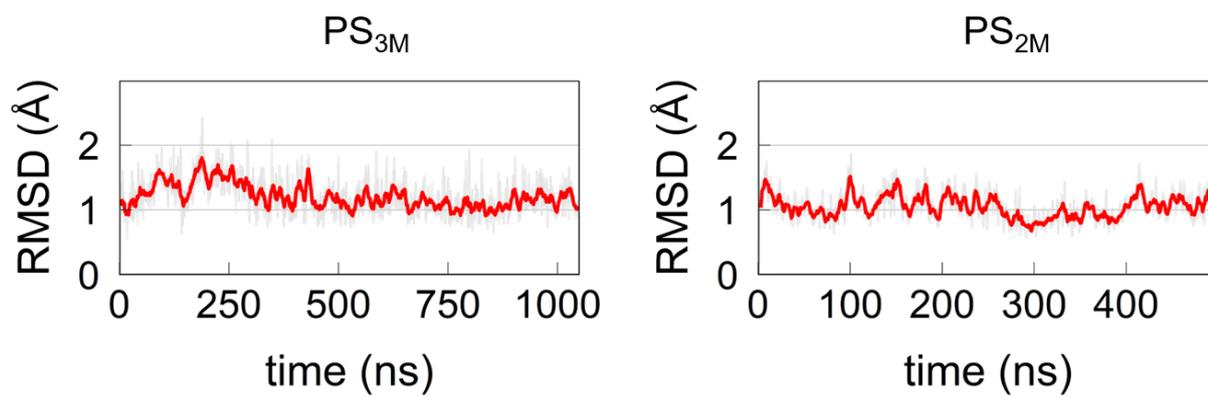


Figure A.10 Time evolution of root mean square deviation (RMSD) of residues 80-125 (i.e. comprising the mobile arch), for the PS_{3M} (left) and PS_{2M} (right) systems.

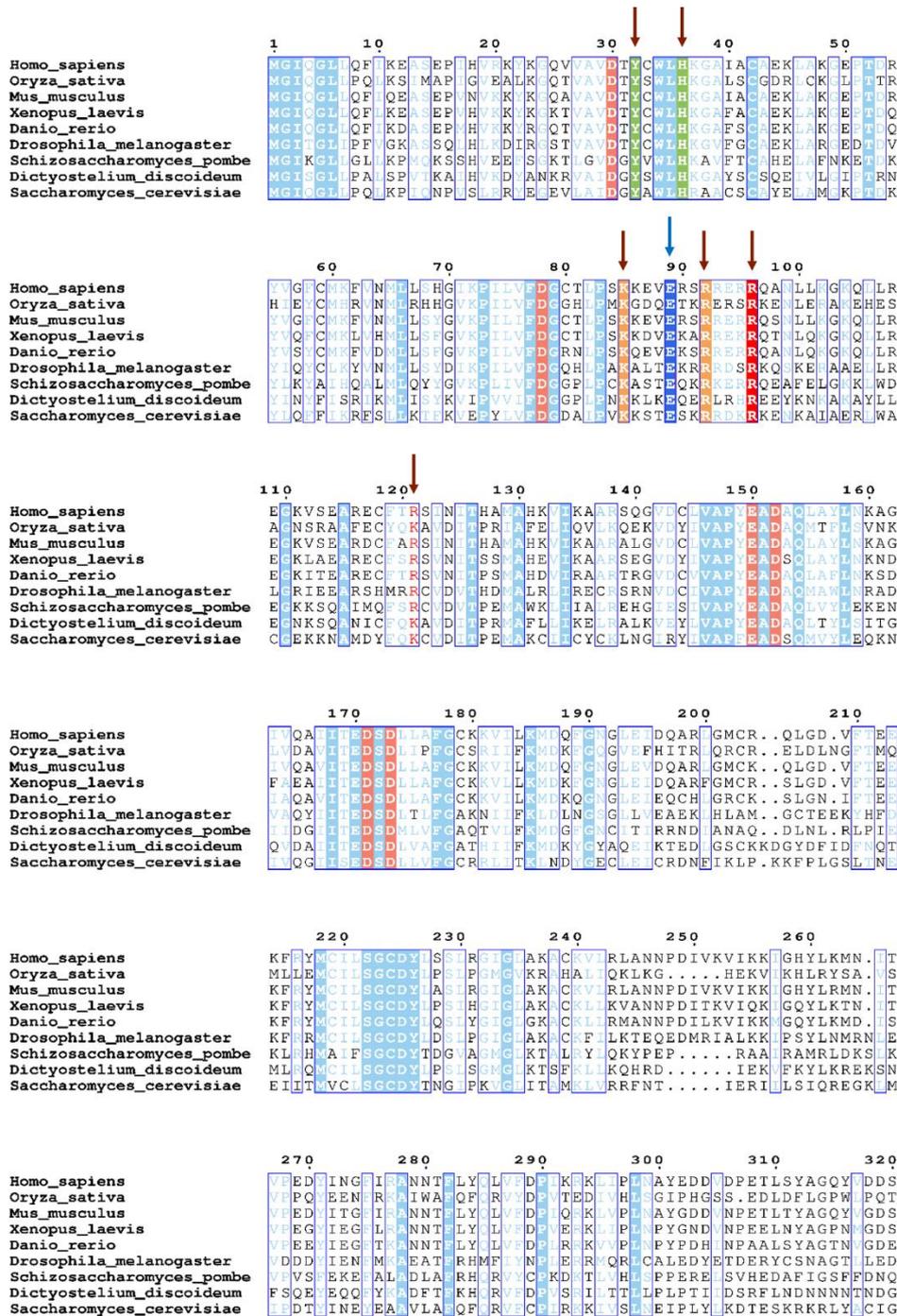


Figure A.11 Sequence alignments of Exonuclease1 from 9 different organisms. Glu89 (blue) is strictly conserved across exonucleases as for the basic residues in the first and second coordination shell of MgA and MgB (light red), catalytic residues Lys85 and Arg92 (orange), guide residues, Tyr32 and His36 (green) and some of the proposed ‘steering’ residues Arg96 and Arg121. Conserved key residues are marked with arrows.

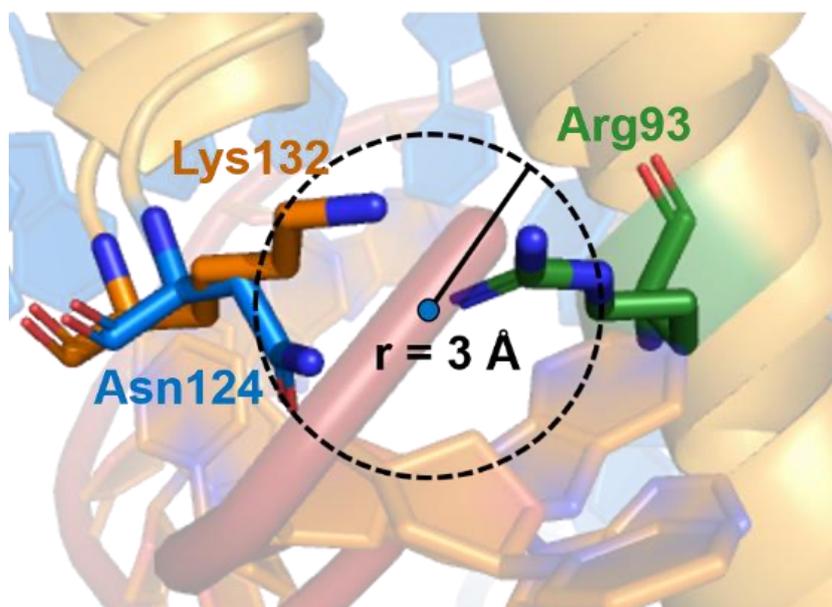


Figure A.12 Structure alignment of the hExo1 crystal structure (PDB ID 5V0E) with the hFEN1 crystal structure (PDB ID 5KSE). The terminal guanidine, amide and amine groups of Arg93, Asn124 (in hExo1) and Lys132 (in hFEN1) residues are within a sphere of $\sim 3 \text{ \AA}$.

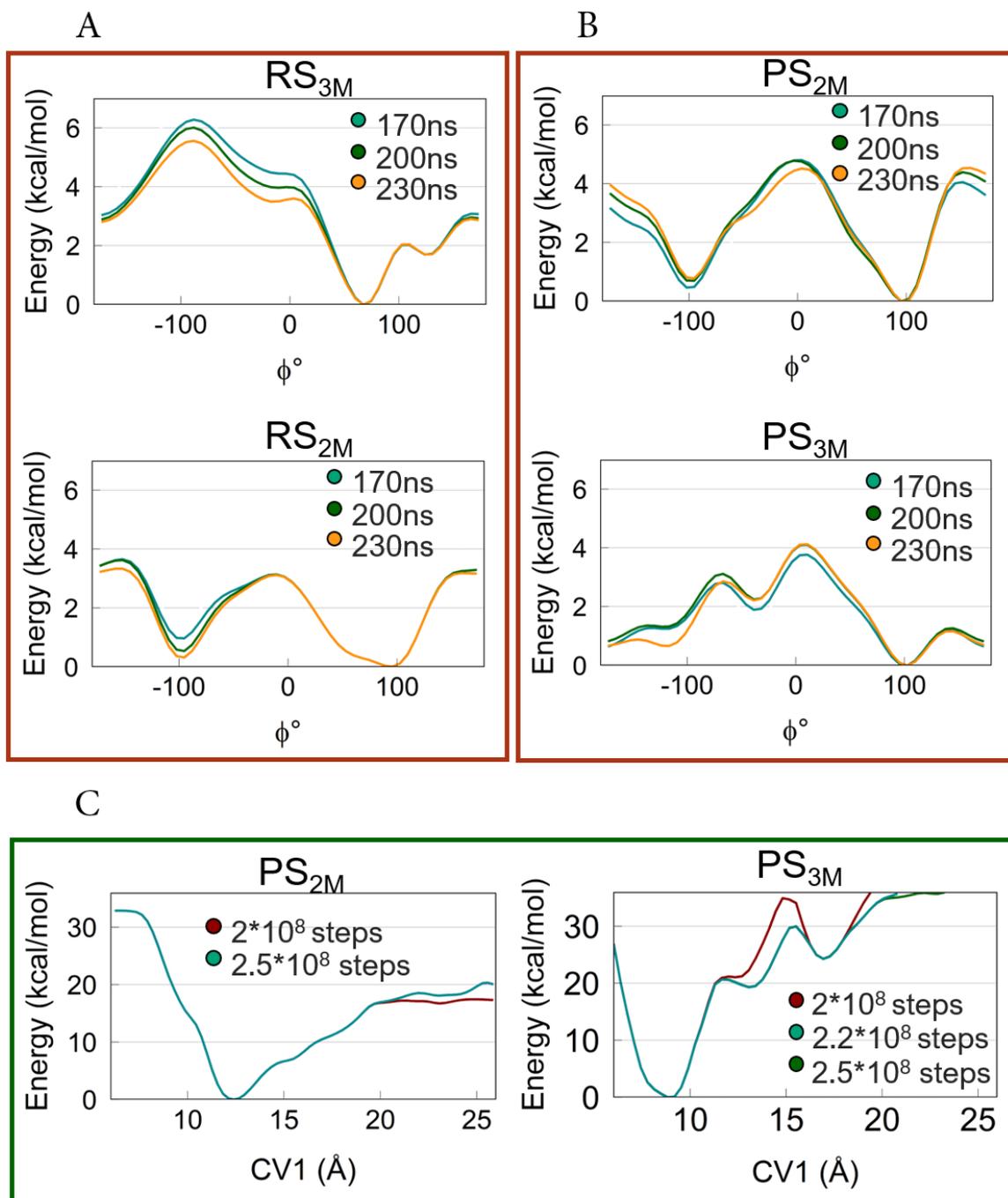


Figure A.13 Convergence of the well-tempered metadynamics simulations for the Glu89 inner/outer flipping in the Reactant States (A), and in the Product States (B). (C) Convergence of the confined well-tempered metadynamics simulations for the leaving group unbinding process. Convergence was checked considering the energy values as a function of time. From ~ 170 ns to ~ 230 ns, no significant changes in energy for Glu89 flipping were detected. From $\sim 2 \cdot 10^8$ steps to $\sim 2.5 \cdot 10^8$ steps, no significant changes in energy for the leaving group release were detected.

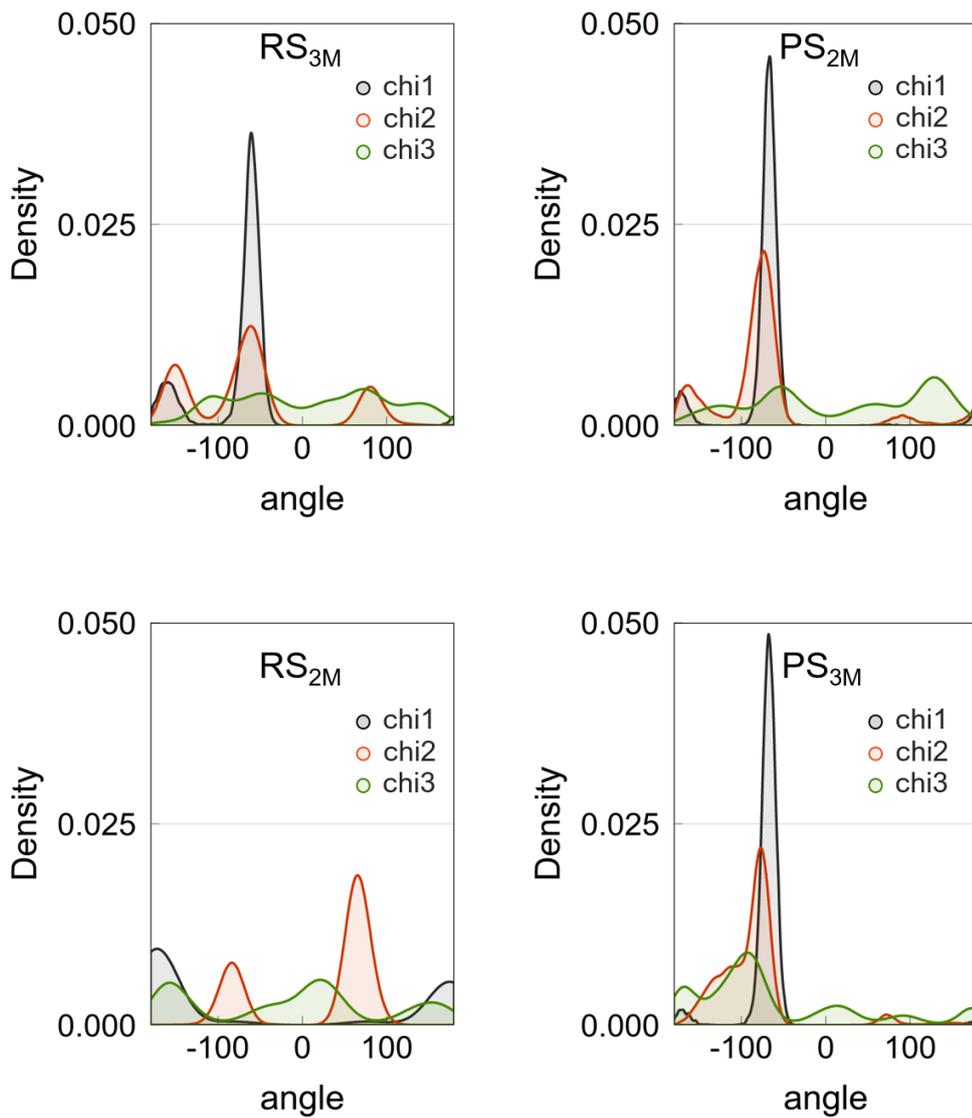


Figure A.14 Population of rotamers (χ_1, χ_2, χ_3) of Glu89 in RS_{3M}, RS_{2M}, PS_{3M} and PS_{2M}.

System name	Method	Catalytic state	*MgC	Mutation
<i>RS_{3M}</i>	Force-field MD	Reactant state	Present	NA
<i>RS_{2M}</i>	Force-field MD	Reactant state	Absent	NA
<i>PS_{2M}</i>	Force-field MD	Product state	Absent	NA
<i>PS_{3M}</i>	Force-field MD	Product state	Present	NA
<i>RS_{Glu89Ala}</i>	Force-field MD	Reactant state	Absent	Glu89Ala
<i>PS_{Glu89Ala}</i>	Force-field MD	Product state	Absent	Glu89Ala
<i>RS_{3M}</i>	Well-tempered Metadynamic	Reactant state	Present	NA
<i>RS_{2M}</i>	Well-tempered Metadynamic	Reactant state	Absent	NA
<i>PS_{2M}</i>	Well-tempered Metadynamic	Product state	Absent	NA
<i>PS_{3M}</i>	Well-tempered Metadynamic	Product state	Present	NA
<i>PS_{3M}</i>	Confined Metadynamic	Product state	Present	NA
<i>PS_{2M}</i>	Confined Metadynamic	Product state	Absent	NA

*Present/Absent is referring to the starting point, when the minimization step starts.

Table A.1 The table summarizes all the simulations performed. Multiple replicas per system were run, as specified in the results section of the manuscript.

Appendix B

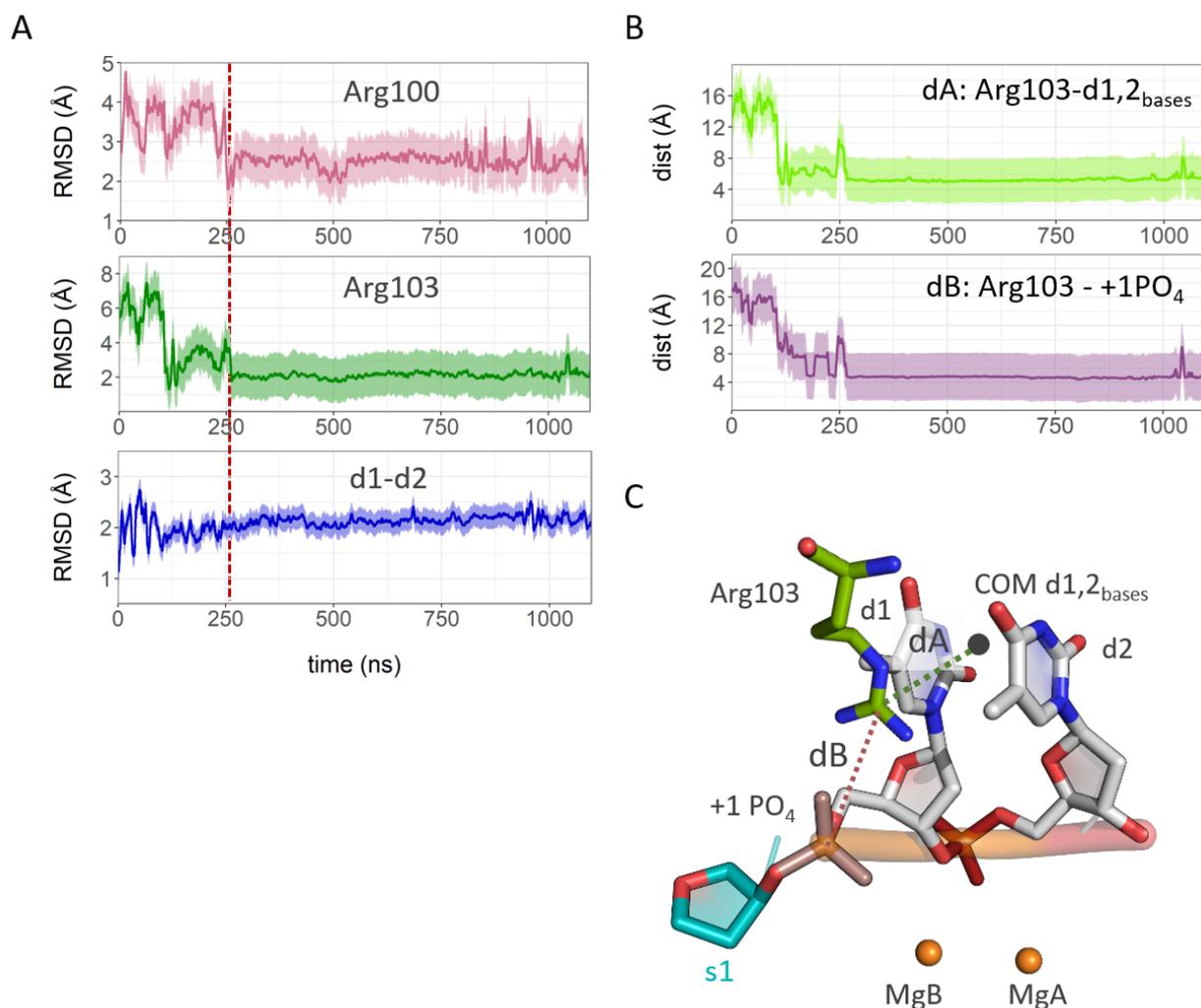


Figure B.1 MD simulation of the wt system. (A) Time evolution of the root mean square deviation (RMSD) for the heavy atoms of Arg100, Arg103 and nucleotides d1 and d2, using as reference the reactant state crystal structure (PDB ID 5UM9). The red dashed line marks the reorganization of Arg100 and Arg103. (B) Time evolution of distances dA and dB. dA is the distance between the center of mass (C.O.M.) of the heavy atoms of the guanidinium group of Arg103 and the C.O.M. of the heavy atoms of both d1 and d2 nucleobases. dB is the distance between the C.O.M. of the heavy atoms of guanidinium group of Arg103 and the C.O.M. of the +1 phosphate. After ~250ns dA and dB stabilize around 4.5 Å and 5 Å, respectively. (C) Schematic representation of nucleotides d2, d1 and main chain atoms of s1 together with the +1 phosphate (highlighted in pink), the Arg103 residue (in licorice, colored in green), the catalytic Mg ions (as spheres) and the distances dA and dB (dashed lines). The C.O.M. of the nucleobases of d1 and d2 is represented as a dark grey sphere.

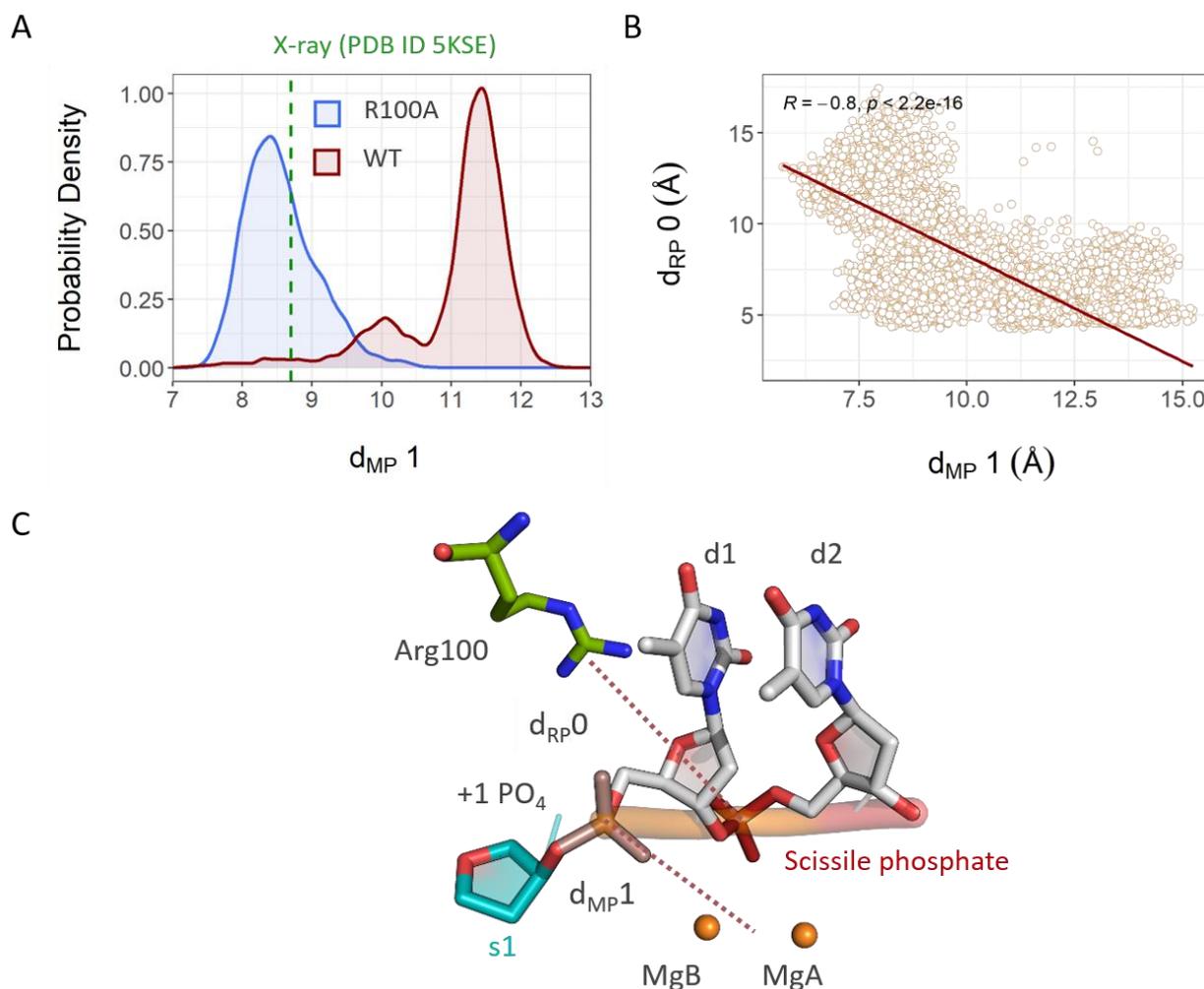


Figure B.2 (A) Probability density of the distance d_{MP1} for the wt and R100A systems during MD simulations. d_{MP1} is the distance between the C.O.M. of the +1 phosphate and the C.O.M. of the catalytic metal ions MgA and MgB. The green dashed line marks the crystallographic value of d_{MP1} (PDB ID 5KSE) where Arg100 is mutated to Ala. (B) Scatter plot of distances d_{MP1} and d_{RP0} (where d_{RP0} represents the distance between the C.O.M. of the heavy atoms of guanidinium group of Arg100 and the C.O.M. of the scissile phosphate) collected from the MD simulations of the wt and RKKK-A/-E systems. The correlation has been calculated using Pearson's method. (C) Schematic representation of nucleotides d2, d1 and main chain atoms of s1, together with the +1 and the scissile phosphates (highlighted in pink and red, respectively), Arg100 (in green licorice), the catalytic Mg ions (as orange sphere). The distances d_{MP1} and d_{RP0} are shown as dashed lines.

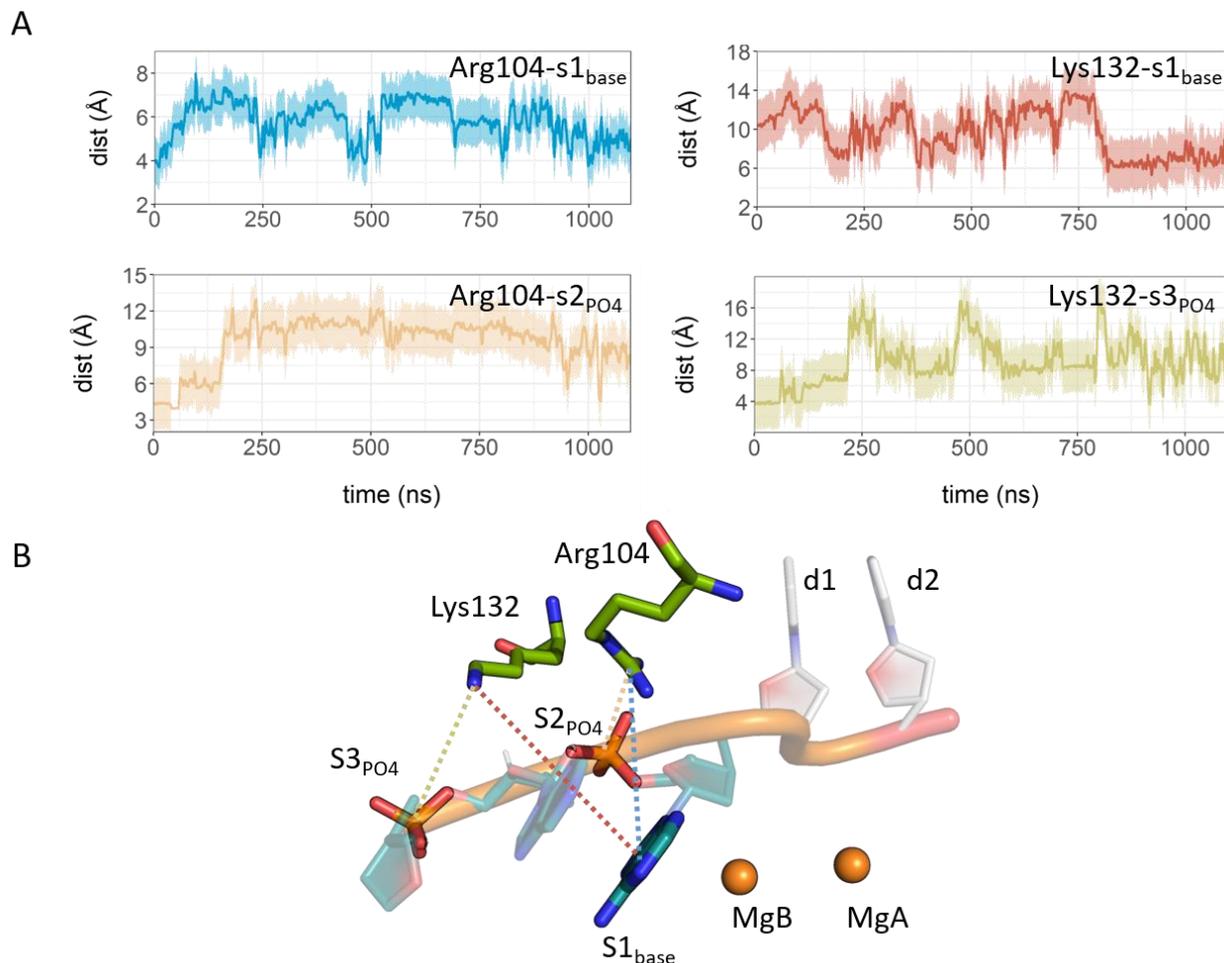


Figure B. 3 (A) On the left, the distance between the C.O.M. of the guanidinium group of Arg104 and the C.O.M. of the nucleobase of nucleotide s1 (blue line), and the distance between the C.O.M. of the guanidinium group of Arg104 and the C.O.M. of the phosphate between nucleotides s1 and s2 (orange line). On the right, the distance between the C.O.M. of the amino group of Lys132 and the C.O.M. of the nucleobase of nucleotide s1 (red line), and the distance between the C.O.M. of the amino group of Lys132 and the C.O.M. of the phosphate between nucleotides s2 and s3 (yellow line). (B) Representative snapshot from the MD simulation of the wt system showing Arg104, Lys132 (green licorice), nucleotides d2 to s3 along the 5' strand of DNA and the two catalytic metal ions (as orange spheres).

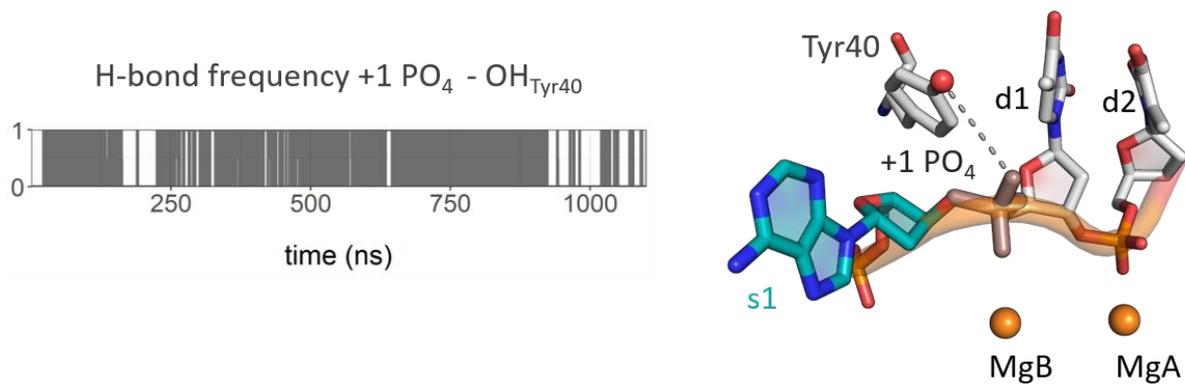
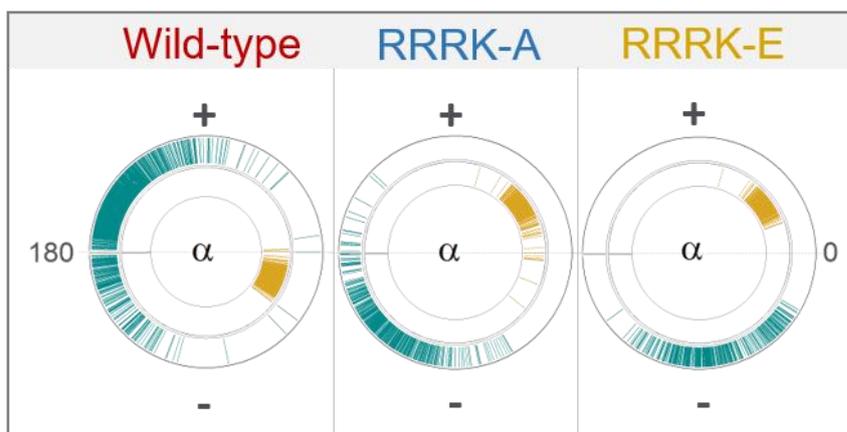


Figure B.4 MD simulations of the wt system. On the left, occupancy of the H-bond between the OH group of Tyr40 side chain and the +1 phosphate. The requirements for the formation of the H-bond were: distance donor-acceptor $< 3.5 \text{ \AA}$ and the cutoff angle $< 25^\circ$. On the right, representative snapshot showing Tyr40 (licorice, with the OH group as red sphere), nucleotides d2, d1 and s1 along the 5'-strand of DNA and the two catalytic metal ion (as orange spheres).

A



B Snapshot from MD simulation (RRRK-E system)

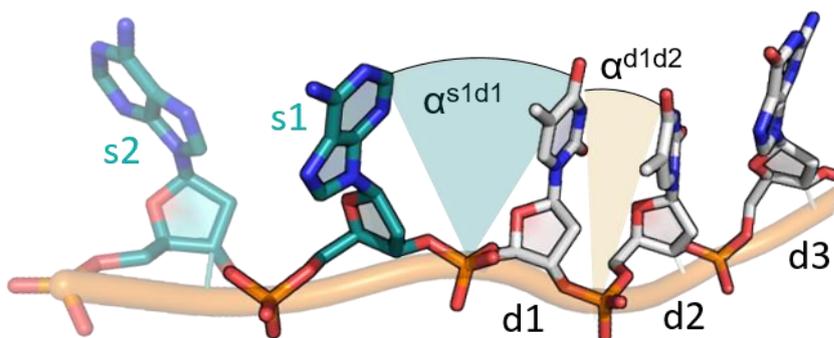
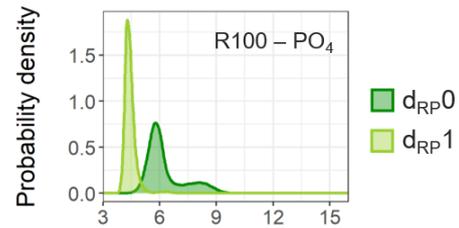
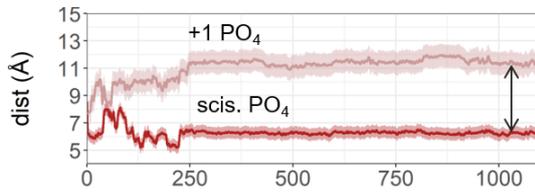


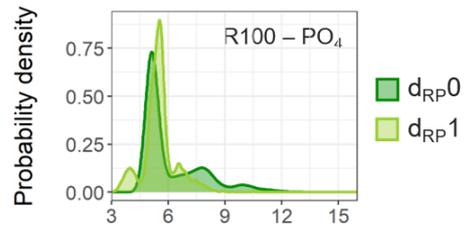
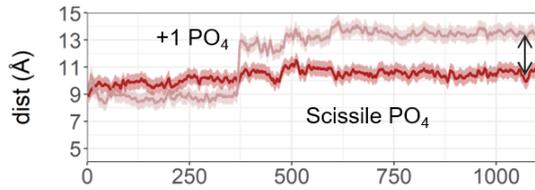
Figure B.5 A) Representation of the α^{s1d1} , α^{d1d2} dihedral angles calculated during MD simulations of the wt, RRRK-A and RRRK-E systems, defined as the angle between the planes of the nucleobases of nucleotides d1 and s1 (α^{s1d1}), and d1 and d2 (α^{d1d2}) within the 5' strand. B) Representative snapshot from the MD simulation of the RRRK-E system showing the non-inverted orientation of nucleotides d3 to s2. The dihedral angles α^{s1d1} and α^{d1d2} are also represented.

A

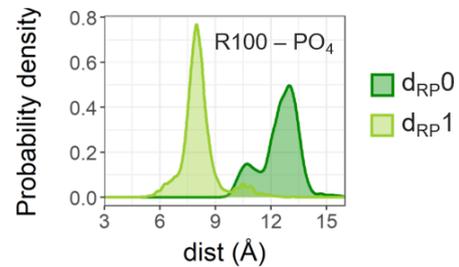
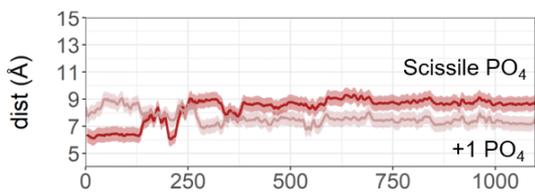
Wild-type



RRRK-A



RRRK-E



time (ns)

dist (Å)

B

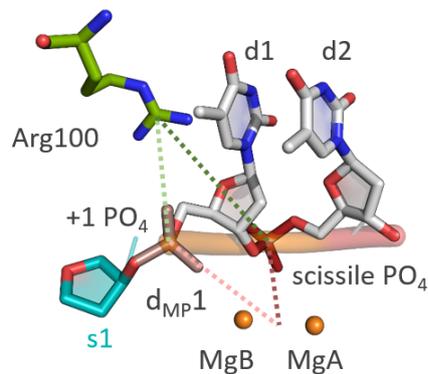


Figure B.6 (A) On the left, time evolution of distances d_{MP0} (red line) and d_{MP1} (pink line); on the right, probability density of d_{RP0} (green line) and d_{RP1} (chartreuse line) from MD simulations of the wt, RRRK-A and RRRK-E systems. d_{MP0} and d_{MP1} are the distances between the C.O.M. of the catalytic MgA and MgB ions and the C.O.M. of the scissile and +1 phosphates, respectively. d_{RP0} and d_{RP1} are the distances between the C.O.M. of the guanidinium group of Arg100 and the C.O.M. of the scissile and +1 phosphates, respectively. (B) Representation of nucleotides d2, d1 and s1 together with the +1 and scissile phosphates (highlighted in pink and red, respectively), Arg100 (green licorice), the catalytic Mg ions (as orange sphere) and the distances d_{RP0} , d_{RP1} , d_{MP0} and d_{MP1} .

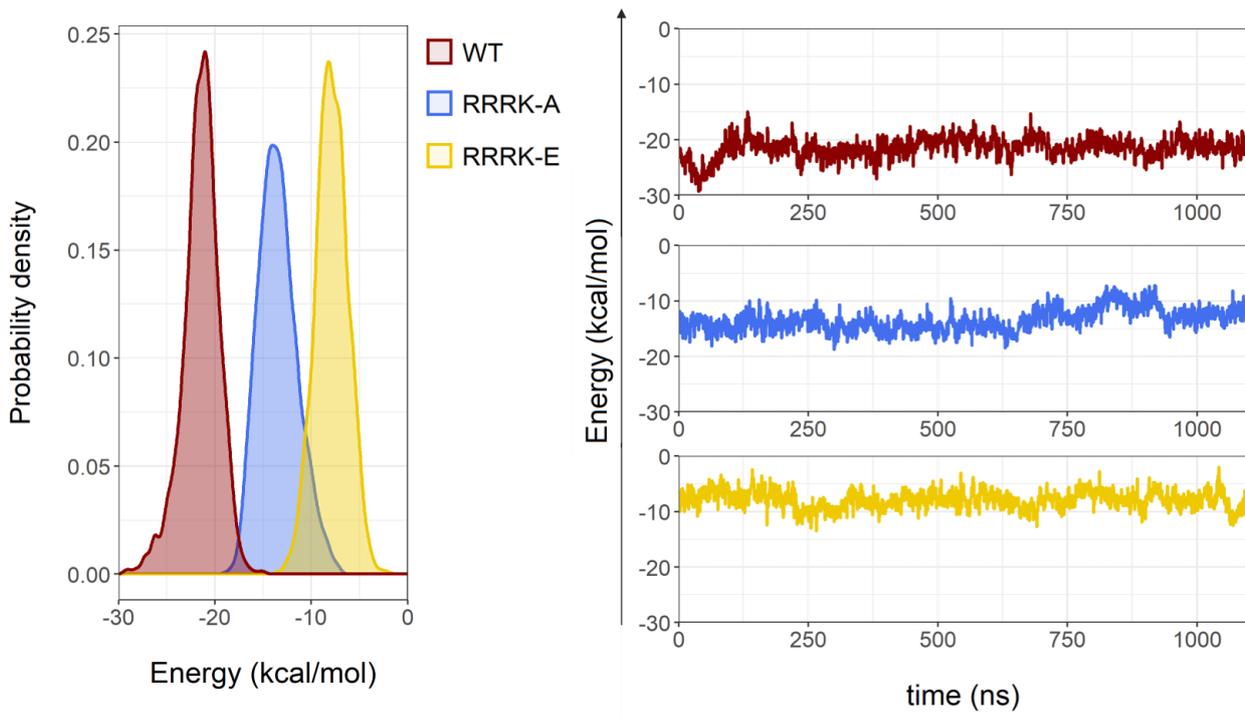


Figure B.7 (Left) Probability density of the Debye-Huckel interaction energy between the 5' flap DNA substrate and the wt and RRRK-A/-E hFEN1 enzymes. The Debye-Huckel interaction energy was calculated using Plumed2.6, setting the temperature to 310K, the ionic strength to 1M and the dielectric constant of the solvent to 80. (Right) Time evolution of the Debye-Huckel interaction energy during the simulations of the wt (dark red line), RRRK-A (blue) and RRRK-E (yellow) systems.

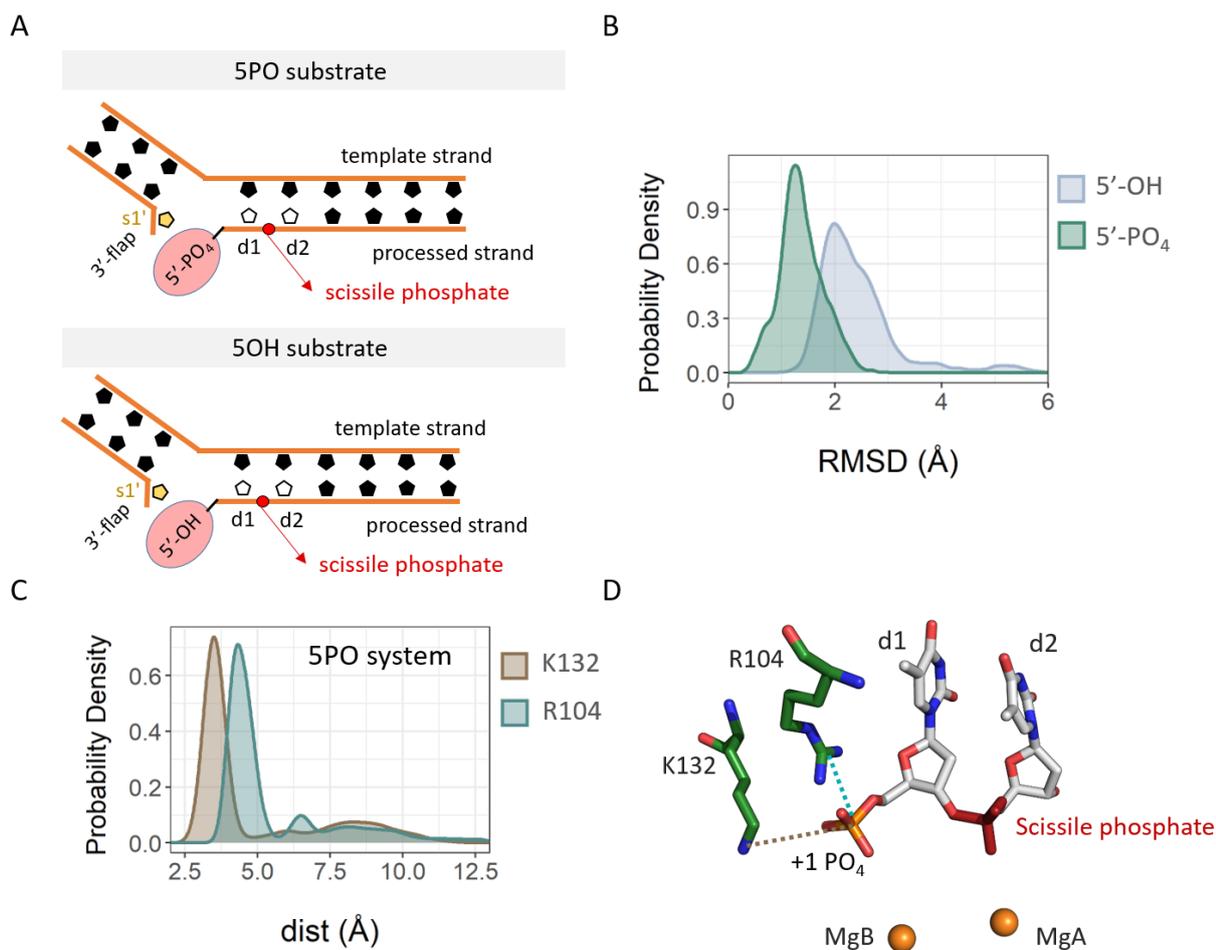


Figure B.8 (A) Schematic representation of the exonucleolytic DNA substrates (i.e. without the 5'-flap), ending with either a 5'-monophosphate (top) or 5'-OH group (bottom). They correspond to the DNA substrates of the 5PO and 5OH systems. The scissile phosphate as for the optimal double-flap DNA substrate is indicated with a red arrow. (B) Probability density of the RMSD values for the heavy atoms of the terminal 5'-nucleotide during the simulations of the 5PO and 5OH systems. (C) Probability density of the distance between the C.O.M. of the +1 phosphate (i.e. the terminal 5'-phosphate) and the C.O.M. of the heavy atoms of the guanidinium group of Arg104 (cyan line) and the distance between the C.O.M. of the +1 phosphate and the C.O.M. of the heavy atoms of the amino group of Lys132 (brown). (D) Representative snapshot from the MD simulation of the 5PO system showing Arg104, Lys132 (in green licorice), nucleotides d1 and d2 along the 5'-strand of DNA and the two catalytic metal ions (as orange spheres).

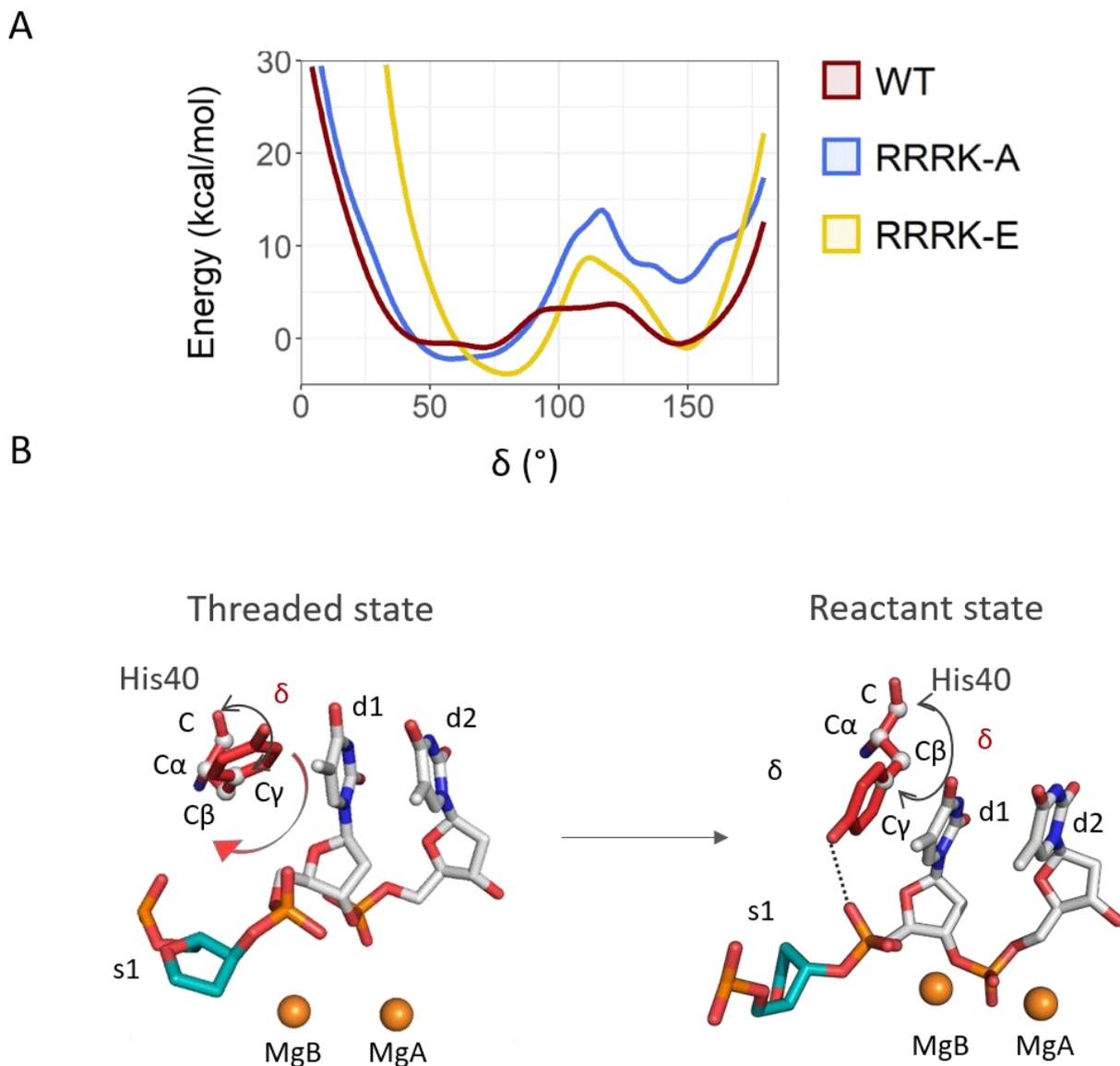


Figure B.9 (A) Free energy profile (in kJ mol^{-1}) corresponding to the dihedral angle (δ) for the wt, RRRK-A/-E systems, as reconstructed from the pMtD simulations reweighting the histograms of the sampled distributions and converting them into free energy values, as implemented in Plumed. (B) Representation of the two conformations assumed by Tyr40 side chain and corresponding to the two minima in the free energy profiles shown in (A). On the left, Tyr40 is not able to form H-bond interactions with the phosphate, while on the right (i.e. in the reactant state) Tyr40 side chain has rotated and thus can interact with the phosphate (dashed black line). The atoms taken into account to calculate the dihedral angle δ (i.e. C, C α , C β and C γ) are shown as white spheres.

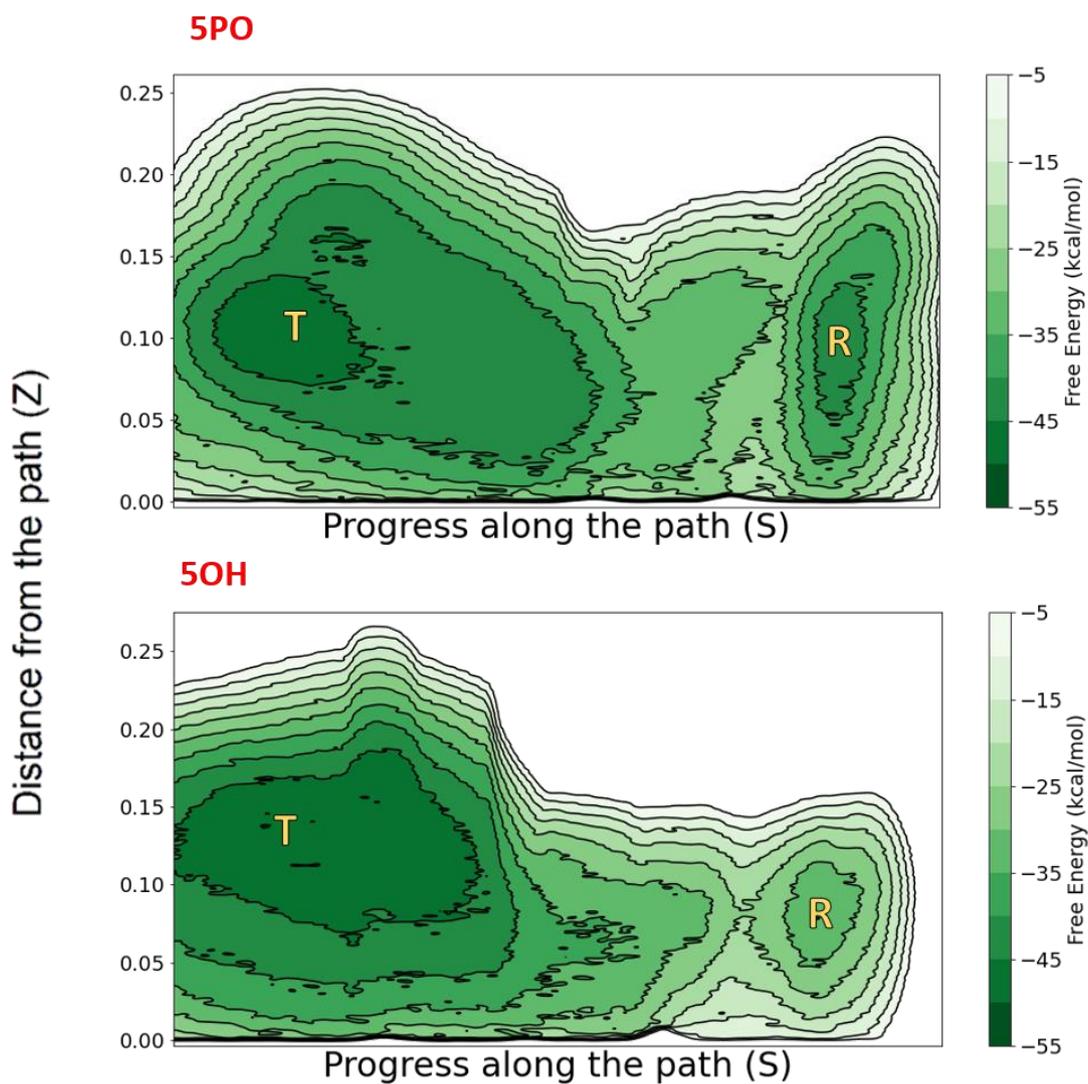


Figure B.10 pMtD free-energy landscape of the 5PO and 5OH systems, where the threaded (T) and reactant (R) states are indicated.

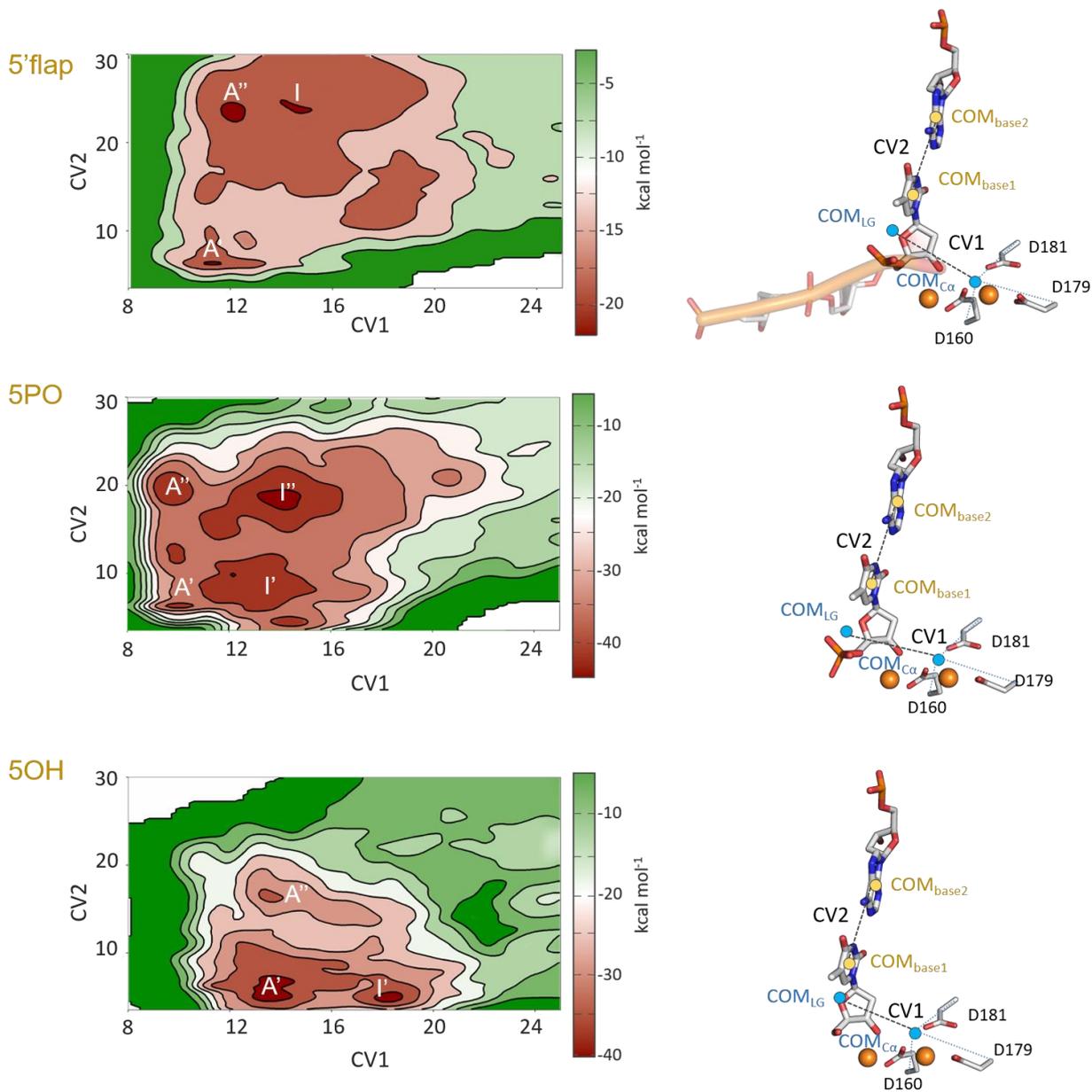


Figure B.11 Free energy landscape for the leaving group release in the 5'-flap, 5PO and 5OH systems from confined well-tempered metadynamics simulations. The minima are indicated with white letters. A representation of the collective variables CV1 and CV2 is also shown for the three systems. CV1 is the distance between the C.O.M. of the heavy atoms of the nucleotide leaving group and the C.O.M. of the C α of Glu160, Asp179, Asp181 in the first coordination shell of ions MgA and MgB (blue circles). CV2 is the distance between the C.O.M. of the nucleobase of d1 and the complementary nucleotide (yellow circles).

Convergence of the pMtD

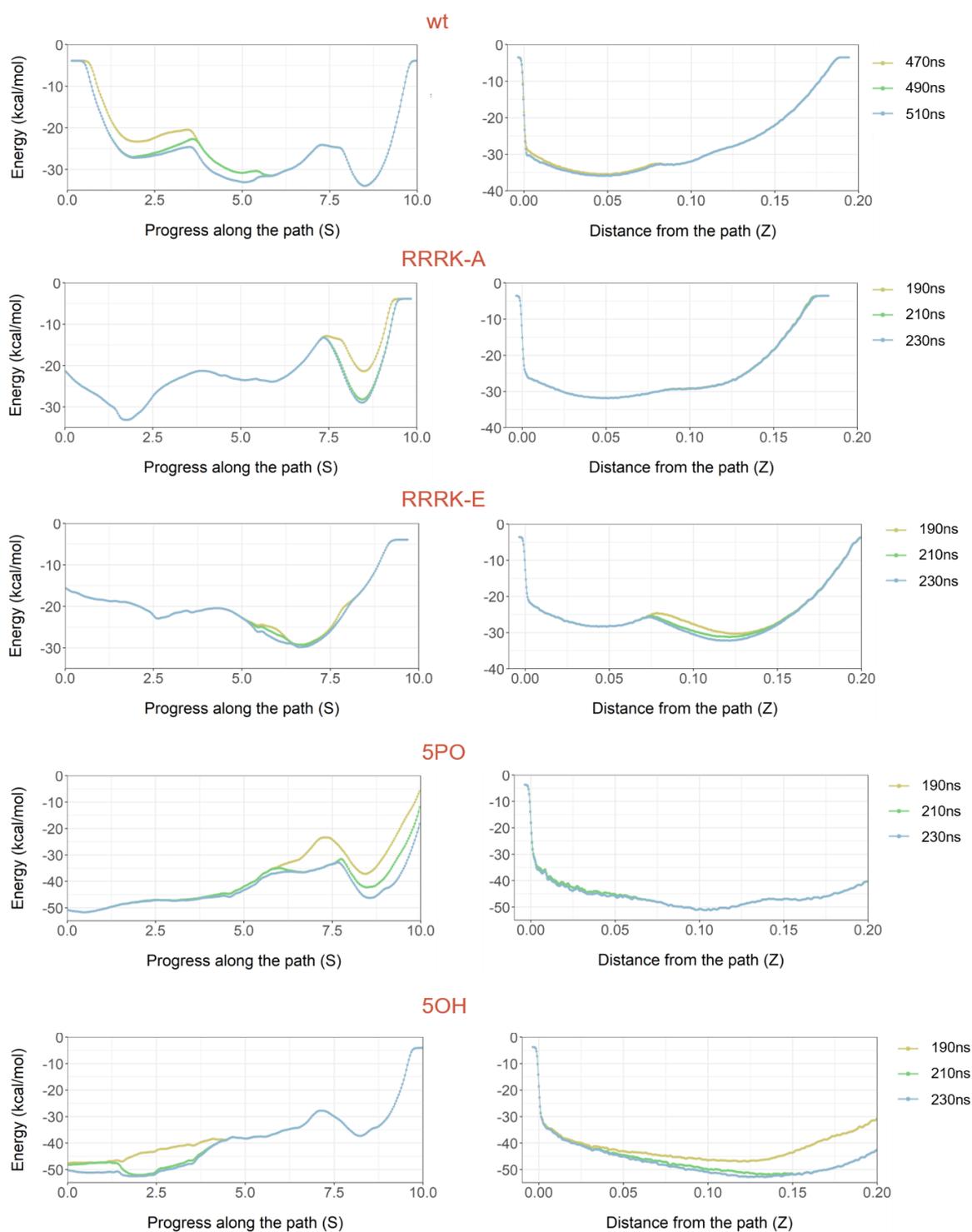


Figure B.12 Convergence of the free energy profiles from the pMtD simulations of the wt, RRRK-A/-E, 5PO and 5OH systems. The projections of the free energy on the path variables S (left panels) and Z (right panels) are shown during the progress of the metadynamics simulations.

Convergence of the leaving group departure

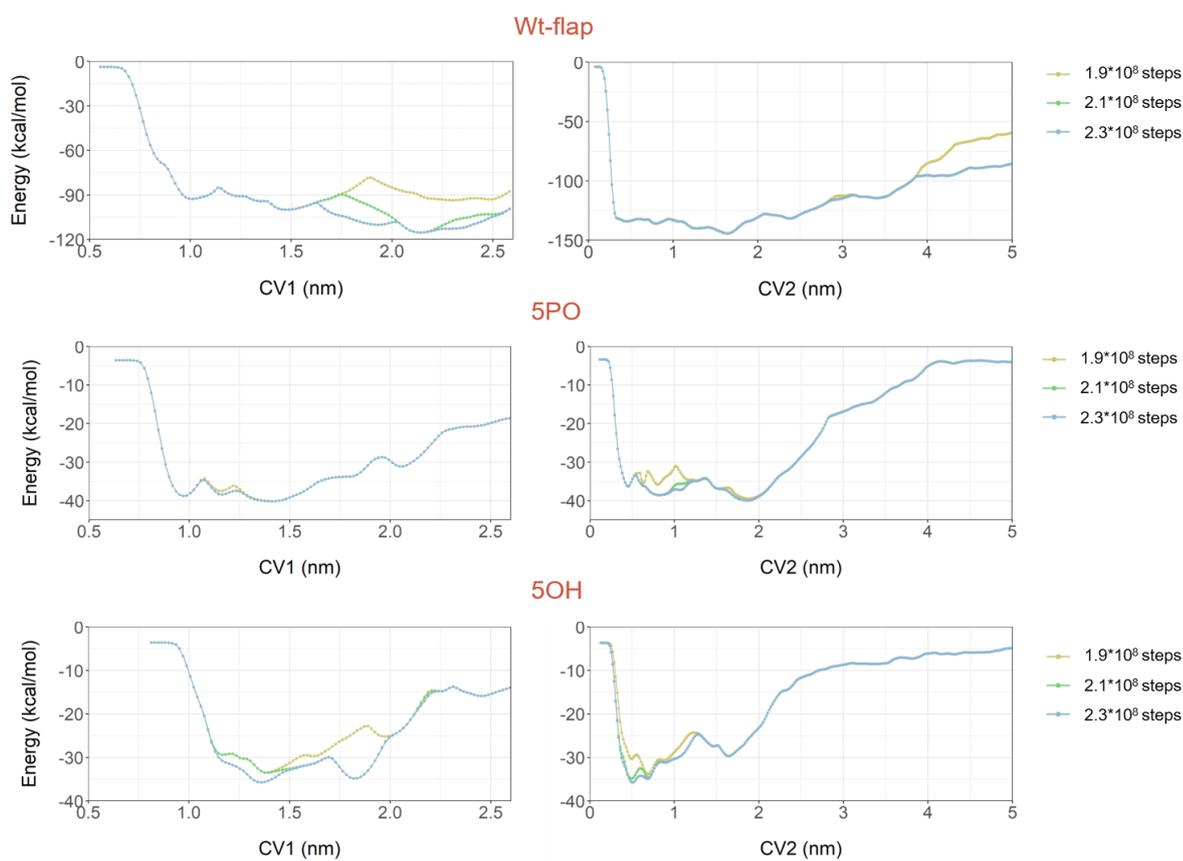


Figure B.13 Convergence of the free energy profiles for the release of the leaving group from metadynamics simulations of the 5'-flap, 5PO and 5OH systems. Free energy profiles projected on CV1 (distance between the C.O.M. of the heavy atoms of the nucleotide leaving group and the C.O.M. of the C α of Glu160, Asp179, Asp181 - left panels) and on CV2 (distance between the C.O.M. of the nucleobases of d1 and the complementary nucleotide - right panels) are shown during the progress of the metadynamics simulations.

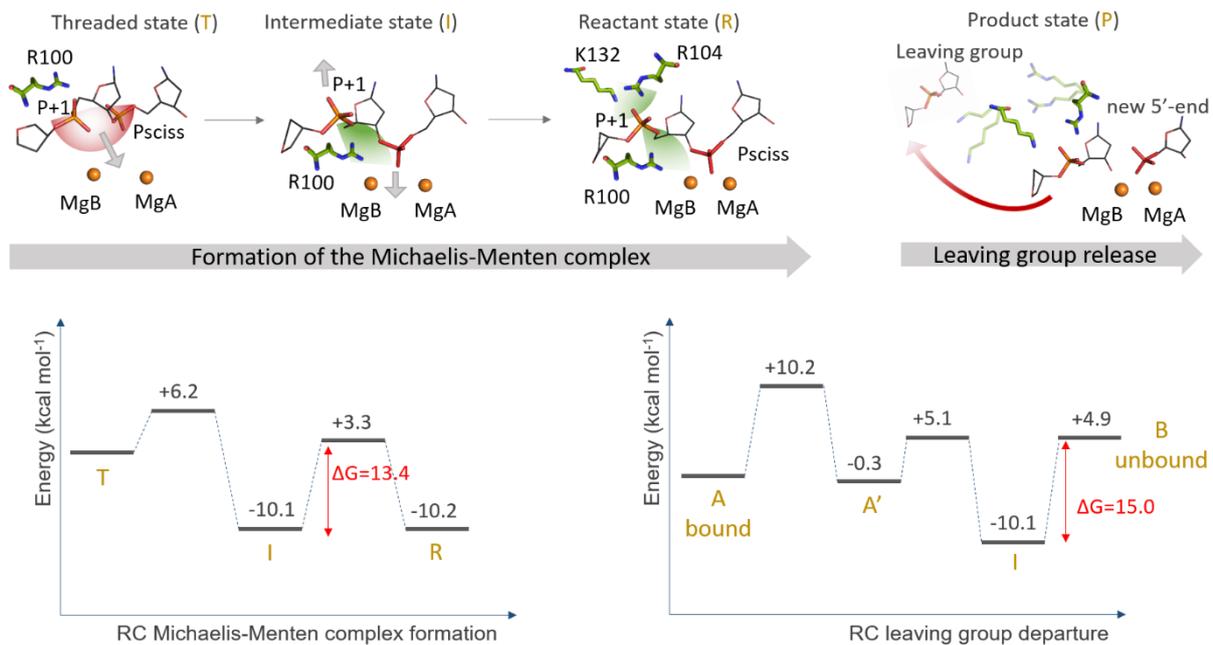


Figure B.14 Energy profiles of the formation of the Michaelis-Menten complex and leaving group departure for the wt hFEN1-5' flap substrate system.

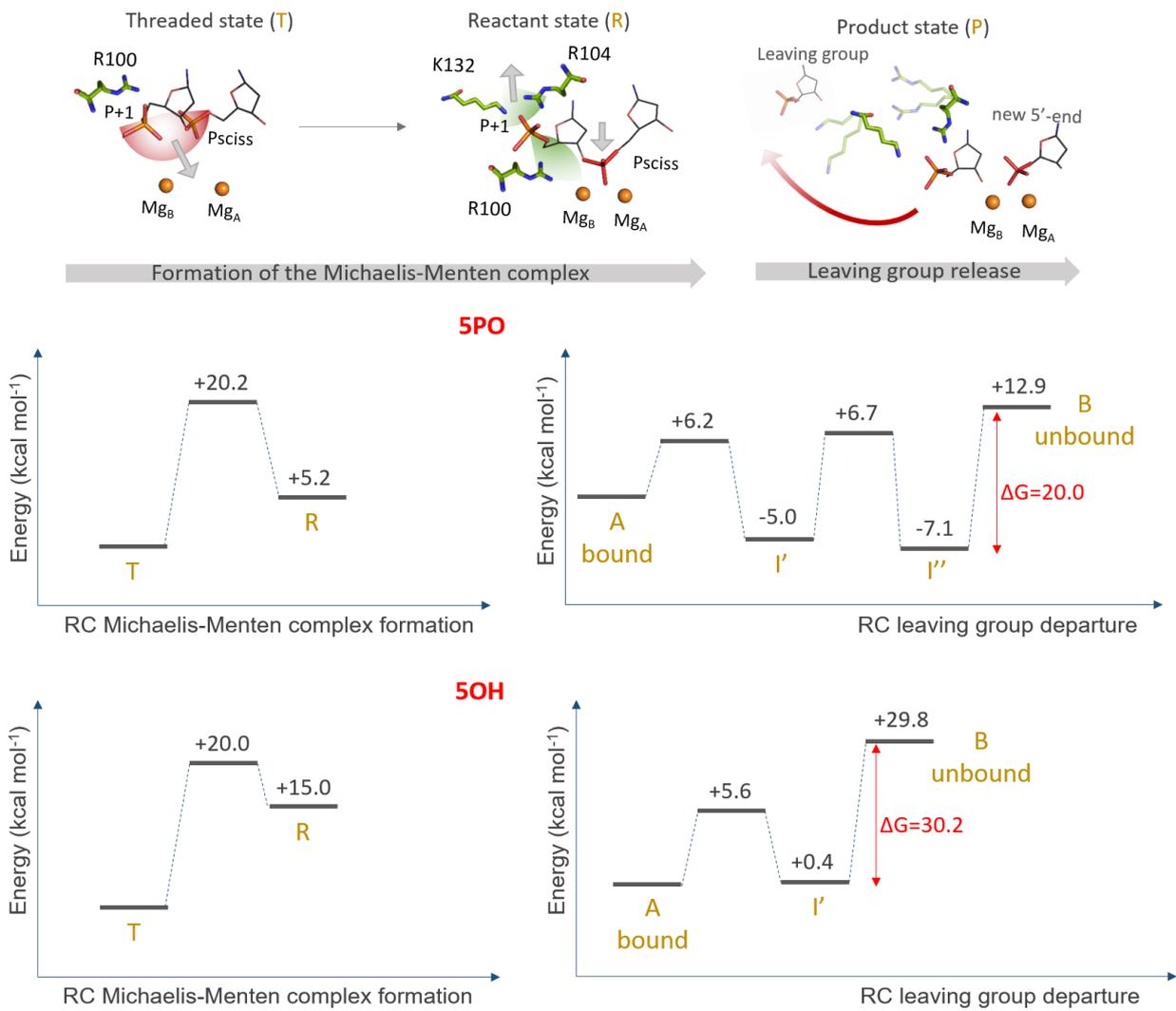
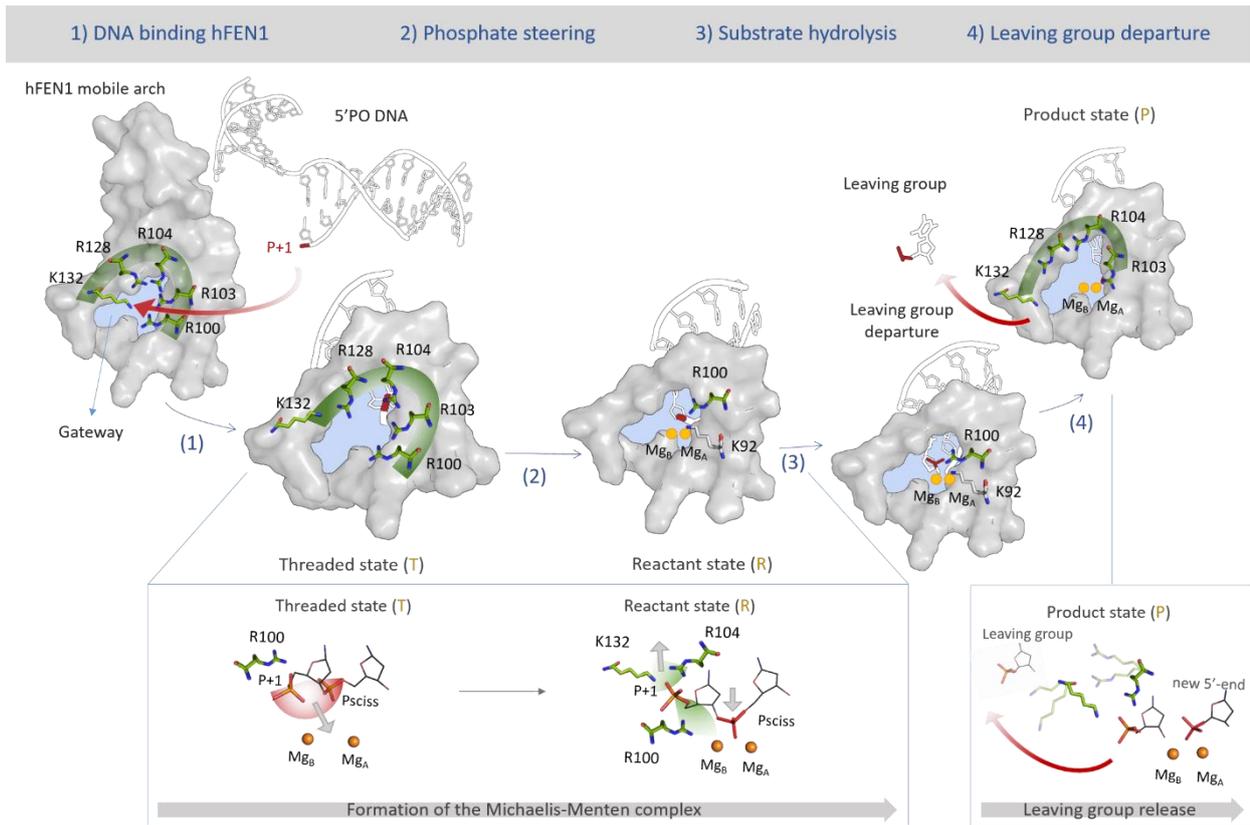


Figure B.15 Energy profiles for the formation of the Michaelis-Menten complex and leaving group departure for both the wt hFEN1-5'PO substrate system (5PO) and the wt hFEN1-5'OH substrate system (5OH).



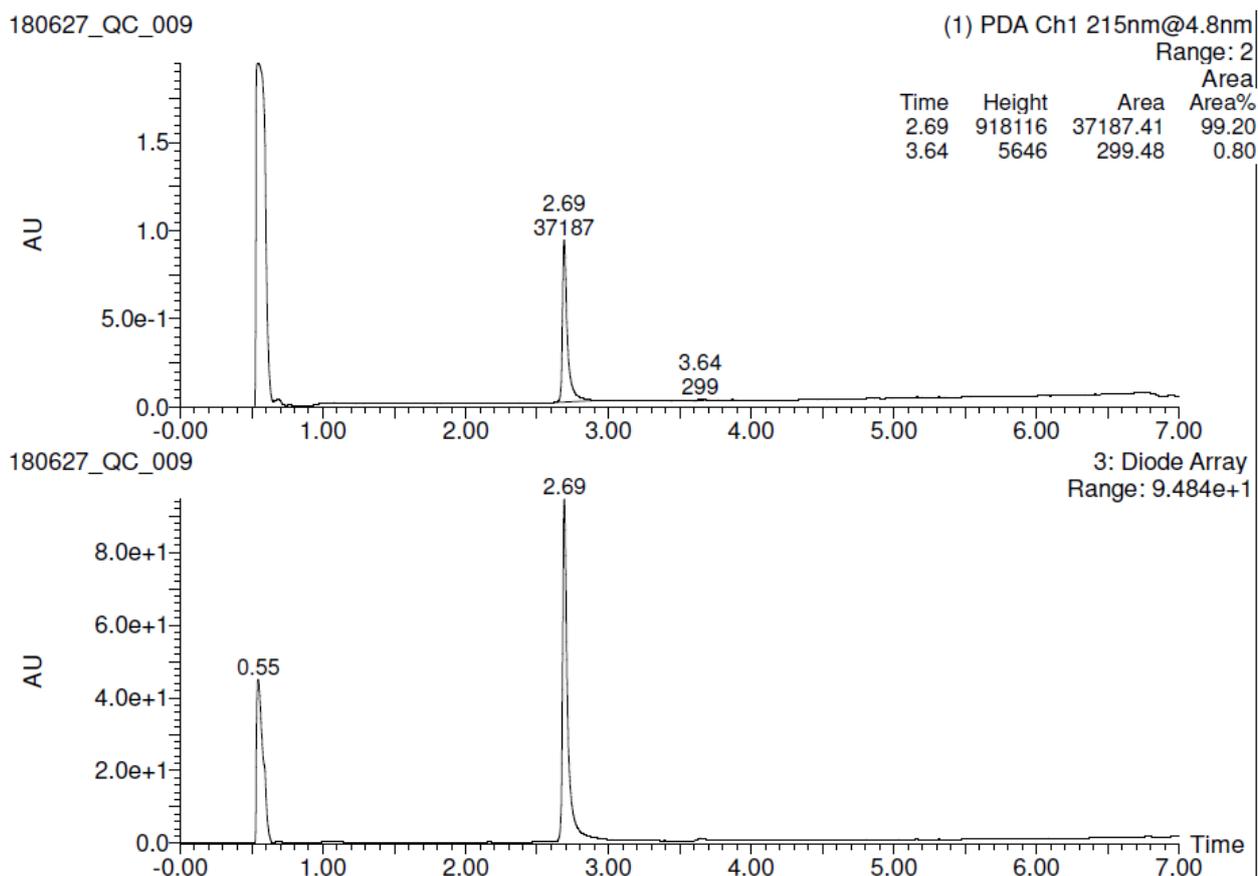
Scheme B.1 Schematic representation of the catalytic steps during hFEN1 catalysis on exonucleolytic DNA substrate (i.e. 5'PO). On the bottom left box, the molecular mechanism of phosphate steering residues Arg104 and Lys132 acting concertedly with Arg100 to promote the Michaelis-Menten complex formation. In the bottom right box, the molecular mechanism for leaving group departure assisted by phosphate steering residues.

Appendix C

Chromatography analysis of luteolin and quercetin

The chromatographic analyses were run on an ACQUITY UPLC BEH C18 (50x2.1mmID, particle size 1.7 μ m) with a VanGuard BEH C18 pre-column (5x2.1mmID, particle size 1.7 μ m) (LogD>1). The mobile phase was 10mM NH₄OAc in H₂O at pH 5 adjusted with AcOH (A) and 10mM NH₄OAc in CH₃CN-H₂O (95:5) at pH 5 (B). The mobile-phase B proportion increased from 10 % to 90 % in 7 min.

Luteolin



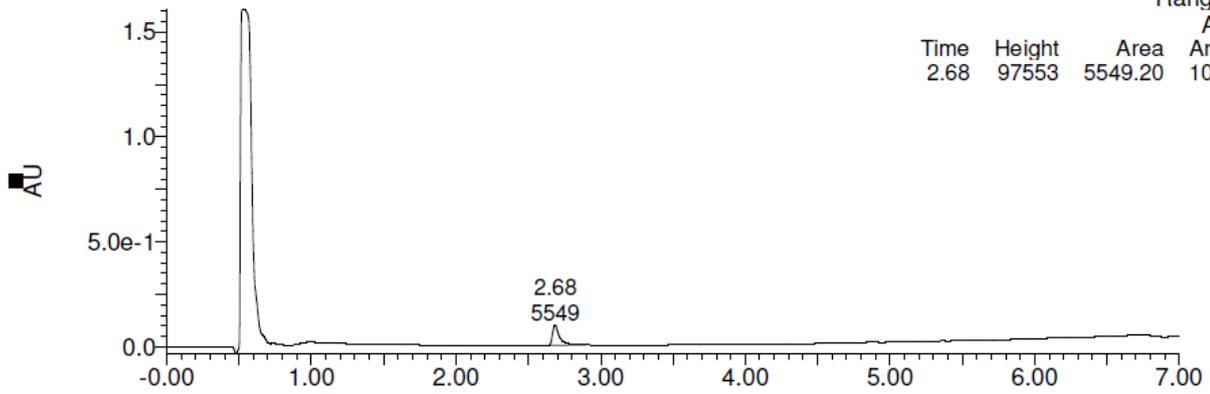
Quercetin

201012_QC_011

(1) PDA Ch1 215nm@4.8nm

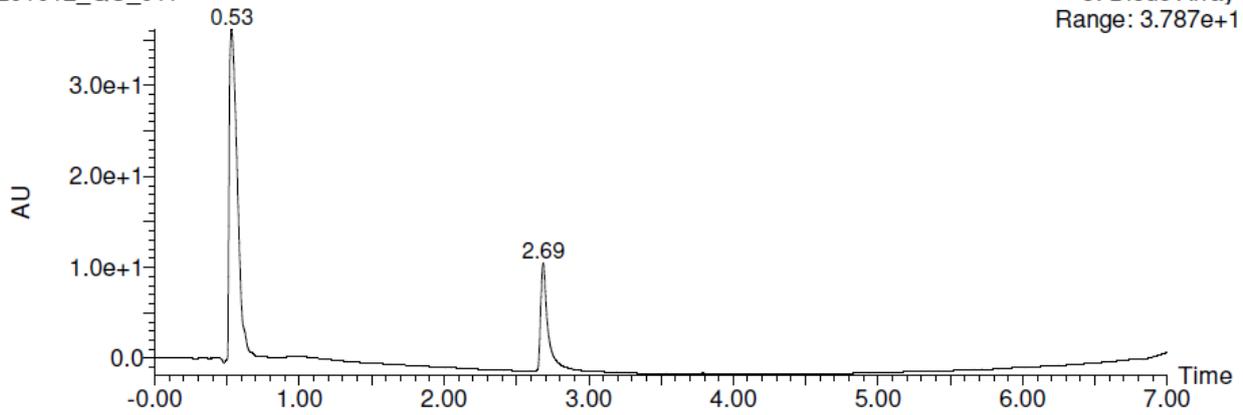
Range: 2

Time	Height	Area	Area%
2.68	97553	5549.20	100.00



201012_QC_011

3: Diode Array
Range: 3.787e+1



Computational studies – Detail

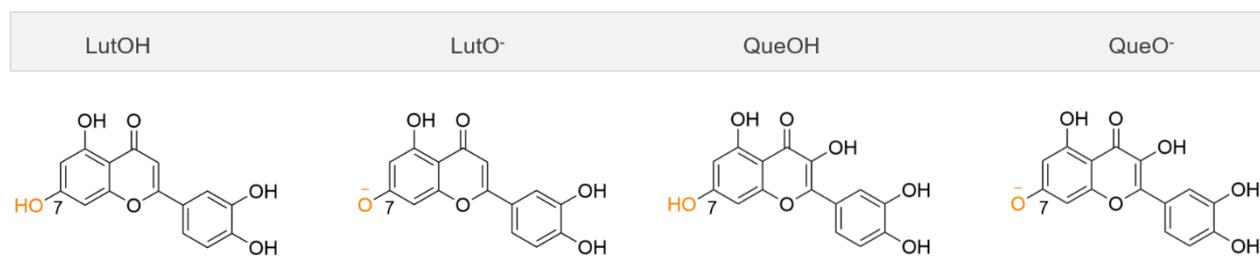


Figure C. 1 Chemical structures of luteolin and quercetin in different protonation states.

B_{RNA} binding pocket		B_{NTP} binding pocket	
Name	GScore	Name	GScore
*QueO ⁻	-7,689	*QueO ⁻	-7,62
QueO ⁻	-7,656	QueOH	-5,443
QueO ⁻	-7,541	*LutOH	-5,228
QueO ⁻	-7,25	LutOH	-5,172
QueO ⁻	-7,191	LutOH	-4,746
QueO ⁻	-7,168	QueOH	-4,534
QueO ⁻	-6,554	QueO ⁻	-4,309
*LutO ⁻	-6,175	QueO ⁻	-4,199
LutO ⁻	-6,122	QueOH	-4,178
LutO ⁻	-6,081	QueOH	-4,098
QueOH	-5,977	QueOH	-4,076
QueOH	-5,939	QueO ⁻	-4,054
QueOH	-5,935	QueO ⁻	-3,807
QueOH	-5,927	LutO ⁻	-3,464
QueOH	-5,923	QueO ⁻	-3,354
LutO ⁻	-5,869	QueOH	-3,351
QueOH	-5,801	QueO ⁻	-3,209
QueOH	-5,722	LutOH	-3,181
QueOH	-5,48	LutOH	-3,077
QueOH	-5,446	LutOH	-3,074
QueOH	-5,439	LutO ⁻	-3,039
LutO ⁻	-5,21	LutOH	-2,902
LutO ⁻	-5,187	LutOH	-2,889
LutO ⁻	-5,161	LutOH	-2,864
LutO ⁻	-5,111		
LutO ⁻	-5,028		
LutO ⁻	-4,99		
LutO ⁻	-4,899		
LutO ⁻	-4,88		
LutO ⁻	-4,603		
LutO ⁻	-4,34		
LutO ⁻	-4,148		
QueO ⁻	-3,95		
QueO ⁻	-3,828		
LutOH	-3,501		
LutOH	-3,009		
LutOH	-2,945		
LutOH	-2,89		
LutO ⁻	-2,469		

*in red selected poses for subsequent MD simulations

Table C. 1 Docking scores of both ligands in each protonation state, obtained from XP Glide for both binding pockets (i.e. B_{RNA} and B_{NTP}).

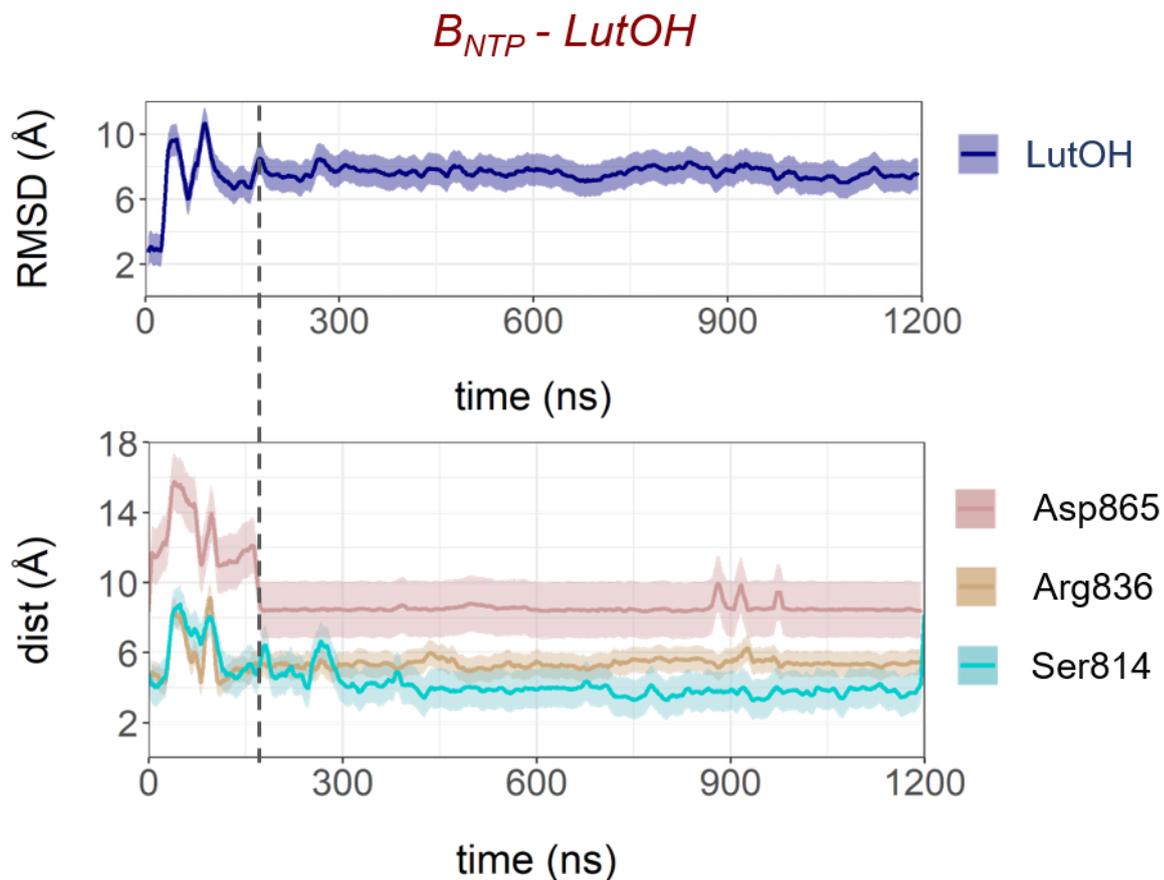


Figure C. 2 MD simulation of the B_{NTP} -LutOH system. (Top) Time evolution of the root mean square deviation (RMSD) for the heavy atoms of LutOH. (Bottom) Time evolution of the distances between the center of mass (C.O.M.) of the heavy atoms of the LutOH ligand and i) the C.O.M. of the heavy atoms of the guanidinium group of Arg836 (in light brown), ii) the C.O.M. of the heavy atoms of the carboxyl group of Asp865 (in pink), and iii) the C.O.M. of the heavy atoms of the side chain of Ser814 (in cyan). The dashed line marks the reorganization of LutOH.

$B_{NTP} - \text{QueO}^-$

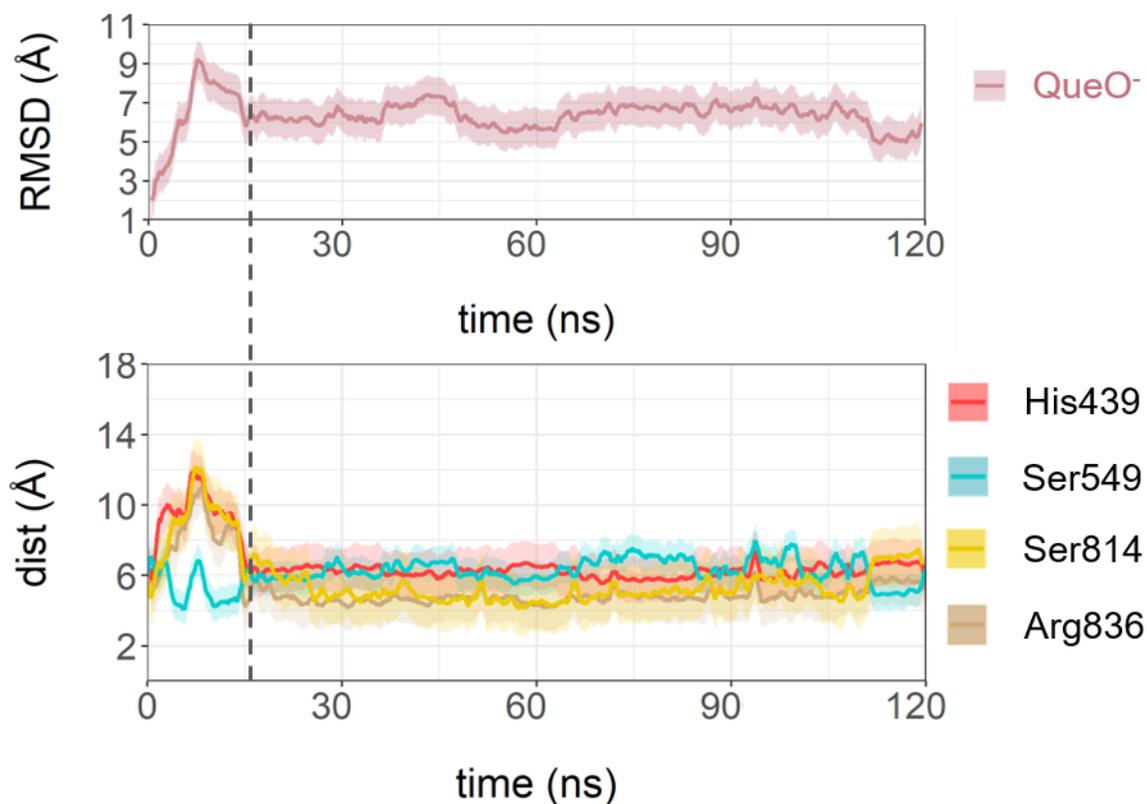


Figure C. 3 MD simulation of the B_{NTP} -QueO⁻ system. (Top) Time evolution of the root mean square deviation (RMSD) for the heavy atoms of QueO⁻. (Bottom) Time evolution of the distances between the center of mass (C.O.M.) of the heavy atoms of the QueO⁻ ligand and i) the C.O.M. of the heavy atoms of the guanidinium group of Arg836 (in light brown), ii) the C.O.M. of the heavy atoms of the side chain of Ser549 (in pink), iii) the C.O.M. of the heavy atoms of the side chain of Ser814 (in cyan). The dashed line marks the reorganization of LutOH, and iv) the C.O.M. of the heavy atoms of imidazole group of His439. The dashed line marks the reorganization of LutOH.

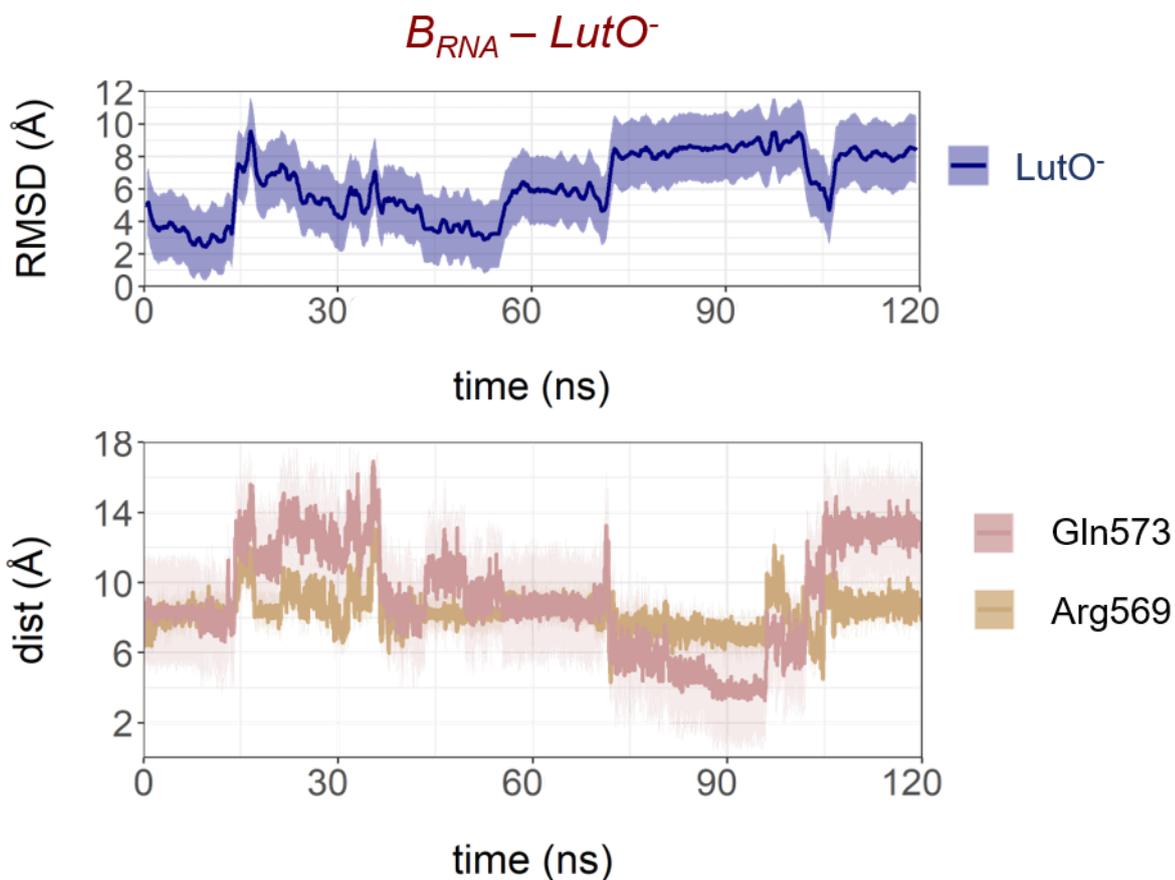


Figure C. 4 MD simulation of the B_{RNA} -LutO⁻ system. (Top) Time evolution of the root mean square deviation (RMSD) for the heavy atoms of LutO⁻. (Bottom) Time evolution of the distances between the center of mass (C.O.M.) of the heavy atoms of the LutO⁻ ligand and i) the C.O.M. of the heavy atoms of the guanidinium group of Arg569 (in light brown), ii) the C.O.M. of the heavy atoms of the side chain of Gln573 (in pink).

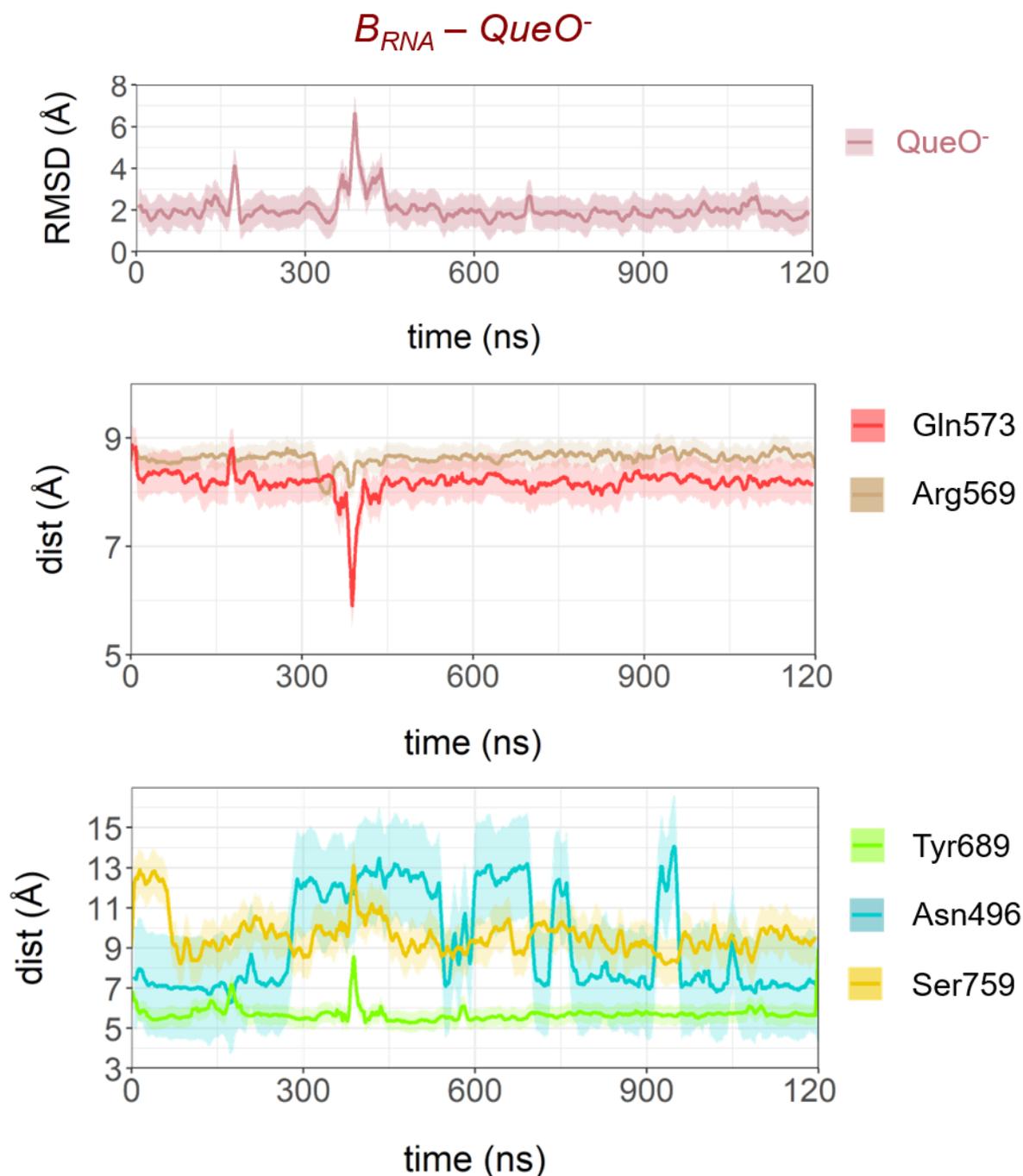


Figure C. 5 MD simulation of the B_{RNA} - $QueO^-$ system. (Top) Time evolution of the root mean square deviation (RMSD) for the heavy atoms of $QueO^-$. (Middle) Time evolution of the distances between the center of mass (C.O.M.) of the heavy atoms of the $QueO^-$ ligand and i) the C.O.M. of the heavy atoms of the guanidinium group of Arg569 (in light brown), ii) the C.O.M. of the heavy atoms of the side chain of Gln573 (in pink). (Bottom) Time evolution of the distances between the center of mass (C.O.M.) of the heavy atoms of the $QueO^-$ ligand and i) the C.O.M. of the heavy atoms of the side chain of Tyr689 (in green), ii) the C.O.M. of the heavy atoms of the side chain of Asn496 (in blue), and iii) the C.O.M. of the heavy atoms of the side chain of Ser759 (in yellow).

Bibliography

- (1) Crick, F. Central Dogma of Molecular Biology. *Nature* **1970**, 227 (2), 561–563. <https://doi.org/10.1002/9781118906545.ch4>.
- (2) Genna, V.; Donati, E.; De Vivo, M. The Catalytic Mechanism of DNA and RNA Polymerases. *ACS Catal.* **2018**, 8, 11103–11118. <https://doi.org/10.1021/acscatal.8b03363>.
- (3) Geronimo, I.; Vidossich, P.; Donati, E.; De Vivo, M. Computational Investigations of Polymerase Enzymes: Structure, Function, Inhibition, and Biotechnology. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2021**, e1534. <https://doi.org/10.1002/wcms.1534>.
- (4) Potapov, V.; Fu, X.; Dai, N.; Corrêa, I. R.; Tanner, N. A.; Ong, J. L. Base Modifications Affecting RNA Polymerase and Reverse Transcriptase Fidelity. *Nucleic Acids Res.* **2018**, 46 (11), 5753–5763. <https://doi.org/10.1093/nar/gky341>.
- (5) Loeb, L. A.; Monnat, R. J. DNA Polymerases and Human Disease. *Nat. Rev. Genet.* **2008**, 9 (8), 594–604. <https://doi.org/10.1038/nrg2345>.
- (6) Cramer, P.; Armache, K. J.; Baumli, S.; Benkert, S.; Brueckner, F.; Buchen, C.; Damsma, G. E.; Dengl, S.; Geiger, S. R.; Jasiak, A. J.; Jawhari, A.; Jennebach, S.; Kamenski, T.; Kettenberger, H.; Kuhn, C. D.; Lehmann, E.; Leike, K.; Sydow, J. F.; Vannini, A. Structure of Eukaryotic RNA Polymerases. *Annu. Rev. Biophys.* **2008**, 37, 337–352. <https://doi.org/10.1146/annurev.biophys.37.032807.130008>.
- (7) Dasari, A.; Deodhar, T.; Berdis, A. J. A Comparative Analysis of Translesion DNA Synthesis Catalyzed by a High-Fidelity DNA Polymerase. *J. Mol. Biol.* **2017**, 429 (15), 2308–2323. <https://doi.org/10.1016/j.jmb.2017.06.003>.
- (8) Freudenthal, B. D.; Beard, W. A.; Perera, L.; Shock, D. D.; Kim, T.; Schlick, T.; Wilson, S. H. Uncovering the Polymerase-Induced Cytotoxicity of an Oxidized Nucleotide. *Nature* **2015**, 517 (7536), 635–639. <https://doi.org/10.1038/nature13886>.
- (9) Lange, S. S.; Takata, K. I.; Wood, R. D. DNA Polymerases and Cancer. *Nat. Rev. Cancer* **2011**, 11 (2), 96–110. <https://doi.org/10.1038/nrc2998>.
- (10) Lang, T.; Maitra, M.; Starcevic, D.; Li, S. X.; Sweasy, J. B. A DNA Polymerase β Mutant from Colon Cancer Cells Induces Mutations. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, 101 (16), 6074–6079. <https://doi.org/10.1073/pnas.0308571101>.
- (11) Kawamura, K.; Bahar, R.; Seimiya, M.; Chiyo, M.; Wada, A.; Okada, S.; Hatano, M.; Tokuhisa, T.; Kimura, H.; Watanabe, S.; Honda, I.; Sakiyama, S.; Tagawa, M.; O-Wang, J. DNA Polymerase θ Is Preferentially Expressed in Lymphoid Tissues and Upregulated in Human Cancers. *Int. J. Cancer* **2004**, 109 (1), 9–16. <https://doi.org/10.1002/ijc.11666>.
- (12) Pillaire, M. J.; Selves, J.; Gordien, K.; Gouraud, P. A.; Gentil, C.; Danjoux, M.; Do, C.; Negre, V.; Bieth, A.; Guimbaud, R.; Trouche, D.; Pasero, P.; Méchali, M.; Hoffmann, J. S.; Cazaux, C. A DNA Replication Signature of Progression and Negative Outcome in Colorectal Cancer. *Oncogene* **2010**, 29 (6), 876–887. <https://doi.org/10.1038/onc.2009.378>.
- (13) O-Wang, J.; Kawamura, K.; Tada, Y.; Ohmori, H.; Kimura, H.; Sakiyama, S.; Tagawa, M. DNA Polymerase κ , Implicated in Spontaneous and DNA Damage-Induced Mutagenesis, Is Overexpressed in Lung Cancer. *Cancer Res.* **2001**, 61, 5366–5369.
- (14) Masutani, C.; Kusumoto Rika; YamadaAyumi; Dohmae, N.; Yokoi, M.; Yuasa, M.; Araki, M.; Iwai, S.; Takio, K.; Hanaoka, F. The XPV (Xeroderma Pigmentosum Variant) Gene

Encodes Human DNA Polymerase Eta. *Nature* **1999**, *399*, 700–704.

- (15) Rahman, S.; Copeland, W. C. POLG-Related Disorders and Their Neurological Manifestations. *Nat. Rev. Neurol.* **2019**, *15* (1), 40–52. <https://doi.org/10.1038/s41582-018-0101-0>.
- (16) Nurminen, A.; Farnum, G. A.; Kaguni, L. S. Pathogenicity in POLG Syndromes: DNA Polymerase Gamma Pathogenicity Prediction Server and Database. *BBA Clin.* **2017**, *7*, 147–156. <https://doi.org/10.1016/j.bbacli.2017.04.001>.
- (17) Zafar, M. K.; Eoff, R. L. Translesion DNA Synthesis in Cancer: Molecular Mechanisms and Therapeutic Opportunities. *Chem. Res. Toxicol.* **2017**, *30* (11), 1942–1955. <https://doi.org/10.1021/ACS.CHEMRESTOX.7B00157>.
- (18) Korzhnev, D. M.; Hadden, M. K. Targeting the Translesion Synthesis Pathway for the Development of Anti-Cancer Chemotherapeutics. *J. Med. Chem.* **2016**, *59* (20), 9321–9336. <https://doi.org/10.1021/acs.jmedchem.6b00596>.
- (19) Berdis, A. J. DNA Polymerases as Therapeutic Targets. *Biochemistry* **2008**, *47* (32), 8253–8260. <https://doi.org/10.1021/bi801179f>.
- (20) Khalil, A. S.; Collins, J. J. Synthetic Biology: Applications Come of Age. *Nat. Rev. Genet.* **2010**, *11* (5), 367–379. <https://doi.org/10.1038/nrg2775>.
- (21) Houlihan, G.; Arangundy-Franklin, S.; Holliger, P. Engineering and Application of Polymerases for Synthetic Genetics. *Curr. Opin. Biotechnol.* **2017**, *48*, 168–179. <https://doi.org/10.1016/J.COPBIO.2017.04.004>.
- (22) Coulther, T. A.; Stern, H. R.; Beuning, P. J. Engineering Polymerases for New Functions. *Trends Biotechnol.* **2019**, *37* (10), 1091–1103. <https://doi.org/10.1016/j.tibtech.2019.03.011>.
- (23) Nikoomanzar, A.; Chim, N.; Yik, E. J.; Chaput, J. C. Engineering Polymerases for Applications in Synthetic Biology. *Q. Rev. Biophys.* **2020**, *53* (e8). <https://doi.org/10.1017/S0033583520000050>.
- (24) Saiki, R. K.; Gelfand, D. H.; Stoffel, S.; Scharf, S. J.; Higuchi, R.; Horn, G. T.; Mullis, K. B.; Erlich, H. A. Primer-Directed Enzymatic Amplification of DNA with a Thermostable DNA Polymerase. *Science* (80-.). **1988**, *239* (4839), 487–491. <https://doi.org/10.1126/science.2448875>.
- (25) Kutzner, C.; Páll, S.; Fechner, M.; Esztermann, A.; de Groot, B. L.; Grubmüller, H. More Bang for Your Buck: Improved Use of GPU Nodes for GROMACS 2018. *J. Comput. Chem.* **2019**, *40* (27), 2418–2431. <https://doi.org/10.1002/jcc.26011>.
- (26) Phillips, J. C.; Hardy, D. J.; Maia, J. D. C.; Stone, J. E.; Ribeiro, J. V.; Bernardi, R. C.; Buch, R.; Fiorin, G.; Hénin, J.; Jiang, W.; McGreevy, R.; Melo, M. C. R.; Radak, B. K.; Skeel, R. D.; Singharoy, A.; Wang, Y.; Roux, B.; Aksimentiev, A.; Luthey-Schulten, Z.; Kalé, L. V.; Schulten, K.; Chipot, C.; Tajkhorshid, E. Scalable Molecular Dynamics on CPU and GPU Architectures with NAMD. *J. Chem. Phys.* **2020**, *153* (4), 44130. <https://doi.org/10.1063/5.0014475>.
- (27) Lee, T.-S.; Cerutti, D. S.; Mermelstein, D.; Lin, C.; LeGrand, S.; Giese, T. J.; Roitberg, A.; Case, D. A.; Walker, R. C.; York, D. M. GPU-Accelerated Molecular Dynamics and Free Energy Methods in Amber18: Performance Enhancements and New Features. *J. Chem. Inf. Model.* **2018**, *58* (10), 2043–2050. <https://doi.org/10.1021/ACS.JCIM.8B00462>.
- (28) Yang, W.; Gao, Y. Translesion and Repair DNA Polymerases: Diverse Structure and

- Mechanism. *Annu. Rev. Biochem.* **2018**, *87*, 239–261. <https://doi.org/10.1146/annurev-biochem-062917-012405>.
- (29) Wu, W. J.; Yang, W.; Tsai, M. D. How DNA Polymerases Catalyse Replication and Repair with Contrasting Fidelity. *Nature Reviews Chemistry*. 2017, pp 1–16. <https://doi.org/10.1038/s41570-017-0068>.
- (30) Kunkel, T. A. DNA Replication Fidelity. *Journal of Biological Chemistry*. Elsevier April 23, 2004, pp 16895–16898. <https://doi.org/10.1074/jbc.R400006200>.
- (31) Lehmann, A. R. New Functions for Y Family Polymerases. *Molecular Cell*. Elsevier November 17, 2006, pp 493–495. <https://doi.org/10.1016/j.molcel.2006.10.021>.
- (32) Yamtich, J.; Sweasy, J. B. DNA Polymerase Family X: Function, Structure, and Cellular Roles. *Biochim. Biophys. Acta - Proteins Proteomics* **2010**, *1804* (5), 1136–1150. <https://doi.org/10.1016/J.BBAPAP.2009.07.008>.
- (33) Raper, A. T.; Reed, A. J.; Suo, Z. Kinetic Mechanism of DNA Polymerases: Contributions of Conformational Dynamics and a Third Divalent Metal Ion. *Chem. Rev.* **2018**, *118* (12), 6000–6025. <https://doi.org/10.1021/ACS.CHEMREV.7B00685>.
- (34) Svetlov, V.; Nudler, E. Basic Mechanism of Transcription by RNA Polymerase II. *Biochim. Biophys. Acta - Gene Regul. Mech.* **2013**, *1829* (1), 20–28. <https://doi.org/10.1016/J.BBAGRM.2012.08.009>.
- (35) 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* (4), 1699–1705. <https://doi.org/10.1021/acs.jctc.7b00157>.
- (36) Smith, E. C.; Sexton, N. R.; Denison, M. R. Thinking Outside the Triangle: Replication Fidelity of the Largest RNA Viruses. *Annu. Rev. Virol.* **2014**, *1* (1), 111–132. <https://doi.org/10.1146/annurev-virology-031413-085507>.
- (37) Warren, J. J.; Forsberg, L. J.; Beese, L. S. The Structural Basis for the Mutagenicity of O6-Methyl-Guanine Lesions. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (52), 19701–19706. <https://doi.org/10.1073/pnas.0609580103>.
- (38) Wang, D.; Bushnell, D. A.; Westover, K. D.; Kaplan, C. D.; Kornberg, R. D. Structural Basis of Transcription: Role of the Trigger Loop in Substrate Specificity and Catalysis. *Cell* **2006**, *127* (5), 941–954. <https://doi.org/10.1016/j.cell.2006.11.023>.
- (39) Basu, R. S.; Warner, B. A.; Molodtsov, V.; Pupov, D.; Esyunina, D.; Fernández-Tornero, C.; Kulbachinskiy, A.; Murakami, K. S. Structural Basis of Transcription Initiation by Bacterial RNA Polymerase Holoenzyme. *J. Biol. Chem.* **2014**, *289* (35), 24549–24559. <https://doi.org/10.1074/jbc.M114.584037>.
- (40) Murakami, K. S. Structural Biology of Bacterial RNA Polymerase. *Biomolecules* **2015**, *5* (2), 848–864. <https://doi.org/10.3390/biom5020848>.
- (41) Yang, W. An Overview of Y-Family DNA Polymerases and a Case Study of Human DNA Polymerase π . *Biochemistry* **2014**, *53* (17), 2793–2803. <https://doi.org/10.1021/bi500019s>.
- (42) Reed, A. J.; Vyas, R.; Raper, A. T.; Suo, Z. Structural Insights into the Post-Chemistry Steps of Nucleotide Incorporation Catalyzed by a DNA Polymerase. *J. Am. Chem. Soc.* **2017**, *139* (1), 465–471. <https://doi.org/10.1021/jacs.6b11258>.
- (43) Patel, S. S.; Wong, I.; Johnson, K. A. Pre-Steady-State Kinetic Analysis of Processive DNA

Replication Including Complete Characterization of an Exonuclease-Deficient Mutant. *Biochemistry* **1991**, *30* (2), 511–525. <https://doi.org/10.1021/bi00216a029>.

- (44) Steitz, T. A. A Mechanism for All Polymerases. *Nature* **1998**, *391*, 231–232.
- (45) Palermo, G.; Cavalli, A.; Klein, M. L.; Alfonso-Prieto, M.; Dal Peraro, M.; De Vivo, M. Catalytic Metal Ions and Enzymatic Processing of DNA and RNA. *Acc. Chem. Res.* **2015**, *48* (2), 220–228. <https://doi.org/10.1021/ar500314j>.
- (46) Yang, W.; Lee, J. Y.; Nowotny, M. Making and Breaking Nucleic Acids: Two-Mg²⁺-Ion Catalysis and Substrate Specificity. *Mol. Cell* **2006**, *22* (1), 5–13. <https://doi.org/10.1016/j.molcel.2006.03.013>.
- (47) 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* (22), 13596–13605. <https://doi.org/10.1021/ACSCATAL.0C03566>.
- (48) Steitz, T. A.; Steitz, J. A. A General Two-Metal-Ion Mechanism for Catalytic RNA. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90* (14), 6498–6502. <https://doi.org/10.1073/pnas.90.14.6498>.
- (49) De Vivo, M.; Dal Peraro, M.; Klein, M. L. Phosphodiester Cleavage in Ribonuclease H Occurs via an Associative Two-Metal-Aided Catalytic Mechanism. *J. Am. Chem. Soc.* **2008**, *130* (33), 10955–10962. <https://doi.org/10.1021/ja8005786>.
- (50) Boehr, D. D.; Arnold, J. J.; Moustafa, I. M.; Cameron, C. E. Structure, Dynamics, and Fidelity of RNA-Dependent RNA Polymerases. In *Nucleic Acids and Molecular Biology*; Springer, Berlin, Heidelberg, 2014; Vol. 30, pp 309–333. https://doi.org/10.1007/978-3-642-39796-7_14.
- (51) Nakamura, T.; Zhao, Y.; Yamagata, Y.; Hua, Y. J.; Yang, W. Watching DNA Polymerase η Make a Phosphodiester Bond. *Nature* **2012**, *487* (7406), 196–201. <https://doi.org/10.1038/nature11181>.
- (52) Wu, S.; Beard, W. A.; Pedersen, L. G.; Wilson, S. H. Structural Comparison of DNA Polymerase Architecture Suggests a Nucleotide Gateway to the Polymerase Active Site. *Chemical Reviews*. American Chemical Society March 12, 2014, pp 2759–2774. <https://doi.org/10.1021/cr3005179>.
- (53) Gao, Y.; Yang, W. Capture of a Third Mg²⁺ Is Essential for Catalyzing DNA Synthesis. *Science* (80-.). **2016**, *352* (6291), 1334–1337. <https://doi.org/10.1126/science.aad9633>.
- (54) Vyas, R.; Reed, A. J.; Tokarsky, E. J.; Suo, Z. Viewing Human DNA Polymerase β Faithfully and Unfaithfully Bypass an Oxidative Lesion by Time-Dependent Crystallography. *J. Am. Chem. Soc.* **2015**, *137* (15), 5225–5230. <https://doi.org/10.1021/jacs.5b02109>.
- (55) Jamsen, J. A.; Beard, W. A.; Pedersen, L. C.; Shock, D. D.; Moon, A. F.; Krahn, J. M.; Bebenek, K.; Kunkel, T. A.; Wilson, S. H. Time-Lapse Crystallography Snapshots of a Double-Strand Break Repair Polymerase in Action. *Nat. Commun.* **2017**, *8* (1). <https://doi.org/10.1038/s41467-017-00271-7>.
- (56) Moon, A. F.; Pryor, J. M.; Ramsden, D. A.; Kunkel, T. A.; Bebenek, K.; Pedersen, L. C. Sustained Active Site Rigidity during Synthesis by Human DNA Polymerase μ . *Nat. Struct. Mol. Biol.* **2014**, *21* (3), 253–260. <https://doi.org/10.1038/nsmb.2766>.
- (57) Chim, N.; Shi, C.; Sau, S. P.; Nikoomanzar, A.; Chaput, J. C. Structural Basis for TNA Synthesis by an Engineered TNA Polymerase. *Nat. Commun.* **2017**, *8* (1).

<https://doi.org/10.1038/s41467-017-02014-0>.

- (58) Molina, R.; Stella, S.; Redondo, P.; Gomez, H.; Marcaida, M. J.; Orozco, M.; Prieto, J.; Montoya, G. Visualizing Phosphodiester-Bond Hydrolysis by an Endonuclease. *Nat. Struct. Mol. Biol.* **2015**, *22* (1), 65–72. <https://doi.org/10.1038/nsmb.2932>.
- (59) Shi, Y.; Hellinga, H. W.; Beese, L. S. Interplay of Catalysis, Fidelity, Threading, and Processivity in the Exo-and Endonucleolytic Reactions of Human Exonuclease I. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (23), 6010–6015. <https://doi.org/10.1073/pnas.1704845114>.
- (60) Manigrasso, J.; De Vivo, M.; Palermo, G. Controlled Trafficking of Multiple and Diverse Cations Prompts Nucleic Acid Hydrolysis. *ACS Catal.* **2021**, *11* (14), 8786–8797. <https://doi.org/10.1021/acscatal.1c01825>.
- (61) Yang, W.; Weng, P. J.; Gao, Y. A New Paradigm of DNA Synthesis: Three-Metal-Ion Catalysis. *Cell Biosci.* **2016**, *6* (1), 51. <https://doi.org/10.1186/s13578-016-0118-2>.
- (62) 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* (6), 2827–2836. <https://doi.org/10.1093/nar/gkw128>.
- (63) Genna, V.; Vidossich, P.; Ippoliti, E.; Carloni, P.; Vivo, M. De. A Self-Activated Mechanism for Nucleic Acid Polymerization Catalyzed by DNA/RNA Polymerases. *J. Am. Chem. Soc.* **2016**, *138* (44), 14592–14598. <https://doi.org/10.1021/jacs.6b05475>.
- (64) 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 Struct. Funct. Bioinforma.* **2017**, *85* (8), 1446–1453. <https://doi.org/10.1002/prot.25305>.
- (65) Stevens, D. R.; 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* (28), 8965–8969. <https://doi.org/10.1021/jacs.8b05177>.
- (66) 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* (9), 3312–3321. <https://doi.org/10.1021/jacs.7b12446>.
- (67) Jorgensen, W. L. Computer-Aided Discovery of Anti-HIV Agents. *Bioorganic Med. Chem.* **2016**, *24* (20), 4768–4778. <https://doi.org/10.1016/j.bmc.2016.07.039>.
- (68) Wang, M.; Cao, R.; Zhang, L.; Yang, X.; Liu, J.; Xu, M.; Shi, Z.; Hu, Z.; Zhong, W.; Xiao, G. Remdesivir and Chloroquine Effectively Inhibit the Recently Emerged Novel Coronavirus (2019-NCoV) in Vitro. *Cell Res.* **2020**, *30* (3), 269–271. <https://doi.org/10.1038/s41422-020-0282-0>.
- (69) Wang, Y.; Zhang, D.; Du, G.; Du, R.; Zhao, J.; Jin, Y.; Fu, S.; Gao, L.; Cheng, Z.; Lu, Q.; Hu, Y.; Luo, G.; Wang, K.; Lu, Y.; Li, H.; Wang, S.; Ruan, S.; Yang, C.; Mei, C.; Wang, Y.; Ding, D.; Wu, F.; Tang, X.; Ye, X.; Ye, Y.; Liu, B.; Yang, J.; Yin, W.; Wang, A.; Fan, G.; Zhou, F.; Liu, Z.; Gu, X.; Xu, J.; Shang, L.; Zhang, Y.; Cao, L.; Guo, T.; Wan, Y.; Qin, H.; Jiang, Y.; Jaki, T.; Hayden, F. G.; Horby, P. W.; Cao, B.; Wang, C. Remdesivir in Adults with Severe COVID-19: A Randomised, Double-Blind, Placebo-Controlled, Multicentre Trial. *Lancet* **2020**, *395* (10236), 1569–1578. [https://doi.org/10.1016/S0140-6736\(20\)31022-9](https://doi.org/10.1016/S0140-6736(20)31022-9).
- (70) Wakchaure, P. D.; 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* (47), 10641–10652. <https://doi.org/10.1021/acs.jpcc.0c06747>.

- (71) Koulgi, S.; Jani, V.; Uppuladinne, M. V. N.; 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* (45), 26792–26803. <https://doi.org/10.1039/d0ra04743k>.
- (72) 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* (32), 6955–6962. <https://doi.org/10.1021/acs.jpcc.0c04198>.
- (73) Gardner, A. F.; Kelman, Z. DNA Polymerases in Biotechnology. *Front. Microbiol.* **2014**, *5*, 659. <https://doi.org/10.3389/fmicb.2014.00659>.
- (74) McDonald, J. P.; 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* (4), 1102–1111. <https://doi.org/10.1093/nar/gkj512>.
- (75) Wu, J.; De Paz, A.; Zamft, B. M.; Marblestone, A. H.; Boyden, E. S.; Kording, K. P.; Tyo, K. E. J. DNA Binding Strength Increases the Processivity and Activity of a Y-Family DNA Polymerase. *Sci. Rep.* **2017**, *7* (1). <https://doi.org/10.1038/s41598-017-02578-3>.
- (76) Grúz, P.; Pisani, F. M.; Shimizu, M.; Yamada, M.; Hayashi, I.; Morikawa, K.; Nohmi, T. 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* (50), 47394–47401. <https://doi.org/10.1074/jbc.M107213200>.
- (77) Wang, L.; Liang, C.; Wu, J.; Liu, L.; Tyo, K. E. J. Increased Processivity, Misincorporation, and Nucleotide Incorporation Efficiency in *Sulfolobus solfataricus* Dpo4 Thumb Domain Mutants. *Appl. Environ. Microbiol.* **2017**, *83* (18). <https://doi.org/10.1128/AEM.01013-17>.
- (78) Marx, A.; Betz, K. The Structural Basis for Processing of Unnatural Base Pairs by DNA Polymerases. *Chemistry - A European Journal*. Wiley-VCH Verlag March 18, 2020, pp 3446–3463. <https://doi.org/10.1002/chem.201903525>.
- (79) Yang, W. Nucleases: Diversity of Structure, Function and Mechanism. *Q. Rev. Biophys.* **2011**, *44* (1), 1–93. <https://doi.org/10.1017/s0033583510000181>.
- (80) Shen, B.; Singh, P.; Liu, R.; Qiu, J.; Zheng, L.; Finger, L. D.; Alas, S. Multiple but Dissectible Functions of FEN-1 Nucleases in Nucleic Acid Processing, Genome Stability and Diseases. *BioEssays*. John Wiley & Sons, Ltd July 1, 2005, pp 717–729. <https://doi.org/10.1002/bies.20255>.
- (81) Kao, H. I.; Bambara, R. A. The Protein Components and Mechanism of Eukaryotic Okazaki Fragment Maturation. *Critical Reviews in Biochemistry and Molecular Biology*. Taylor & Francis 2003, pp 433–452. <https://doi.org/10.1080/10409230390259382>.
- (82) Marti, T. M.; Fleck, O. DNA Repair Nucleases. *Cell. Mol. Life Sci.* **2004**, *61* (3), 336–354. <https://doi.org/10.1007/s00018-003-3223-4>.
- (83) Mimitou, E. P.; Symington, L. S. DNA End Resection: Many Nucleases Make Light Work. *DNA Repair (Amst)*. **2009**, *8* (9), 983–995. <https://doi.org/10.1016/j.dnarep.2009.04.017>.
- (84) Doherty, R.; Madhusudan, S. DNA Repair Endonucleases: Physiological Roles and Potential

as Drug Targets. *J. Biomol. Screen.* **2015**, *20* (7), 829–841.
<https://doi.org/10.1177/1087057115581581>.

- (85) Ameisen, J. C. On the Origin, Evolution, and Nature of Programmed Cell Death: A Timeline of Four Billion Years. *Cell Death Differ.* **2002**, *9* (4), 367–393.
<https://doi.org/10.1038/sj/cdd/4400950>.
- (86) Schoeffler, A. J.; Berger, J. M. DNA Topoisomerases: Harnessing and Constraining Energy to Govern Chromosome Topology. *Q. Rev. Biophys.* **2008**, *41* (1), 41–101.
<https://doi.org/10.1017/S003358350800468X>.
- (87) Grindley, N. D. F.; Whiteson, K. L.; Rice, P. A. Mechanisms of Site-Specific Recombination. *Annu. Rev. Biochem.* **2006**, *75*, 567–607.
<https://doi.org/10.1146/annurev.biochem.73.011303.073908>.
- (88) Nowotny, M.; Yang, W. Structural and Functional Modules in RNA Interference. *Curr. Opin. Struct. Biol.* **2009**, *19* (3), 286–293. <https://doi.org/10.1016/J.SBI.2009.04.006>.
- (89) Chu, C.-Y.; Rana, T. M. Small RNAs: Regulators and Guardians of the Genome. *J. Cell. Physiol.* **2007**, *213* (2), 412–419. <https://doi.org/10.1002/JCP.21230>.
- (90) Abelson, J.; Trotta, C. R.; Li, H. tRNA Splicing. *J. Biol. Chem.* **1998**, *273* (21), 12685–12688. <https://doi.org/10.1074/JBC.273.21.12685>.
- (91) Patel, A. A.; Steitz, J. A. Splicing Double: Insights from the Second Spliceosome. *Nat. Rev. Mol. Cell Biol.* **2003**, *4* (12), 960–970. <https://doi.org/10.1038/nrm1259>.
- (92) Tarnauskaitė, Ž.; Bicknell, L. S.; Marsh, J. A.; Murray, J. E.; Parry, D. A.; Logan, C. V.; Bober, M. B.; de Silva, D. C.; Duker, A. L.; Sillence, D.; Wise, C.; Jackson, A. P.; Murina, O.; Reijns, M. A. M. Biallelic Variants in DNA2 Cause Microcephalic Primordial Dwarfism. *Hum. Mutat.* **2019**, *40* (8), 1063–1070. <https://doi.org/10.1002/humu.23776>.
- (93) Falquet, B.; Ölmezer, G.; Enkner, F.; Klein, D.; Challa, K.; Appanah, R.; Gasser, S. M.; Rass, U. Disease-Associated DNA2 Nuclease–Helicase Protects Cells from Lethal Chromosome under-Replication. *Nucleic Acids Res.* **2020**, *48* (13), 7265–7278.
<https://doi.org/10.1093/nar/gkaa524>.
- (94) Strauss, C.; Kornowski, M.; Benvenisty, A.; Shahar, A.; Masury, H.; Ben-Porath, I.; Ravid, T.; Arbel-Eden, A.; Goldberg, M. The DNA2 Nuclease/Helicase Is an Estrogen-Dependent Gene Mutated in Breast and Ovarian Cancers. *Oncotarget* **2014**, *5* (19), 9396–9409.
<https://doi.org/10.18632/oncotarget.2414>.
- (95) Deshmukh, A. L.; Porro, A.; Mohiuddin, M.; Lanni, S.; Panigrahi, G. B.; Caron, M. C.; Masson, J. Y.; Sartori, A. A.; Pearson, C. E. FAN1, a DNA Repair Nuclease, as a Modifier of Repeat Expansion Disorders. *J. Huntingtons. Dis.* **2021**, *10* (1), 95–122.
<https://doi.org/10.3233/JHD-200448>.
- (96) Sun, H.; He, L.; Wu, H.; Pan, F.; Wu, X.; Zhao, J.; Hu, Z.; Sekhar, C.; Li, H.; Zheng, L.; Chen, H.; Shen, B. H.; Guo, Z. The FEN1 L209P Mutation Interferes with Long-Patch Base Excision Repair and Induces Cellular Transformation. *Oncogene* **2017**, *36* (2), 194–207.
<https://doi.org/10.1038/onc.2016.188>.
- (97) Liu, S.; Lu, G.; Ali, S.; Liu, W.; Zheng, L.; Dai, H.; Li, H.; Xu, H.; Hua, Y.; Zhou, Y.; Ortega, J.; Li, G.-M.; Kunkel, T. A.; Shen, B. Okazaki Fragment Maturation Involves A-segment Error Editing by the Mammalian FEN1/MutSα Functional Complex. *EMBO J.* **2015**, *34* (13), 1829–1843. <https://doi.org/10.15252/embj.201489865>.

- (98) Pizzolato, J.; Mukherjee, S.; Schärer, O. D.; Jiricny, J. FANCD2-Associated Nuclease 1, but Not Exonuclease 1 or Flap Endonuclease 1, Is Able to Unhook DNA Interstrand Cross-Links in Vitro. *J. Biol. Chem.* **2015**, *290* (37), 22602–22611. <https://doi.org/10.1074/jbc.M115.663666>.
- (99) Zheng, L.; Dai, H.; Zhou, M.; Li, M.; Singh, P.; Qiu, J.; Tsark, W.; Huang, Q.; Kernstine, K.; Zhang, X.; Lin, D.; Shen, B. Fen1 Mutations Result in Autoimmunity, Chronic Inflammation and Cancers. *Nat. Med.* **2007**, *13* (7), 812–819. <https://doi.org/10.1038/nm1599>.
- (100) Brosh, R. M.; Von Kobbe, C.; Sommers, J. A.; Karmakar, P.; Opresko, P. L.; Piotrowski, J.; Dianova, I.; Dianov, G. L.; Bohr, V. A. Werner Syndrome Protein Interacts with Human Flap Endonuclease 1 and Stimulates Its Cleavage Activity. *EMBO J.* **2001**, *20* (20), 5791–5801. <https://doi.org/10.1093/emboj/20.20.5791>.
- (101) Stöhr, H.; Marquardt, A.; Rivera, A.; Cooper, P. R.; Nowak, N. J.; Shows, T. B.; Gerhard, D. S.; Weber, B. H. F. A Gene Map of the Best's Vitelliform Macular Dystrophy Region in Chromosome 11q12-Q13.1. *Genome Res.* **1998**, *8* (1), 48–56. <https://doi.org/10.1101/gr.8.1.48>.
- (102) Hengel, S. R.; Spies, M. A.; Spies, M. Small-Molecule Inhibitors Targeting DNA Repair and DNA Repair Deficiency in Research and Cancer Therapy. *Cell Chem. Biol.* **2017**, *24* (9), 1101–1119. <https://doi.org/10.1016/j.chembiol.2017.08.027>.
- (103) Lodovichi, S.; Cervelli, T.; Pelliccioli, A.; Galli, A. Inhibition of Dna Repair in Cancer Therapy: Toward a Multi-Target Approach. *Int. J. Mol. Sci.* **2020**, *21* (18), 1–29. <https://doi.org/10.3390/ijms21186684>.
- (104) Canestrari, E.; Paroo, Z. Ribonucleases as Drug Targets. *Trends Pharmacol. Sci.* **2018**, *39* (10), 855–866. <https://doi.org/10.1016/J.TIPS.2018.07.005>.
- (105) Guha, T. K.; Edgell, D. R. Applications of Alternative Nucleases in the Age of CRISPR/Cas9. *Int. J. Mol. Sci.* **2017**, *18* (12), 2565. <https://doi.org/10.3390/IJMS18122565>.
- (106) Normile, D. China Sprints Ahead in CRISPR Therapy Race. *Science* (80-.). **2017**, *358* (6359), 20–21. <https://doi.org/10.1126/science.358.6359.20>.
- (107) Zhou, C.; Pourmal, S.; Pavletich, N. P. Dna2 Nuclease-Helicase Structure, Mechanism and Regulation by Rpa. *Elife* **2015**, *4* (e09832). <https://doi.org/10.7554/eLife.09832>.
- (108) Hyjek, M.; Figiel, M.; Nowotny, M. RNases H: Structure and Mechanism. *DNA Repair (Amst)*. **2019**, *84*, 102672. <https://doi.org/10.1016/j.dnarep.2019.102672>.
- (109) Rangarajan, E. S.; Shankar, V. Sugar Non-Specific Endonucleases. *FEMS Microbiol. Rev.* **2001**, *25* (5), 583–613. <https://doi.org/10.1111/j.1574-6976.2001.tb00593.x>.
- (110) Hsia, K. C.; Li, C. L.; Yuan, H. S. Structural and Functional Insight into Sugar-Nonspecific Nucleases in Host Defense. *Curr. Opin. Struct. Biol.* **2005**, *15* (1), 126–134. <https://doi.org/10.1016/J.SBI.2005.01.015>.
- (111) Deltcheva, E.; Chylinski, K.; Sharma, C. M.; Gonzales, K.; Chao, Y.; Pirzada, Z. A.; Eckert, M. R.; Vogel, J.; Charpentier, E. CRISPR RNA Maturation by Trans-Encoded Small RNA and Host Factor RNase III. *Nature* **2011**, *471* (7340), 602–607. <https://doi.org/10.1038/nature09886>.
- (112) Urnov, F. D.; Rebar, E. J.; Holmes, M. C.; Zhang, H. S.; Gregory, P. D. Genome Editing with Engineered Zinc Finger Nucleases. *Nat. Rev. Genet.* **2010**, *11* (9), 636–646. <https://doi.org/10.1038/nrg2842>.

- (113) Nemudryi, A. A.; Valetdinova, K. R.; Medvedev, S. P.; Zakian, S. M. TALEN and CRISPR/Cas Genome Editing Systems: Tools of Discovery. *Acta Naturae* **2014**, *6* (22), 19–40. <https://doi.org/10.32607/20758251-2014-6-3-19-40>.
- (114) Nishino, T.; Ishino, Y.; Morikawa, K. Structure-Specific DNA Nucleases: Structural Basis for 3D-Scissors. *Curr. Opin. Struct. Biol.* **2006**, *16* (1), 60–67. <https://doi.org/10.1016/j.sbi.2006.01.009>.
- (115) Yang, W. An Equivalent Metal Ion in One- and Two-Metal-Ion Catalysis. *Nat. Struct. Mol. Biol.* **2008**, *15* (11), 1228–1231. <https://doi.org/10.1038/nsmb.1502>.
- (116) Cooper, D. A.; Jha, B. K.; Silverman, R. H.; Hesselberth, J. R.; Barton, D. J. Ribonuclease L and Metal-Ion-Independent Endoribonuclease Cleavage Sites in Host and Viral RNAs. *Nucleic Acids Res.* **2014**, *42* (8), 5202–5216. <https://doi.org/10.1093/nar/gku118>.
- (117) Grazulis, S.; Manakova, E.; Roessle, M.; Bochtler, M.; Tamulaitiene, G.; Huber, R.; Siksnys, V. Structure of the Metal-Independent Restriction Enzyme BfiI Reveals Fusion of a Specific DNA-Binding Domain with a Nonspecific Nuclease. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (44), 15797–15802. <https://doi.org/10.1073/pnas.0507949102>.
- (118) Van Gent, D. C.; Mizuuchi, K.; Gellert, M. Similarities between Initiation of V(D)J Recombination and Retroviral Integration. *Science* (80-.). **1996**, *271* (5255), 1592–1594. <https://doi.org/10.1126/science.271.5255.1592>.
- (119) Doudna, J. A.; Cech, T. R. The Chemical Repertoire of Natural Ribozymes. *Nature* **2002**, *418* (6894), 222–228. <https://doi.org/10.1038/418222a>.
- (120) Stuckey, J. A.; Dixon, J. E. Crystal Structure of a Phospholipase D Family Member. *Nat. Struct. Biol.* **1999**, *6* (3), 278–284. <https://doi.org/10.1038/6716>.
- (121) Raines, R. T. Ribonuclease A. *Chem. Rev.* **1998**, *98* (3), 1045–1065. <https://doi.org/10.1021/cr960427h>.
- (122) Deutscher, M. P.; Marshall, G. T.; Cudny, H. RNase PH: An Escherichia Coli Phosphate-Dependent Nuclease Distinct from Polynucleotide Phosphorylase. *Proc. Natl. Acad. Sci. U. S. A.* **1988**, *85* (13), 4710–4714. <https://doi.org/10.1073/pnas.85.13.4710>.
- (123) Marcia, M.; Pyle, A. M. Visualizing Group II Intron Catalysis through the Stages of Splicing. *Cell* **2012**, *151* (3), 497–507. <https://doi.org/10.1016/j.cell.2012.09.033>.
- (124) Marcia, M.; Humphris-Narayanan, E.; Keating, K. S.; Somarowthu, S.; Rajashankar, K.; Pyle, A. M. Solving Nucleic Acid Structures by Molecular Replacement: Examples from Group II Intron Studies. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2013**, *69* (11), 2174–2185. <https://doi.org/10.1107/S0907444913013218>.
- (125) 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* (1), 40–50. <https://doi.org/10.1016/j.str.2017.11.008>.
- (126) Dupureur, C. M. An Integrated Look at Metallonuclease Mechanism. *Curr. Chem. Biol.* **2008**, *2* (2), 159–173. <https://doi.org/10.2174/187231308784220491>.
- (127) Dupureur, C. M. Roles of Metal Ions in Nucleases. *Curr. Opin. Chem. Biol.* **2008**, *12* (2), 250–255. <https://doi.org/10.1016/J.CBPA.2008.01.012>.
- (128) Cowan, J. A. Structural and Catalytic Chemistry of Magnesium-Dependent Enzymes. *BioMetals* **2002**, *15* (3), 225–235. <https://doi.org/10.1023/A:1016022730880>.

- (129) Harding, M. M. The Geometry of Metal-Ligand Interactions Relevant to Proteins. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **1999**, *55* (8), 1432–1443. <https://doi.org/10.1107/S0907444499007374>.
- (130) Pidcock, E.; Moore, G. R. Structural Characteristics of Protein Binding Sites for Calcium and Lanthanide Ions. *J. Biol. Inorg. Chem.* **2001**, *6* (5), 479–489. <https://doi.org/10.1007/S007750100214>.
- (131) Christianson, D. W. Structural Biology of Zinc. *Adv. Protein Chem.* **1991**, *42*, 281–355. [https://doi.org/10.1016/S0065-3233\(08\)60538-0](https://doi.org/10.1016/S0065-3233(08)60538-0).
- (132) Ivanov, I.; Tainer, J. A.; McCammon, J. A. Unraveling the Three-Metal-Ion Catalytic Mechanism of the DNA Repair Endonuclease IV. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (5), 1465–1470. <https://doi.org/10.1073/pnas.0603468104>.
- (133) Syson, K.; Tomlinson, C.; Chapados, B. R.; Sayers, J. R.; Tainer, J. A.; Williams, N. H.; Grasby, J. A. Three Metal Ions Participate in the Reaction Catalyzed by T5 Flap Endonuclease. *J. Biol. Chem.* **2008**, *283* (42), 28741–28746. <https://doi.org/10.1074/jbc.M801264200>.
- (134) Palermo, G.; Cavalli, A.; Klein, M. L.; Alfonso-Prieto, M.; Dal Peraro, M.; De Vivo, M. Catalytic Metal Ions and Enzymatic Processing of DNA and RNA. *Acc. Chem. Res.* **2015**, *48* (2), 220–228. <https://doi.org/10.1021/ar500314j>.
- (135) Beese, L. S.; Steitz, T. A. Structural Principles for the Inhibition of the 3'-5' Exonuclease Activity of Escherichia Coli DNA Polymerase I by Phosphorothioates. *J. Mol. Biol.* **1998**, *277* (2), 363–377. <https://doi.org/10.1006/jmbi.1997.1586>.
- (136) Nowotny, M.; Yang, W. Stepwise Analyses of Metal Ions in RNase H Catalysis from Substrate Destabilization to Product Release. *EMBO J.* **2006**, *25* (9), 1924–1933. <https://doi.org/10.1038/sj.emboj.7601076>.
- (137) Hsia, K. C.; Li, C. L.; Yuan, H. S. Structural and Functional Insight into Sugar-Nonspecific Nucleases in Host Defense. *Curr. Opin. Struct. Biol.* **2005**, *15* (1), 126–134. <https://doi.org/10.1016/j.sbi.2005.01.015>.
- (138) Champoux, J. J. DNA Topoisomerases: Structure, Function, and Mechanism. *Annual Review of Biochemistry*. Annual Reviews 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139, USA November 28, 2001, pp 369–413. <https://doi.org/10.1146/annurev.biochem.70.1.369>.
- (139) Gottlin, E. B.; Rudolph, A. E.; Zhao, Y.; Matthews, H. R.; Dixon, J. E. Catalytic Mechanism of the Phospholipase D Superfamily Proceeds via a Covalent Phosphohistidine Intermediate. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95* (16), 9202–9207. <https://doi.org/10.1073/pnas.95.16.9202>.
- (140) Buckle, A. M.; Fersht, A. R. Subsite Binding in an RNase: Structure of a Barnase-Tetranucleotide Complex at 1.76-Å Resolution. *Biochemistry* **1994**, *33* (7), 1644–1653. <https://doi.org/10.1021/bi00173a005>.
- (141) Correll, C. C.; Yang, X.; Gerczei, T.; Beneken, J.; Plantinga, M. J. RNA Recognition and Base Flipping by the Toxin Sarcin. *J. Synchrotron Radiat.* **2004**, *11* (1), 93–96. <https://doi.org/10.1107/S0909049503023896>.
- (142) Cochrane, J. C.; Strobel, S. A. Catalytic Strategies of Self-Cleaving Ribozymes. *Acc. Chem. Res.* **2008**, *41* (8), 1027–1035. <https://doi.org/10.1021/ar800050c>.

- (143) Calvin, K.; Li, H. RNA-Splicing Endonuclease Structure and Function. *Cell. Mol. Life Sci.* **2008**, *65* (7), 1176–1185. <https://doi.org/10.1007/S00018-008-7393-Y>.
- (144) Prieto, J.; Redondo, P.; Merino, N.; Villate, M.; Montoya, G.; Blanco, F. J.; Molina, R. Structure of the I-SceI Nuclease Complexed with Its DsDNA Target and Three Catalytic Metal Ions. *Acta Crystallogr. Sect. Struct. Biol. Commun.* **2016**, *72* (6), 473–479. <https://doi.org/10.1107/S2053230X16007512>.
- (145) Garcin, E. D.; Hosfield, D. J.; Desai, S. A.; Haas, B. J.; Björas, M.; Cunningham, R. P.; Tainer, J. A. DNA Apurinic-Apyrimidinic Site Binding and Excision by Endonuclease IV. *Nat. Struct. Mol. Biol.* **2008**, *15* (5), 515–522. <https://doi.org/10.1038/nsmb.1414>.
- (146) Romier, C.; Dominguez, R.; Lahm, A.; Dahl, O.; Suck, D. Recognition of Single-Stranded DNA by Nuclease P1: High Resolution Crystal Structures of Complexes with Substrate Analogs. *Proteins Struct. Funct. Genet.* **1998**, *32* (4), 414–424. [https://doi.org/10.1002/\(SICI\)1097-0134\(19980901\)32:4<414::AID-PROT2>3.0.CO;2-G](https://doi.org/10.1002/(SICI)1097-0134(19980901)32:4<414::AID-PROT2>3.0.CO;2-G).
- (147) AlMalki, F. A.; Flemming, C. S.; Zhang, J.; Feng, M.; Sedelnikova, S. E.; Ceska, T.; Rafferty, J. B.; Sayers, J. R.; Artymiuk, P. J. Direct Observation of DNA Threading in Flap Endonuclease Complexes. *Nat. Struct. Mol. Biol.* **2016**, *23* (7), 640–646. <https://doi.org/10.1038/nsmb.3241>.
- (148) Shi, Y.; Hellinga, H. W.; Beese, L. S. Interplay of Catalysis, Fidelity, Threading, and Processivity in the Exo- and Endonucleolytic Reactions of Human Exonuclease I. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (23), 6010–6015. <https://doi.org/10.1073/pnas.1704845114>.
- (149) Horton, N. C.; Perona, J. J. DNA Cleavage by EcoRV Endonuclease: Two Metal Ions in Three Metal Ion Binding Sites. *Biochemistry* **2004**, *43* (22), 6841–6857. <https://doi.org/10.1021/bi0499056>.
- (150) Samara, N. L.; Yang, W. Cation Trafficking Propels RNA Hydrolysis. *Nat. Struct. Mol. Biol.* **2018**, *25* (8), 715–721. <https://doi.org/10.1038/s41594-018-0099-4>.
- (151) Manigrasso, J.; De Vivo, M.; Palermo, G. Controlled Trafficking of Multiple and Diverse Cations Prompts Nucleic Acid Hydrolysis. *ACS Catal.* **2021**, *11* (14), 8786–8797. <https://doi.org/10.1021/acscatal.1c01825>.
- (152) Davis, L. G.; Dibner, M. D.; Battey, J. F. Restriction Endonucleases (REs) and Their Use. In *Basic Methods in Molecular Biology*; Elsevier, 1986; pp 51–57. <https://doi.org/10.1016/b978-0-444-01082-7.50021-7>.
- (153) Tsutakawa, S. E.; Classen, S.; Chapados, B. R.; Arvai, A. S.; Finger, L. D.; Guenther, G.; Tomlinson, C. G.; Thompson, P.; Sarker, A. H.; Shen, B.; Cooper, P. K.; Grasby, J. A.; Tainer, J. A. Human Flap Endonuclease Structures, DNA Double-Base Flipping, and a Unified Understanding of the FEN1 Superfamily. *Cell* **2011**, *145* (2), 198–211. <https://doi.org/10.1016/j.cell.2011.03.004>.
- (154) Park, J. E.; Heo, I.; Tian, Y.; Simanshu, D. K.; Chang, H.; Jee, D.; Patel, D. J.; Kim, V. N. Dicer Recognizes the 5' End of RNA for Efficient and Accurate Processing. *Nature* **2011**, *475* (7355), 201–205. <https://doi.org/10.1038/nature10198>.
- (155) Tsutakawa, S. E.; Lafrance-Vanasse, J.; Tainer, J. A. The Cutting Edges in DNA Repair, Licensing, and Fidelity: DNA and RNA Repair Nucleases Sculpt DNA to Measure Twice, Cut Once. *DNA Repair (Amst)*. **2014**, *19*, 95–107. <https://doi.org/10.1016/j.dnarep.2014.03.022>.

- (156) Williams, R. S.; Kunkel, T. A. FEN Nucleases: Bind, Bend, Fray, Cut. *Cell* **2011**, *145* (2), 171–172. <https://doi.org/10.1016/j.cell.2011.03.039>.
- (157) Grasby, J. A.; Finger, L. D.; Tsutakawa, S. E.; Atack, J. M.; Tainer, J. A. Unpairing and Gating: Sequence-Independent Substrate Recognition by FEN Superfamily Nucleases. *Trends in Biochemical Sciences*. Elsevier Current Trends February 1, 2012, pp 74–84. <https://doi.org/10.1016/j.tibs.2011.10.003>.
- (158) Tsutakawa, S. E.; Thompson, M. J.; Arvai, A. S.; Neil, A. J.; Shaw, S. J.; Algasier, S. I.; Kim, J. C.; Finger, L. D.; Jardine, E.; Gotham, V. J. B.; Sarker, A. H.; Her, M. Z.; Rashid, F.; Hamdan, S. M.; Mirkin, S. M.; Grasby, J. A.; Tainer, J. A. Phosphate Steering by Flap Endonuclease 1 Promotes 5'-Flap Specificity and Incision to Prevent Genome Instability. *Nat. Commun.* **2017**, *8* (1), 1–15. <https://doi.org/10.1038/ncomms15855>.
- (159) González-Corrochano, R.; Ruiz, F. M.; Taylor, N. M. I.; Huecas, S.; Drakulic, S.; Spínola-Amilibia, M.; Fernández-Tornero, C. The Crystal Structure of Human XPG, the Xeroderma Pigmentosum Group G Endonuclease, Provides Insight into Nucleotide Excision DNA Repair. *Nucleic Acids Res.* **2020**, *48* (17), 9943–9958. <https://doi.org/10.1093/nar/gkaa688>.
- (160) Tsutakawa, S. E.; Sarker, A. H.; Ng, C.; Arvai, A. S.; Shin, D. S.; Shih, B.; Jiang, S.; Thwin, A. C.; Tsai, M. S.; Willcox, A.; Her, M. Z.; Trego, K. S.; Raetz, A. G.; Rosenberg, D.; Bacolla, A.; Hammel, M.; Griffith, J. D.; Cooper, P. K.; Tainer, J. A. Human XPG Nuclease Structure, Assembly, and Activities with Insights for Neurodegeneration and Cancer from Pathogenic Mutations. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117* (25), 14127–14138. <https://doi.org/10.1073/pnas.1921311117>.
- (161) Liu, Y.; Freeman, A. D.; Déclais, A. C.; Lilley, D. M. J. A Monovalent Ion in the DNA Binding Interface of the Eukaryotic Junction-Resolving Enzyme GEN1. *Nucleic Acids Res.* **2018**, *46* (20), 11089–11098. <https://doi.org/10.1093/nar/gky863>.
- (162) Shaw, S. J.; Finger, L. D.; Grasby, J. A. Human Exonuclease 1 Threads 5'-Flap Substrates through Its Helical Arch. *Biochemistry* **2017**, *56* (29), 3704–3707. <https://doi.org/10.1021/acs.biochem.7b00507>.
- (163) Patel, N.; Atack, J. M.; Finger, L. D.; Exell, J. C.; Thompson, P.; Tsutakawa, S.; Tainer, J. A.; Williams, D. M.; Grasby, J. A. Flap Endonucleases Pass 5'-Flaps through a Flexible Arch Using a Disorder-Thread-Order Mechanism to Confer Specificity for Free 5'-Ends. *Nucleic Acids Res.* **2012**, *40* (10), 4507–4519. <https://doi.org/10.1093/nar/gks051>.
- (164) Orans, J.; McSweeney, E. A.; Iyer, R. R.; Hast, M. A.; Hellinga, H. W.; Modrich, P.; Beese, L. S. Structures of Human Exonuclease 1 DNA Complexes Suggest a Unified Mechanism for Nuclease Family. *Cell* **2011**, *145* (2), 212–223. <https://doi.org/10.1016/j.cell.2011.03.005>.
- (165) Lee, B. I.; Wilson, D. M. The RAD2 Domain of Human Exonuclease 1 Exhibits 5' to 3' Exonuclease and Flap Structure-Specific Endonuclease Activities. *J. Biol. Chem.* **1999**, *274* (53), 37763–37769. <https://doi.org/10.1074/jbc.274.53.37763>.
- (166) Tsutakawa, S. E.; Tainer, J. A. Double Strand Binding-Single Strand Incision Mechanism for Human Flap Endonuclease: Implications for the Superfamily. *Mech. Ageing Dev.* **2012**, *133* (4), 195–202. <https://doi.org/10.1016/j.mad.2011.11.009>.
- (167) Xu, Y.; Li, Z. CRISPR-Cas Systems: Overview, Innovations and Applications in Human Disease Research and Gene Therapy. *Comput. Struct. Biotechnol. J.* **2020**, *18*, 2401–2415. <https://doi.org/10.1016/j.csbj.2020.08.031>.
- (168) Li, H.; Yang, Y.; Hong, W.; Huang, M.; Wu, M.; Zhao, X. Applications of Genome Editing

Technology in the Targeted Therapy of Human Diseases: Mechanisms, Advances and Prospects. *Signal Transduct. Target. Ther.* **2020**, *5* (1), 1–23. <https://doi.org/10.1038/s41392-019-0089-y>.

- (169) Scott, T. A.; Morris, K. V. Designer Nucleases to Treat Malignant Cancers Driven by Viral Oncogenes. *Virolog. J.* **2021**, *18* (1), 1–15. <https://doi.org/10.1186/s12985-021-01488-1>.
- (170) Dash, P. K.; Kaminski, R.; Bella, R.; Su, H.; Mathews, S.; Ahooyi, T. M.; Chen, C.; Mancuso, P.; Sariyer, R.; Ferrante, P.; Donadoni, M.; Robinson, J. A.; Sillman, B.; Lin, Z.; Hilaire, J. R.; Banoub, M.; Elango, M.; Gautam, N.; Mosley, R. L.; Poluektova, L. Y.; McMillan, J. E.; Bade, A. N.; Gorantla, S.; Sariyer, I. K.; Burdo, T. H.; Young, W. Bin; Amini, S.; Gordon, J.; Jacobson, J. M.; Edagwa, B.; Khalili, K.; Gendelman, H. E. Sequential LASER ART and CRISPR Treatments Eliminate HIV-1 in a Subset of Infected Humanized Mice. *Nat. Commun.* **2019**, *10* (1), 1–20. <https://doi.org/10.1038/s41467-019-10366-y>.
- (171) Kostyushev, D.; Brezgin, S.; Kostyusheva, A.; Zarifyan, D.; Goptar, I.; Chulanov, V. Orthologous CRISPR/Cas9 Systems for Specific and Efficient Degradation of Covalently Closed Circular DNA of Hepatitis B Virus. *Cell. Mol. Life Sci.* **2019**, *76* (9), 1779–1794. <https://doi.org/10.1007/s00018-019-03021-8>.
- (172) Kuscü, C.; Arslan, S.; Singh, R.; Thorpe, J.; Adli, M. Genome-Wide Analysis Reveals Characteristics of off-Target Sites Bound by the Cas9 Endonuclease. *Nat. Biotechnol.* **2014**, *32* (7), 677–683. <https://doi.org/10.1038/nbt.2916>.
- (173) Chew, W. L.; Tabebordbar, M.; Cheng, J. K. W.; Mali, P.; Wu, E. Y.; Ng, A. H. M.; Zhu, K.; Wagers, A. J.; Church, G. M. A Multifunctional AAV-CRISPR-Cas9 and Its Host Response. *Nat. Methods* **2016**, *13* (10), 868–874. <https://doi.org/10.1038/nmeth.3993>.
- (174) Haapaniemi, E.; Botla, S.; Persson, J.; Schmierer, B.; Taipale, J. CRISPR-Cas9 Genome Editing Induces a P53-Mediated DNA Damage Response. *Nat. Med.* **2018**, *24* (7), 927–930. <https://doi.org/10.1038/s41591-018-0049-z>.
- (175) Xu, X.; Wan, T.; Xin, H.; Li, D.; Pan, H.; Wu, J.; Ping, Y. Delivery of CRISPR/Cas9 for Therapeutic Genome Editing. *J. Gene Med.* **2019**, *21* (7), e3107. <https://doi.org/10.1002/jgm.3107>.
- (176) Kumar, S.; Peng, X.; Daley, J.; Yang, L.; Shen, J.; Nguyen, N.; Bae, G.; Niu, H.; Peng, Y.; Hsieh, H. J.; Wang, L.; Rao, C.; Stephan, C. C.; Sung, P.; Ira, G.; Peng, G. Inhibition of DNA2 Nuclease as a Therapeutic Strategy Targeting Replication Stress in Cancer Cells. *Oncogenesis* **2017**, *6* (4), e319–e319. <https://doi.org/10.1038/oncsis.2017.15>.
- (177) Bartosova, Z.; Krejci, L. Nucleases in Homologous Recombination as Targets for Cancer Therapy. *FEBS Lett.* **2014**, *588* (15), 2446–2456. <https://doi.org/10.1016/j.febslet.2014.06.010>.
- (178) Zheng, L.; Dai, H.; Zhou, M.; Li, M.; Singh, P.; Qiu, J.; Tsark, W.; Huang, Q.; Kernstine, K.; Zhang, X.; Lin, D.; Shen, B. Fen1 Mutations Result in Autoimmunity, Chronic Inflammation and Cancers. *Nat. Med.* **2007**, *13* (7), 812–819. <https://doi.org/10.1038/nm1599>.
- (179) Liu, S.; Lu, G.; Ali, S.; Liu, W.; Zheng, L.; Dai, H.; Li, H.; Xu, H.; Hua, Y.; Zhou, Y.; Ortega, J.; Li, G.-M.; Kunkel, T. A.; Shen, B. Okazaki Fragment Maturation Involves A-segment Error Editing by the Mammalian FEN 1/MutS α Functional Complex. *EMBO J.* **2015**, *34* (13), 1829–1843. <https://doi.org/10.15252/embj.201489865>.
- (180) Sun, X.; Zheng, L.; Shen, B. Functional Alterations of Human Exonuclease 1 Mutants Identified in Atypical Hereditary Nonpolyposis Colorectal Cancer Syndrome. *Cancer Res.*

2002, 62 (21), 6026–6030.

- (181) Karplus, M.; McCammon, J. A. Molecular Dynamics Simulations of Biomolecules. *Nat. Struct. Biol.* **2002**, 9 (9), 646–652. <https://doi.org/10.1038/nsb0902-646>.
- (182) Jones, J. E. On the Determination of Molecular Fields.—II. From the Equation of State of a Gas. In *Proc. R. Soc. A Math. Phys. Eng. Sci.*; 1924; pp 463–477.
- (183) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, 25 (9), 1157–1174. <https://doi.org/10.1002/jcc.20035>.
- (184) Schmid, N.; Christ, C. D.; Christen, M.; Eichenberger, A. P.; Van Gunsteren, W. F. Architecture, Implementation and Parallelisation of the GROMOS Software for Biomolecular Simulation. *Comput. Phys. Commun.* **2012**, 183 (4), 890–903. <https://doi.org/10.1016/j.cpc.2011.12.014>.
- (185) Brooks, B. R.; Brucoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* **1983**, 4 (2), 187–217. <https://doi.org/10.1002/jcc.540040211>.
- (186) Anderson, C. F.; Record, M. T. Salt-Nucleic Acid Interactions. *Annu. Rev. Phys. Chem.* **1995**, 46, 657–700.
- (187) Baldwin, R. L. How Hofmeister Ion Interactions Affect Protein Stability. *Biophys. J.* **1996**, 71 (4), 2056–2063. [https://doi.org/10.1016/S0006-3495\(96\)79404-3](https://doi.org/10.1016/S0006-3495(96)79404-3).
- (188) Li, P.; Merz, K. M. Metal Ion Modeling Using Classical Mechanics. *Chem. Rev.* **2017**, 117 (3), 1564–1686. <https://doi.org/10.1021/acs.chemrev.6b00440>.
- (189) Panteva, M. T.; Giambaşu, G. M.; York, D. M. Comparison of Structural, Thermodynamic, Kinetic and Mass Transport Properties of Mg²⁺ Ion Models Commonly Used in Biomolecular Simulations. *J. Comput. Chem.* **2015**, 36 (13), 970–982. <https://doi.org/10.1002/jcc.23881>.
- (190) Panteva, M. T.; Giambaşu, G. M.; York, D. M. Force Field for Mg²⁺, Mn²⁺, Zn²⁺, and Cd²⁺ Ions That Have Balanced Interactions with Nucleic Acids. *J. Phys. Chem. B* **2015**, 119 (50), 15460–15470. <https://doi.org/10.1021/acs.jpcc.5b10423>.
- (191) Lemkul, J. A.; Mackerell, A. D. Balancing the Interactions of Mg²⁺ in Aqueous Solution and with Nucleic Acid Moieties for a Polarizable Force Field Based on the Classical Drude Oscillator Model. *J. Phys. Chem. B* **2016**, 120 (44), 11436–11448. <https://doi.org/10.1021/acs.jpcc.6b09262>.
- (192) Peters, M. B.; Yang, Y.; Wang, B.; Füsti-Molnár, L.; Weaver, M. N.; Merz, K. M. Structural Survey of Zinc-Containing Proteins and Development of the Zinc AMBER Force Field (ZAFF). *J. Chem. Theory Comput.* **2010**, 6 (9), 2935–2947. <https://doi.org/10.1021/ct1002626>.
- (193) Hoops, S. C.; Anderson, K. W.; Merz, K. M. Force Field Design for Metalloproteins. *J. Am. Chem. Soc.* **1991**, 113 (22), 8262–8270. <https://doi.org/10.1021/ja00022a010>.
- (194) Stote, R. H.; Karplus, M. Zinc Binding in Proteins and Solution: A Simple but Accurate Nonbonded Representation. *Proteins Struct. Funct. Bioinforma.* **1995**, 23 (1), 12–31. <https://doi.org/10.1002/prot.340230104>.
- (195) Li, P.; Merz, K. M. Taking into Account the Ion-Induced Dipole Interaction in the

Nonbonded Model of Ions. *J. Chem. Theory Comput.* **2014**, *10* (1), 289–297.
<https://doi.org/10.1021/ct400751u>.

- (196) Joung, I. S.; Cheatham, T. E. Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem. B* **2008**, *112* (30), 9020–9041. <https://doi.org/10.1021/jp8001614>.
- (197) Åqvist, J. Ion-Water Interaction Potentials Derived from Free Energy Perturbation Simulations. *J. Phys. Chem.* **1990**, *94* (21), 8021–8024. <https://doi.org/10.1021/j100384a009>.
- (198) Allnér, O.; Nilsson, L.; Villa, A. Magnesium Ion-Water Coordination and Exchange in Biomolecular Simulations. *J. Chem. Theory Comput.* **2012**, *8* (4), 1493–1502. <https://doi.org/10.1021/ct3000734>.
- (199) Li, P.; Roberts, B. P.; Chakravorty, D. K.; Merz, K. M. Rational Design of Particle Mesh Ewald Compatible Lennard-Jones Parameters for +2 Metal Cations in Explicit Solvent. *J. Chem. Theory Comput.* **2013**, *9* (6), 2733–2748. <https://doi.org/10.1021/ct400146w>.
- (200) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79* (2), 926–935. <https://doi.org/10.1063/1.445869>.
- (201) Kusalik, P. G.; Patey, G. N. The Thermodynamic Properties of Electrolyte Solutions: Some Formal Results. *J. Chem. Phys.* **1987**, *86* (9), 5110–5116. <https://doi.org/10.1063/1.452629>.
- (202) Lyubartsev, A. P.; Laaksonen, A. Osmotic and Activity Coefficients from Effective Potentials for Hydrated Ions. *Phys. Rev. E - Stat. Physics, Plasmas, Fluids, Relat. Interdiscip. Top.* **1997**, *55* (5), 5689–5696. <https://doi.org/10.1103/PhysRevE.55.5689>.
- (203) Hart, K.; Foloppe, N.; Baker, C. M.; Denning, E. J.; Nilsson, L.; MacKerell, A. D. Optimization of the CHARMM Additive Force Field for DNA: Improved Treatment of the BI/BII Conformational Equilibrium. *J. Chem. Theory Comput.* **2012**, *8* (1), 348–362. <https://doi.org/10.1021/ct200723y>.
- (204) Krepl, M.; Zgarbová, M.; Stadlbauer, P.; Otyepka, M.; Banáš, P.; Koča, J.; Cheatham, T. E.; Jurečka, P.; Šponer, J. Reference Simulations of Noncanonical Nucleic Acids with Different χ Variants of the AMBER FORCE Field: Quadruplex DNA, Quadruplex RNA, and Z-DNA. *J. Chem. Theory Comput.* **2012**, *8* (7), 2506–2520. <https://doi.org/10.1021/ct300275s>.
- (205) Langley, D. R. Molecular Dynamic Simulations of Environment and Sequence Dependent Dna Conformations: The Development of the Bms Nucleic Acid Force Field and Comparison with Experimental Results. *J. Biomol. Struct. Dyn.* **1998**, *16* (3), 487–509. <https://doi.org/10.1080/07391102.1998.10508265>.
- (206) Ivani, I.; Dans, P. D.; Noy, A.; Pérez, A.; Faustino, I.; Hospital, A.; Walther, J.; Andrio, P.; Goñi, R.; Balaceanu, A.; Portella, G.; Battistini, F.; Gelpí, J. L.; González, C.; Vendruscolo, M.; Loughton, C. A.; Harris, S. A.; Case, D. A.; Orozco, M. Parmbsc1: A Refined Force Field for DNA Simulations. *Nat. Methods* **2015**, *13* (1), 55–58. <https://doi.org/10.1038/nmeth.3658>.
- (207) Zgarbová, M.; Šponer, J.; Otyepka, M.; Cheatham, T. E.; Galindo-Murillo, R.; Jurečka, P. Refinement of the Sugar-Phosphate Backbone Torsion Beta for AMBER Force Fields Improves the Description of Z- and B-DNA. *J. Chem. Theory Comput.* **2015**, *11* (12), 5723–5736. <https://doi.org/10.1021/acs.jctc.5b00716>.
- (208) Pérez, A.; Marchán, I.; Svozil, D.; Šponer, J.; Cheatham, T. E.; Loughton, C. A.; Orozco, M.

Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of α/γ Conformers. *Biophys. J.* **2007**, *92* (11), 3817–3829.
<https://doi.org/10.1529/biophysj.106.097782>.

- (209) Galindo-Murillo, R.; Robertson, J. C.; Zgarbová, M.; Šponer, J.; Otyepka, M.; Jurečka, P.; Cheatham, T. E. Assessing the Current State of Amber Force Field Modifications for DNA. *J. Chem. Theory Comput.* **2016**, *12* (8), 4114–4127. <https://doi.org/10.1021/acs.jctc.6b00186>.
- (210) Dans, P. D.; Ivani, I.; Hospital, A.; Portella, G.; González, C.; Orozco, M. How Accurate Are Accurate Force-Fields for B-DNA? *Nucleic Acids Res.* **2017**, *45* (7), 4217–4230.
<https://doi.org/10.1093/nar/gkw1355>.
- (211) Bussi, G.; Laio, A.; Parrinello, M. Equilibrium Free Energies from Nonequilibrium Metadynamics. *Phys. Rev. Lett.* **2006**, *96* (9).
<https://doi.org/10.1103/PhysRevLett.96.090601>.
- (212) Bussi, G.; Laio, A. Using Metadynamics to Explore Complex Free-Energy Landscapes. *Nat. Rev. Phys.* **2020**, *2* (4), 200–212. <https://doi.org/10.1038/s42254-020-0153-0>.
- (213) Bernardi, R. C.; Melo, M. C. R.; Schulten, K. Enhanced Sampling Techniques in Molecular Dynamics Simulations of Biological Systems. *Biochim. Biophys. Acta - Gen. Subj.* **2015**, *1850* (5), 872–877. <https://doi.org/10.1016/J.BBAGEN.2014.10.019>.
- (214) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (20), 12562–12566. <https://doi.org/10.1073/pnas.202427399>.
- (215) Laio, A.; Rodriguez-Forteza, A.; Gervasio, F. L.; Ceccarelli, M.; Parrinello, M. Assessing the Accuracy of Metadynamics. *J. Phys. Chem. B* **2005**, *109* (14), 6714–6721.
<https://doi.org/10.1021/jp045424k>.
- (216) Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **2008**, *100* (2), 020603.
<https://doi.org/10.1103/PhysRevLett.100.020603>.
- (217) La Sala, G.; Riccardi, L.; Gaspari, R.; Cavalli, A.; Hantschel, O.; De Vivo, M. HRD Motif as the Central Hub of the Signaling Network for Activation Loop Autophosphorylation in Abl Kinase. *J. Chem. Theory Comput.* **2016**, *12* (11), 5563–5574.
<https://doi.org/10.1021/acs.jctc.6b00600>.
- (218) Branduardi, D.; Gervasio, F. L.; Parrinello, M. From A to B in Free Energy Space. *J. Chem. Phys.* **2007**, *126* (5), 054103. <https://doi.org/10.1063/1.2432340>.
- (219) Pavlov, Y. I.; Shcherbakova, P. V.; Rogozin, I. B. Roles of DNA Polymerases in Replication, Repair, and Recombination in Eukaryotes. *International Review of Cytology*. Academic Press January 1, 2006, pp 41–132. [https://doi.org/10.1016/S0074-7696\(06\)55002-8](https://doi.org/10.1016/S0074-7696(06)55002-8).
- (220) Nishino, T.; Morikawa, K. Structure and Function of Nucleases in DNA Repair: Shape, Grip and Blade of the DNA Scissors. *Oncogene*. Nature Publishing Group December 16, 2002, pp 9022–9032. <https://doi.org/10.1038/sj.onc.1206135>.
- (221) T. M. Marti, O. F. DNA Repair Nucleases. *Cell. Mol. Life Sci.* **2004**, *61* (3), 336–354.
- (222) Potapov, V.; Fu, X.; Dai, N.; Corrêa, I. R.; Tanner, N. A.; Ong, J. L. Base Modifications Affecting RNA Polymerase and Reverse Transcriptase Fidelity. *Nucleic Acids Res.* **2018**, *46* (11), 5753–5763. <https://doi.org/10.1093/nar/gky341>.
- (223) Nimonkar, A. V.; Ozsoy, A. Z.; Genschel, J.; Modrich, P.; Kowalczykowski, S. C. Human

- Exonuclease 1 and BLM Helicase Interact to Resect DNA and Initiate DNA Repair. *Proc. Natl. Acad. Sci.* **2008**, *105* (44), 16906–16911. <https://doi.org/10.1073/pnas.0809380105>.
- (224) Zheng, L.; Jia, J.; Finger, L. D.; Guo, Z.; Zer, C.; Shen, B. Functional Regulation of FEN1 Nuclease and Its Link to Cancer. *Nucleic Acids Research*. 2011, pp 781–794. <https://doi.org/10.1093/nar/gkq884>.
- (225) Peltomäki, P. Role of DNA Mismatch Repair Defects in the Pathogenesis of Human Cancer. *J. Clin. Oncol.* **2003**, *21* (6), 1174–1179. <https://doi.org/10.1200/JCO.2003.04.060>.
- (226) Dai, Y.; Tang, Z.; Yang, Z.; Zhang, L.; Deng, Q.; Zhang, X.; Yu, Y.; Liu, X.; Zhu, J. EXO1 Overexpression Is Associated with Poor Prognosis of Hepatocellular Carcinoma Patients. *Cell Cycle* **2018**, *17* (19–20), 2386–2397. <https://doi.org/10.1080/15384101.2018.1534511>.
- (227) Uson, M. L.; Carl, A.; Goldgur, Y.; Shuman, S. Crystal Structure and Mutational Analysis of Mycobacterium Smegmatis FenA Highlight Active Site Amino Acids and Three Metal Ions Essential for Flap Endonuclease and 5 Exonuclease Activities. *Nucleic Acids Res.* **2018**, *46* (8), 4164–4175. <https://doi.org/10.1093/nar/gky238>.
- (228) Palermo, G.; Stenta, M.; Cavalli, A.; Dal Peraro, M.; De Vivo, M. Molecular Simulations Highlight the Role of Metals in Catalysis and Inhibition of Type II Topoisomerase. *J. Chem. Theory Comput.* **2013**, *9* (2), 857–862. <https://doi.org/10.1021/ct300691u>.
- (229) Schmidt, B. H.; Burgin, A. B.; Deweese, J. E.; Osheroff, N.; Berger, J. M. A Novel and Unified Two-Metal Mechanism for DNA Cleavage by Type II and IA Topoisomerases. *Nature* **2010**, *465* (7298), 641–644. <https://doi.org/10.1038/nature08974>.
- (230) Perera, L.; Freudenthal, B. D.; Beard, W. A.; Shock, D. D.; Pedersen, L. G.; Wilson, S. H.; Broyde, S. Requirement for Transient Metal Ions Revealed through Computational Analysis for DNA Polymerase Going in Reverse. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (38), E5228–E5236. <https://doi.org/10.1073/pnas.1511207112>.
- (231) Jimeno, S.; Herrera-Moyano, E.; Ortega, P.; Aguilera, A. Differential Effect of the Overexpression of Rad2/XPG Family Endonucleases on Genome Integrity in Yeast and Human Cells. *DNA Repair (Amst)*. **2017**, *57*, 66–75. <https://doi.org/10.1016/j.dnarep.2017.06.030>.
- (232) Emmert, S. The Human XPG Gene: Gene Architecture, Alternative Splicing and Single Nucleotide Polymorphisms. *Nucleic Acids Res.* **2001**, *29* (7), 1443–1452. <https://doi.org/10.1093/nar/29.7.1443>.
- (233) Qiu, J.; Qian, Y.; Chen, V.; Guan, M.-X.; Shen, B. Human Exonuclease 1 Functionally Complements Its Yeast Homologues in DNA Recombination, RNA Primer Removal, and Mutation Avoidance. *J. Biol. Chem.* **1999**, *274* (25), 17893–17900. <https://doi.org/10.1074/jbc.274.25.17893>.
- (234) Genschel, J.; Bazemore, L. R.; Modrich, P. Human Exonuclease I Is Required for 5' and 3' Mismatch Repair. *J. Biol. Chem.* **2002**, *277* (15), 13302–13311. <https://doi.org/10.1074/jbc.M111854200>.
- (235) Wei, K.; Clark, A. B.; Wong, E.; Kane, M. F.; Mazur, D. J.; Parris, T.; Kolas, N. K.; Russell, R.; Hou, H.; Kneitz, B.; Yang, G.; Kunkel, T. A.; Kolodner, R. D.; Cohen, P. E.; Edelman, W. Inactivation of Exonuclease I in Mice Results in DNA Mismatch Repair Defects, Increased Cancer Susceptibility, and Male and Female Sterility. *Genes Dev.* **2003**, *17* (5), 603–614. <https://doi.org/10.1101/gad.1060603>.

- (236) Zhu, Z.; Chung, W. H.; Shim, E. Y.; Lee, S. E.; Ira, G. Sgs1 Helicase and Two Nucleases Dna2 and Exo1 Resect DNA Double-Strand Break Ends. *Cell* **2008**, *134* (6), 981–994. <https://doi.org/10.1016/j.cell.2008.08.037>.
- (237) Vallur, A. C.; Maizels, N. Complementary Roles for Exonuclease 1 and Flap Endonuclease 1 in Maintenance of Triplet Repeats. *J. Biol. Chem.* **2010**, *285* (37), 28514–28519. <https://doi.org/10.1074/jbc.M110.132738>.
- (238) Keijzers, G.; Bohr, V.; Juel Rasmussen, L. Human Exonuclease 1 (EXO1) Activity Characterization and Its Function on FLAP Structures. *Biosci. Rep.* **2015**, *1*, 1–13. <https://doi.org/10.1042/BSR20150058>.
- (239) Dupureur, C. M. One Is Enough: Insights into the Two-Metal Ion Nuclease Mechanism from Global Analysis and Computational Studies. *Metallomics*. 2010, pp 609–620. <https://doi.org/10.1039/c0mt00013b>.
- (240) Beese, L. S.; Steitz, T. A. Structural Basis for the 3'-5' Exonuclease Activity of Escherichia Coli DNA Polymerase I: A Two Metal Ion Mechanism. *EMBO J.* **1991**, *10* (1), 25–33. <https://doi.org/10.1002/j.1460-2075.1991.tb07917.x>.
- (241) Tomlinson, C. G.; Atack, J. M.; Chapados, B.; Tainer, J. A.; Grasby, J. A. Substrate Recognition and Catalysis by Flap Endonucleases and Related Enzymes. *Biochem. Soc. Trans.* **2010**, *38* (2), 433–437. <https://doi.org/10.1042/BST0380433>.
- (242) Bennet, I. A.; David Finger, L.; Baxter, N. J.; Ambrose, B.; Hounslow, A. M.; Thompson, M. J.; Exell, J. C.; Nur Nazihah, N. N. B.; Craggs, T. D.; Waltho, J. P.; Grasby, J. A. Regional Conformational Flexibility Couples Substrate Specificity and Scissile Phosphate Diester Selectivity in Human Flap Endonuclease 1. *Nucleic Acids Res.* **2018**, *46* (11), 5618–5633. <https://doi.org/10.1093/nar/gky293>.
- (243) Black, C. B.; Huang, H. W.; Cowan, J. A. Biological Coordination Chemistry of Magnesium, Sodium, and Potassium Ions. Protein and Nucleotide Binding Sites. *Coord. Chem. Rev.* **1994**, *135–136* (C), 165–202. [https://doi.org/10.1016/0010-8545\(94\)80068-5](https://doi.org/10.1016/0010-8545(94)80068-5).
- (244) Ho, M.-H.; De Vivo, M.; Dal Peraro, M.; Klein, M. L. Understanding the Effect of Magnesium Ion Concentration on the Catalytic Activity of Ribonuclease H through Computation: Does a Third Metal Binding Site Modulate Endonuclease Activity? *J. Am. Chem. Soc.* **2011**, *132* (39), 13702–13712. <https://doi.org/10.1126/scisignal.2001449.Engineering>.
- (245) Hwang, W.; Yoo, J.; Lee, Y.; Park, S.; Hoang, P. L.; Cho, H.; Yu, J.; Hoa Vo, T. M.; Shin, M.; Jin, M. S.; Park, D.; Hyeon, C.; Lee, G. Dynamic Coordination of Two-Metal-Ions Orchestrates λ -Exonuclease Catalysis. *Nat. Commun.* **2018**, *9* (1), 4404. <https://doi.org/10.1038/s41467-018-06750-9>.
- (246) Sasnauskas, G.; Jeltsch, A.; Pingoud, A.; Siksnys, V. Plasmid DNA Cleavage by MunI Restriction Enzyme: Single-Turnover and Steady-State Kinetic Analysis. *Biochemistry* **1999**, *38* (13), 4028–4036. <https://doi.org/10.1021/bi982456n>.
- (247) Xie, F.; Dupureur, C. M. Kinetic Analysis of Product Release and Metal Ions in a Metallonuclease. *Arch. Biochem. Biophys.* **2009**, *483* (1), 1–9. <https://doi.org/10.1016/j.abb.2009.01.001>.
- (248) Sengerová, B.; Tomlinson, C.; Atack, J. M.; Williams, R.; Sayers, J. R.; Williams, N. H.; Grasby, J. A. Brønsted Analysis and Rate-Limiting Steps for the T5 Flap Endonuclease Catalyzed Hydrolysis of Exonucleolytic Substrates. *Biochemistry* **2010**, *49* (37), 8085–8093.

<https://doi.org/10.1021/bi100895j>.

- (249) Taylor, J. D.; Halford, S. E. Discrimination between DNA Sequences by the EcoRV Restriction Endonuclease. *Biochemistry* **1989**, *28* (15), 6198–6207. <https://doi.org/10.1021/bi00441a011>.
- (250) Yang, C. C.; Baxter, B. K.; Topal, M. D. DNA Cleavage by NaeI: Protein Purification, Rate-Limiting Step, and Accuracy. *Biochemistry* **1994**, *33* (49), 14918–14925. <https://doi.org/10.1021/bi00253a031>.
- (251) Williams, R.; Sengerová, B.; Osborne, S.; Syson, K.; Ault, S.; Kilgour, A.; Chapados, B. R.; Tainer, J. A.; Sayers, J. R.; Grasby, J. A. Comparison of the Catalytic Parameters and Reaction Specificities of a Phage and an Archaeal Flap Endonuclease. *J. Mol. Biol.* **2007**, *371* (1), 34–48. <https://doi.org/10.1016/j.jmb.2007.04.063>.
- (252) Blanchard, M. S.; Theimer, C. A.; Sengerová, B.; Singh, P.; Chavez, V.; Liu, F.; Grasby, J. A.; Shen, B. The 3-Flap Pocket of Human Flap Endonuclease 1 Is Critical for Substrate Binding and Catalysis * □ S L. David Finger ‡1. **2009**. <https://doi.org/10.1074/jbc.M109.015065>.
- (253) Kuznetsova, A. A.; Fedorova, O. S.; Kuznetsov, N. A. Kinetic Features of 3'-5' Exonuclease Activity of Human AP-Endonuclease Ape1. *Molecules* **2018**, *23* (9), 2101. <https://doi.org/10.3390/molecules23092101>.
- (254) Warshel, A.; Levitt, M. Theoretical Studies of Enzymic Reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme. *J. Mol. Biol.* **1976**, *103*, 227–249. <https://doi.org/10.1007/s002149900077>.
- (255) Brunk, E.; Rothlisberger, U. Mixed Quantum Mechanical/Molecular Mechanical Molecular Dynamics Simulations of Biological Systems in Ground and Electronically Excited States. *Chem. Rev.* **2015**, *115* (12), 6217–6263. <https://doi.org/10.1021/cr500628b>.
- (256) Field, M. J.; Bash, P. A.; Karplus, M. A Combined Quantum Mechanical and Molecular Mechanical Potential for Molecular Dynamics Simulations. *J. Comput. Chem.* **1990**, *11* (6), 700–733. <https://doi.org/10.1002/jcc.540110605>.
- (257) Saul, N. B.; Christian, W. A General Method Applicable To Search for Similarities in Amino Acid Sequence of Two Proteins. *J. Mol. Biol.* **1970**, *48* (3), 443–453. [https://doi.org/10.1016/0022-2836\(70\)90057-4](https://doi.org/10.1016/0022-2836(70)90057-4).
- (258) Szymanski, M. R.; Yu, W.; Gmyrek, A. M.; White, M. A.; Molineux, I. J.; Lee, J. C.; Yin, Y. W. A Domain in Human EXOG Converts Apoptotic Endonuclease to DNA-Repair Exonuclease. *Nat. Commun.* **2017**, *8*, 14959. <https://doi.org/10.1038/ncomms14959>.
- (259) Wu, C.-C.; Lin, J. L. J.; Yang-Yen, H.-F.; Yuan, H. S. A Unique Exonuclease ExoG Cleaves between RNA and DNA in Mitochondrial DNA Replication. *Nucleic Acids Res.* **2019**, *1*. <https://doi.org/10.1093/nar/gkz189>.
- (260) Zhang, J.; McCabe, K. A.; Bell, C. E. Crystal Structures of λ Exonuclease in Complex with DNA Suggest an Electrostatic Ratchet Mechanism for Processivity. *Proc. Natl. Acad. Sci.* **2011**, *108* (29), 11872–11877. <https://doi.org/10.1073/pnas.1103467108/-/DCSupplemental>. www.pnas.org/cgi/doi/10.1073/pnas.1103467108.
- (261) Zhang, J.; Pan, X.; Bell, C. E. Crystal Structure of λ Exonuclease in Complex with DNA and Ca²⁺. *Biochemistry* **2014**, *53* (47), 7415–7425. <https://doi.org/10.1021/bi501155q>.
- (262) Cheng, K.; Xu, H.; Chen, X.; Wang, L.; Tian, B.; Zhao, Y.; Hua, Y. Structural Basis for

DNA 5'-End Resection by RecJ. *Elife* **2016**, 5 (e14294), 1–21.
<https://doi.org/10.7554/eLife.14294>.

- (263) Zhang, F.; Shi, J.; Chen, S. H.; Bian, C.; Yu, X. The PIN Domain of EXO1 Recognizes Poly(ADP-Ribose) in DNA Damage Response. *Nucleic Acids Res.* **2015**, 43 (22), 10782–10794. <https://doi.org/10.1093/nar/gkv939>.
- (264) Palermo, G.; Miao, Y.; Walker, R. C.; Jinek, M.; McCammon, J. A. CRISPR-Cas9 Conformational Activation as Elucidated from Enhanced Molecular Simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, 114 (28), 7260–7265. <https://doi.org/10.1073/pnas.1707645114>.
- (265) Mulholland, A. J.; Roitberg, A. E.; Tuñón, I. Enzyme Dynamics and Catalysis in the Mechanism of DNA Polymerase. *Theor. Chem. Acc.* **2012**, 131 (12), 1–4. <https://doi.org/10.1007/s00214-012-1286-8>.
- (266) Riccardi, L.; Genna, V.; De Vivo, M. Metal–Ligand Interactions in Drug Design. *Nat. Rev. Chem.* **2018**, 2 (7), 100–112. <https://doi.org/10.1038/s41570-018-0018-6>.
- (267) De Vivo, M. Bridging Quantum Mechanics and Structure-Based Drug Design. *Front. Biosci.* **2011**, 16 (5), 1619–1633. <https://doi.org/10.2741/3809>.
- (268) Genna, V.; Marcia, M.; Vivo, M. De. A Transient and Flexible Cation– π Interaction Promotes Hydrolysis of Nucleic Acids in DNA and RNA Nucleases. *J. Am. Chem. Soc.* **2019**, 141 (27), 10770–10776. <https://doi.org/10.1021/jacs.9b03663>.
- (269) Yan, C.; Dodd, T.; He, Y.; Tainer, J. A.; Tsutakawa, S. E.; Ivanov, I. Transcription Preinitiation Complex Structure and Dynamics Provide Insight into Genetic Diseases. *Nat. Struct. Mol. Biol.* **2019**, 26 (6), 397–406. <https://doi.org/10.1038/s41594-019-0220-3>.
- (270) Casalino, L.; Palermo, G.; Spinello, A.; Rothlisberger, U.; Magistrato, A. All-Atom Simulations Disentangle the Functional Dynamics Underlying Gene Maturation in the Intron Lariat Spliceosome. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, 115 (26), 6584–6589. <https://doi.org/10.1073/pnas.1802963115>.
- (271) Orellana, L.; Thorne, A. H.; Lema, R.; Gustavsson, J.; Parisian, A. D.; Hospital, A.; Cordeiro, T. N.; Bernadó, P.; Scott, A. M.; Brun-Heath, I.; Lindahl, E.; Cavenee, W. K.; Furnari, F. B.; Orozco, M. Oncogenic Mutations at the EGFR Ectodomain Structurally Converge to Remove a Steric Hindrance on a Kinase-Coupled Cryptic Epitope. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, 116 (20), 10009–10018. <https://doi.org/10.1073/pnas.1821442116>.
- (272) Palermo, G.; Casalino, L.; Magistrato, A.; Andrew McCammon, J. Understanding the Mechanistic Basis of Non-Coding RNA through Molecular Dynamics Simulations. *J. Struct. Biol.* **2019**, 206 (3), 267–279. <https://doi.org/10.1016/j.jsb.2019.03.004>.
- (273) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. Ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99SB. *J. Chem. Theory Comput.* **2015**, 11 (8), 3696–3713. <https://doi.org/10.1021/acs.jctc.5b00255>.
- (274) Galindo-Murillo, R.; Robertson, J. C.; Zgarbová, M.; Šponer, J.; Otyepka, M.; Jurečka, P.; Cheatham, T. E. Assessing the Current State of Amber Force Field Modifications for DNA. *J. Chem. Theory Comput.* **2016**, 12 (8), 4114–4127. <https://doi.org/10.1021/acs.jctc.6b00186>.
- (275) Dans, P. D.; Ivani, I.; Hospital, A.; Portella, G.; González, C.; Orozco, M. How Accurate Are Accurate Force-Fields for B-DNA? *Nucleic Acids Res.* **2017**, 45 (7), 4217–4230. <https://doi.org/10.1093/nar/gkw1355>.

- (276) Singh, U. C.; Kollman, P. A. An Approach to Computing Electrostatic Charges for Molecules. *J. Comput. Chem.* **1984**, *5* (2), 129–145. <https://doi.org/10.1002/jcc.540050204>.
- (277) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. *J. Phys. Chem.* **1993**, *97* (40), 10269–10280. <https://doi.org/10.1021/j100142a004>.
- (278) Hess, B. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4* (1), 116–122. <https://doi.org/10.1021/ct700200b>.
- (279) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindah, E. Gromacs: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25. <https://doi.org/10.1016/j.softx.2015.06.001>.
- (280) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An $N \cdot \log(N)$ Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98* (12), 10089–10092. <https://doi.org/10.1063/1.464397>.
- (281) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103* (19), 8577–8593. <https://doi.org/10.1063/1.470117>.
- (282) Bader, R. F. W. Atoms in Molecules. *Acc. Chem. Res.* **1985**, *18*, 9–15. <https://doi.org/10.1002/prot.10414>.
- (283) Peraro, M. D.; Spiegel, K.; Lamoureux, G.; Vivo, M. De; DeGrado, W. F.; Klein, M. L. Modeling the Charge Distribution at Metal Sites in Proteins for Molecular Dynamics Simulations. *J. Struct. Biol.* **2007**, *157* (3), 444–453. <https://doi.org/10.1016/j.jsb.2006.10.019>.
- (284) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126* (1), 14101. <https://doi.org/10.1063/1.2408420>.
- (285) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52* (12), 7182–7190. <https://doi.org/10.1063/1.328693>.
- (286) Desai, N. A.; Shankar, V. *Single-Strand-Specific Nucleases*. <https://doi.org/10.1111/j.1574-6976.2003.tb00626.x>.
- (287) Zheng, L.; Shen, B. Okazaki Fragment Maturation: Nucleases Take Centre Stage. *J. Mol. Cell Biol.* **2011**, *3* (1), 23–30. <https://doi.org/10.1093/jmcb/mjq048>.
- (288) Ceska, T. Structure-Specific DNA Cleavage by 5' Nucleases. *Trends Biochem. Sci.* **1998**, *23* (9), 331–336. [https://doi.org/10.1016/S0968-0004\(98\)01259-6](https://doi.org/10.1016/S0968-0004(98)01259-6).
- (289) Craggs, T. D.; Sustarsic, M.; Plochowietz, A.; Mosayebi, M.; Kaju, H.; Cuthbert, A.; Hohlbein, J.; Domicевичa, L.; Biggin, P. C.; Doye, J. P. K.; Kapanidis, A. N. Substrate Conformational Dynamics Facilitate Structure-Specific Recognition of Gapped DNA by DNA Polymerase. *Nucleic Acids Res.* **2019**, *47* (20), 10788–10800. <https://doi.org/10.1093/nar/gkz797>.
- (290) Craggs, T. D.; Sustarsic, M.; Plochowietz, A.; Mosayebi, M.; Kaju, H.; Cuthbert, A.; Hohlbein, J.; Domicевичa, L.; Biggin, P. C.; Doye, J. P. K.; Kapanidis, A. N. Substrate Conformational Dynamics Facilitate Structure-Specific Recognition of Gapped DNA by DNA Polymerase. *Nucleic Acids Res.* **2019**, *47* (20), 10788–10800. <https://doi.org/10.1093/nar/gkz797>.

- (291) Falquet, B.; Rass, U. Structure-Specific Endonucleases and the Resolution of Chromosome Underreplication. *Genes (Basel)*. **2019**, *10* (3). <https://doi.org/10.3390/genes10030232>.
- (292) Wu, T.; Zhu, H.; Zhang, M.; Sun, Y.; Yang, Y.; Gu, L.; Zhang, J.; Mu, D.; Wu, C.; Hu, Z.; Jiang, L.; Jia, S.; Zhang, Y.; He, L.; Pan, F. Y.; Guo, Z. FEN1 Inhibitor Synergizes with Low-Dose Camptothecin to Induce Increased Cell Killing via the Mitochondria Mediated Apoptotic Pathway. *Gene Ther.* **2021**, 1–11. <https://doi.org/10.1038/s41434-020-00215-9>.
- (293) Guo, E.; Ishii, Y.; Mueller, J.; Srivatsan, A.; Gahman, T.; Putnam, C. D.; Wang, J. Y. J.; Kolodner, R. D. FEN1 Endonuclease as a Therapeutic Target for Human Cancers with Defects in Homologous Recombination. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117* (32), 19415–19424. <https://doi.org/10.1073/pnas.2009237117>.
- (294) Exell, J. C.; Thompson, M. J.; Finger, L. D.; Shaw, Steven J.; Debreczeni, J.; Ward, T. A.; McWhirter, C.; Siöberg, C. L. B.; Molina, D. M.; Abbott, W. M.; Jones, C. D.; Nissink, J. W. M.; Durant, S. T.; Grasby, J. A. Cellularly Active N-Hydroxyurea FEN1 Inhibitors Block Substrate Entry to the Active Site. *Nature Chemical Biology*. 2016, pp 815–821. <https://doi.org/10.1038/nchembio.2148>.
- (295) Kathera, C.; Zhang, J.; Janardhan, A.; Sun, H.; Ali, W.; Zhou, X.; He, L.; Guo, Z. *Interacting Partners of FEN1 and Its Role in the Development of Anticancer Therapeutics*; 2017; Vol. 8.
- (296) Balakrishnan, L.; Bambara, R. A. Flap Endonuclease 1. *Annu. Rev. Biochem.* **2013**, *82* (1), 119–138. <https://doi.org/10.1146/annurev-biochem-072511-122603>.
- (297) Zaher, M. S.; Rashid, F.; Song, B.; Joudeh, L. I.; Sobhy, M. A.; Tehseen, M.; Hingorani, M. M.; Hamdan, S. M. Missed Cleavage Opportunities by FEN1 Lead to Okazaki Fragment Maturation via the Long-Flap Pathway. *Nucleic Acids Res.* **2018**, *46* (6), 2956–2974. <https://doi.org/10.1093/nar/gky082>.
- (298) Li, X.; Li, J.; Harrington, J.; Lieber, M. R.; Burgers, P. M. J. Lagging Strand DNA Synthesis at the Eukaryotic Replication Fork Involves Binding and Stimulation of FEN-1 by Proliferating Cell Nuclear Antigen. *J. Biol. Chem.* **1995**, *270* (38), 22109–22112. <https://doi.org/10.1074/jbc.270.38.22109>.
- (299) Liu, B.; Hu, J.; Wang, J.; Kong, D. Direct Visualization of RNA-DNA Primer Removal from Okazaki Fragments Provides Support for Flap Cleavage and Exonucleolytic Pathways in Eukaryotic Cells. *J. Biol. Chem.* **2017**, *292* (12), 4777–4788. <https://doi.org/10.1074/jbc.M116.758599>.
- (300) Xu, H.; Shi, R.; Han, W.; Cheng, J.; Xu, X.; Cheng, K.; Wang, L.; Tian, B.; Zheng, L.; Shen, B.; Hua, Y.; Zhao, Y. Structural Basis of 5' Flap Recognition and Protein–Protein Interactions of Human Flap Endonuclease 1. *Nucleic Acids Res.* **2018**, *46* (21), 11315–11325. <https://doi.org/10.1093/nar/gky911>.
- (301) Chapados, B. R.; Hosfield, D. J.; Han, S.; Qiu, J.; Yelent, B.; Shen, B.; Tainer, J. A. Structural Basis for FEN-1 Substrate Specificity and PCNA-Mediated Activation in DNA Replication and Repair. *Cell* **2004**, *116* (1), 39–50. [https://doi.org/10.1016/S0092-8674\(03\)01036-5](https://doi.org/10.1016/S0092-8674(03)01036-5).
- (302) Kovall, R.; Matthews, B. W. Toroidal Structure of λ -Exonuclease. *Science (80-.)*. **1997**, *277* (5333), 1824–1827. <https://doi.org/10.1126/science.277.5333.1824>.
- (303) Debye-Hückel Approximation. In *Encyclopedia of Microfluidics and Nanofluidics*; Li, D., Ed.; Springer US: Boston, MA, 2008; p 334. https://doi.org/10.1007/978-0-387-48998-8_296.

- (304) Tzul, F. O.; Schweiker, K. L.; Makhatadze, G. I. Modulation of Folding Energy Landscape by Charge-Charge Interactions: Linking Experiments with Computational Modeling. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (3), E259–E266. <https://doi.org/10.1073/pnas.1410424112>.
- (305) Mereghetti, P.; Martinez, M.; Wade, R. C. Long Range Debye-Hückel Correction for Computation of Grid-Based Electrostatic Forces between Biomacromolecules. *BMC Biophys.* **2014**, *7* (1), 1–11. <https://doi.org/10.1186/2046-1682-7-4>.
- (306) Vidossich, P.; Castañeda Moreno, L. E.; Mota, C.; De Sanctis, D.; Miscione, G. Pietro; De Vivo, M. Functional Implications of Second-Shell Basic Residues for DUTPase DR2231 Enzymatic Specificity. *ACS Catal.* **2020**, *10* (23), 13825–13833. <https://doi.org/10.1021/acscatal.0c04148>.
- (307) Wu, T.; Yang, Y.; Chen, W.; Wang, J.; Yang, Z.; Wang, S.; Xiao, X.; Li, M.; Zhao, M. Noncanonical Substrate Preference of Lambda Exonuclease for 5'-Nonphosphate-Ended DsDNA and a Mismatch-Induced Acceleration Effect on the Enzymatic Reaction. *Nucleic Acids Res.* **2018**, *46* (6), 3119–3129. <https://doi.org/10.1093/nar/gky154>.
- (308) Cheng, K.; Xu, H.; Chen, X.; Wang, L.; Tian, B.; Zhao, Y.; Hua, Y. Structural Basis for DNA 5'-End Resection by RecJ. *Elife* **2016**, *5*, e14294. <https://doi.org/10.7554/eLife.14294>.
- (309) Cheng, K.; Xu, Y.; Chen, X.; Lu, H.; He, Y.; Wang, L.; Hua, Y. Participation of RecJ in the Base Excision Repair Pathway of Deinococcus Radiodurans. *Nucleic Acids Res.* **2020**, *48* (17), 9859–9871. <https://doi.org/10.1093/nar/gkaa714>.
- (310) Li, M. J.; Yi, G. S.; Yu, F.; Zhou, H.; Chen, J. N.; Xu, C. Y.; Wang, F. P.; Xiao, X.; He, J. H.; Liu, X. P. The Crystal Structure of Pyrococcus Furiosus RecJ Implicates It as an Ancestor of Eukaryotic Cdc45. *Nucleic Acids Res.* **2017**, *45* (21), 12551–12564. <https://doi.org/10.1093/nar/gkx887>.
- (311) Viadiu, H.; Aggarwal, A. K. The Role of Metals in Catalysis by the Restriction Endonuclease BamHI. *Nature Structural Biology*. Nature Publishing Group October 1998, pp 910–916. <https://doi.org/10.1038/2352>.
- (312) 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* (6), 2823–2834. <https://doi.org/10.1021/jacs.9b10656>.
- (313) Lee, B. I.; Wilson, D. M. The RAD2 Domain of Human Exonuclease 1 Exhibits 5' to 3' Exonuclease and Flap Structure-Specific Endonuclease Activities. *J. Biol. Chem.* **1999**, *274* (53), 37763–37769. <https://doi.org/10.1074/jbc.274.53.37763>.
- (314) Zheng, P.; Arantes, G. M.; Field, M. J.; Li, H. Force-Induced Chemical Reactions on the Metal Centre in a Single Metalloprotein Molecule. *Nat. Commun.* **2015**, *6* (1), 1–9. <https://doi.org/10.1038/ncomms8569>.
- (315) Aktah, D.; Frank, I. Breaking Bonds by Mechanical Stress: When Do Electrons Decide for the Other Side? *J. Am. Chem. Soc.* **2002**, *124* (13), 3402–3406. <https://doi.org/10.1021/ja004010b>.
- (316) Vidossich, P.; De Vivo, M. The Role of First Principles Simulations in Studying (Bio)Catalytic Processes. *Chem Catal.* **2021**, *1* (1), 69–87. <https://doi.org/10.1016/j.checat.2021.04.009>.
- (317) Wang, L.; Zhang, M.; Alexov, E. DelPhiPKa Web Server: Predicting PK a of Proteins,

RNAs and DNAs. *Bioinformatics* **2016**, *32* (4), 614–615.
<https://doi.org/10.1093/bioinformatics/btv607>.

- (318) Mark, P.; Nilsson, L. Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *J. Phys. Chem. A* **2001**, *105* (43), 9954–9960. <https://doi.org/10.1021/jp003020w>.
- (319) Manigrasso, J.; Chillón, I.; Genna, V.; Vidossich, P.; Somarowthu, S.; Pyle, A. M.; De Vivo, M.; Marcia, M. Visualizing Group II Intron Dynamics between the First and Second Steps of Splicing. *Nat. Commun.* **2020**, *11* (1), 1–15. <https://doi.org/10.1038/s41467-020-16741-4>.
- (320) Ricci, C. G.; Chen, J. S.; Miao, Y.; Jinek, M.; Doudna, J. A.; McCammon, J. A.; Palermo, G. Deciphering Off-Target Effects in CRISPR-Cas9 through Accelerated Molecular Dynamics. *ACS Cent. Sci.* **2019**, *5* (4), 651–662. <https://doi.org/10.1021/acscentsci.9b00020>.
- (321) Joung, I. S.; Cheatham, T. E. Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem. B* **2008**, *112* (30), 9020–9041. <https://doi.org/10.1021/jp8001614>.
- (322) Li, P.; Roberts, B. P.; Chakravorty, D. K.; Merz, K. M. Rational Design of Particle Mesh Ewald Compatible Lennard-Jones Parameters for +2 Metal Cations in Explicit Solvent. *J. Chem. Theory Comput.* **2013**, *9* (6), 2733–2748. <https://doi.org/10.1021/ct400146w>.
- (323) Díaz Leines, G.; Ensing, B. Path Finding on High-Dimensional Free Energy Landscapes. *Phys. Rev. Lett.* **2012**, *109* (2). <https://doi.org/10.1103/PhysRevLett.109.020601>.
- (324) Phillips, N. The Coronavirus Is Here to Stay - Here's What That Means. *Nature* **2021**, *590* (7846), 382–384. <https://doi.org/10.1038/d41586-021-00396-2>.
- (325) Wu, F.; Zhao, S.; Yu, B.; Chen, Y. M.; Wang, W.; Song, Z. G.; Hu, Y.; Tao, Z. W.; Tian, J. H.; Pei, Y. Y.; Yuan, M. L.; Zhang, Y. L.; Dai, F. H.; Liu, Y.; Wang, Q. M.; Zheng, J. J.; Xu, L.; Holmes, E. C.; Zhang, Y. Z. A New Coronavirus Associated with Human Respiratory Disease in China. *Nature* **2020**, *579* (7798), 265–269. <https://doi.org/10.1038/s41586-020-2008-3>.
- (326) Wang, Q.; Wu, J.; Wang, H.; Gao, Y.; Liu, Q.; Mu, A.; Ji, W.; Yan, L.; Zhu, Y.; Zhu, C.; Fang, X.; Yang, X.; Huang, Y.; Gao, H.; Liu, F.; Ge, J.; Sun, Q.; Yang, X.; Xu, W.; Liu, Z.; Yang, H.; Lou, Z.; Jiang, B.; Guddat, L. W.; Gong, P.; Rao, Z. Structural Basis for RNA Replication by the SARS-CoV-2 Polymerase. *Cell* **2020**, *182* (2), 417-428.e13. <https://doi.org/10.1016/j.cell.2020.05.034>.
- (327) Huang, J.; Song, W.; Huang, H.; Sun, Q. Pharmacological Therapeutics Targeting RNA-Dependent RNA Polymerase, Proteinase and Spike Protein: From Mechanistic Studies to Clinical Trials for COVID-19. *J. Clin. Med.* **2020**, *9* (4), 1131. <https://doi.org/10.3390/jcm9041131>.
- (328) Casalino, L.; Dommer, A. C.; Gaieb, Z.; Barros, E. P.; Sztain, T.; Ahn, S. H.; Trifan, A.; Brace, A.; Bogetti, A. T.; Clyde, A.; Ma, H.; Lee, H.; Turilli, M.; Khalid, S.; Chong, L. T.; Simmerling, C.; Hardy, D. J.; Maia, J. D. C.; Phillips, J. C.; Kurth, T.; Stern, A. C.; Huang, L.; McCalpin, J. D.; Tatineni, M.; Gibbs, T.; Stone, J. E.; Jha, S.; Ramanathan, A.; Amaro, R. E. AI-Driven Multiscale Simulations Illuminate Mechanisms of SARS-CoV-2 Spike Dynamics. *Int. J. High Perform. Comput. Appl.* **2021**, *35* (5), 432–451. <https://doi.org/10.1177/10943420211006452>.
- (329) Shang, J.; Ye, G.; Shi, K.; Wan, Y.; Luo, C.; Aihara, H.; Geng, Q.; Auerbach, A.; Li, F. Structural Basis of Receptor Recognition by SARS-CoV-2. *Nature* **2020**, *581* (7807), 221–224. <https://doi.org/10.1038/s41586-020-2179-y>.

- (330) Zhu, W.; Chen, C. Z.; Gorshkov, K.; Xu, M.; Lo, D. C.; Zheng, W. RNA-Dependent RNA Polymerase as a Target for COVID-19 Drug Discovery. *SLAS Discov.* **2020**, *25* (10), 1141–1151. <https://doi.org/10.1177/2472555220942123>.
- (331) Picarazzi, F.; Vicenti, I.; Saladini, F.; Zazzi, M.; Mori, M. Targeting the RdRp of Emerging RNA Viruses: The Structure-Based Drug Design Challenge. *Mol. 2020, Vol. 25, Page 5695* **2020**, *25* (23), 5695. <https://doi.org/10.3390/MOLECULES25235695>.
- (332) Tian, L.; Qiang, T.; Liang, C.; Ren, X.; Jia, M.; Zhang, J.; Li, J.; Wan, M.; YuWen, X.; Li, H.; Cao, W.; Liu, H. RNA-Dependent RNA Polymerase (RdRp) Inhibitors: The Current Landscape and Repurposing for the COVID-19 Pandemic. *Eur. J. Med. Chem.* **2021**, *213*, 113201. <https://doi.org/10.1016/j.ejmech.2021.113201>.
- (333) Chien, M.; Anderson, T. K.; Jockusch, S.; Tao, C.; Li, X.; Kumar, S.; Russo, J. J.; Kirchdoerfer, R. N.; Ju, J. Nucleotide Analogues as Inhibitors of SARS-CoV-2 Polymerase, a Key Drug Target for COVID-19. *J. Proteome Res.* **2020**, *19* (11), 4690–4697. <https://doi.org/10.1021/acs.jproteome.0c00392>.
- (334) Jockusch, S.; Tao, C.; Li, X.; Anderson, T. K.; Chien, M.; Kumar, S.; Russo, J. J.; Kirchdoerfer, R. N.; Ju, J. A Library of Nucleotide Analogues Terminate RNA Synthesis Catalyzed by Polymerases of Coronaviruses That Cause SARS and COVID-19. *Antiviral Res.* **2020**, *180*, 104857. <https://doi.org/10.1016/j.antiviral.2020.104857>.
- (335) Khan, S.; Attar, F.; Bloukh, S. H.; Sharifi, M.; Nabi, F.; Bai, Q.; Khan, R. H.; Falahati, M. A Review on the Interaction of Nucleoside Analogues with SARS-CoV-2 RNA Dependent RNA Polymerase. *Int. J. Biol. Macromol.* **2021**, *181*, 605–611. <https://doi.org/10.1016/J.IJBIOMAC.2021.03.112>.
- (336) Li, G.; De Clercq, E. Therapeutic Options for the 2019 Novel Coronavirus (2019-NCoV). *Nat. Rev. Drug Discov.* **2021**, *193* **2020**, *19* (3), 149–150. <https://doi.org/10.1038/d41573-020-00016-0>.
- (337) Wang, M.; Cao, R.; Zhang, L.; Yang, X.; Liu, J.; Xu, M.; Shi, Z.; Hu, Z.; Zhong, W.; Xiao, G. Remdesivir and Chloroquine Effectively Inhibit the Recently Emerged Novel Coronavirus (2019-NCoV) in Vitro. *Cell Res.* **2020**, *303* **2020**, *30* (3), 269–271. <https://doi.org/10.1038/s41422-020-0282-0>.
- (338) Wang, Y.; Zhang, D.; Du, G.; Du, R.; Zhao, J.; Jin, Y.; Fu, S.; Gao, L.; Cheng, Z.; Lu, Q.; Hu, Y.; Luo, G.; Wang, K.; Lu, Y.; Li, H.; Wang, S.; Ruan, S.; Yang, C.; Mei, C.; Wang, Y.; Ding, D.; Wu, F.; Tang, X.; Ye, X.; Ye, Y.; Liu, B.; Yang, J.; Yin, W.; Wang, A.; Fan, G.; Zhou, F.; Liu, Z.; Gu, X.; Xu, J.; Shang, L.; Zhang, Y.; Cao, L.; Guo, T.; Wan, Y.; Qin, H.; Jiang, Y.; Jaki, T.; Hayden, F. G.; Horby, P. W.; Cao, B.; Wang, C. Remdesivir in Adults with Severe COVID-19: A Randomised, Double-Blind, Placebo-Controlled, Multicentre Trial. *Lancet* **2020**, *395* (10236), 1569–1578. [https://doi.org/10.1016/S0140-6736\(20\)31022-9](https://doi.org/10.1016/S0140-6736(20)31022-9).
- (339) Malone, B.; Campbell, E. A. Molnupiravir: Coding for Catastrophe. *Nat. Struct. Mol. Biol.* **2021**, *289* **2021**, *28* (9), 706–708. <https://doi.org/10.1038/s41594-021-00657-8>.
- (340) Carta, A.; Briguglio, I.; Piras, S.; Corona, P.; Boatto, G.; Nieddu, M.; Giunchedi, P.; Marongiu, M. E.; Giliberti, G.; Iuliano, F.; Blois, S.; Ibba, C.; Busonera, B.; La Colla, P. Quinoline Tricyclic Derivatives. Design, Synthesis and Evaluation of the Antiviral Activity of Three New Classes of RNA-Dependent RNA Polymerase Inhibitors. *Bioorganic Med. Chem.* **2011**, *19* (23), 7070–7084. <https://doi.org/10.1016/j.bmc.2011.10.009>.

- (341) Zhao, J.; Zhang, Y.; Wang, M.; Liu, Q.; Lei, X.; Wu, M.; Guo, S.; Yi, D.; Li, Q.; Ma, L.; Liu, Z.; Guo, F.; Wang, J.; Li, X.; Wang, Y.; Cen, S. Quinoline and Quinazoline Derivatives Inhibit Viral RNA Synthesis by SARS-CoV-2 RdRp. *ACS Infect. Dis.* **2021**, *7* (6), 1535–1544. <https://doi.org/10.1021/acsinfecdis.1c00083>.
- (342) Newman, D. J.; Cragg, G. M. Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. *J. Nat. Prod.* **2020**, *83* (3), 770–803. https://doi.org/10.1021/ACS.JNATPROD.9B01285/SUPPL_FILE/NP9B01285_SI_009.PDF.
- (343) Steinmann, J.; Buer, J.; Pietschmann, T.; Steinmann, E. Anti-Infective Properties of Epigallocatechin-3-Gallate (EGCG), a Component of Green Tea. *Br. J. Pharmacol.* **2013**, *168* (5), 1059–1073. <https://doi.org/10.1111/BPH.12009>.
- (344) Saeedi-Boroujeni, A.; Mahmoudian-Sani, M. R. Anti-Inflammatory Potential of Quercetin in COVID-19 Treatment. *J. Inflamm. (United Kingdom)* **2021**, *18* (1), 1–9. <https://doi.org/10.1186/s12950-021-00268-6>.
- (345) Liskova, A.; Samec, M.; Koklesova, L.; Samuel, S. M.; Zhai, K.; Al-Ishaq, R. K.; Abotaleb, M.; Nosal, V.; Kajo, K.; Ashrafizadeh, M.; Zarrabi, A.; Brockmueller, A.; Shakibaei, M.; Sabaka, P.; Mozos, I.; Ullrich, D.; Prosecky, R.; La Rocca, G.; Caprnda, M.; Büsselberg, D.; Rodrigo, L.; Kruzliak, P.; Kubatka, P. Flavonoids against the SARS-CoV-2 Induced Inflammatory Storm. *Biomed. Pharmacother.* **2021**, *138*, 111430. <https://doi.org/10.1016/J.BIOPHA.2021.111430>.
- (346) Derosa, G.; Maffioli, P.; D'Angelo, A.; Di Pierro, F. A Role for Quercetin in Coronavirus Disease 2019 (COVID-19). *Phyther. Res.* **2021**, *35* (3), 1230–1236. <https://doi.org/10.1002/PTR.6887>.
- (347) Goris, T.; Pérez-Valero, Á.; Martínez, I.; Yi, D.; Fernández-Calleja, L.; San León, D.; Bornscheuer, U. T.; Magadán-Corpas, P.; Lombó, F.; Nogales, J. Repositioning Microbial Biotechnology against COVID-19: The Case of Microbial Production of Flavonoids. *Microb. Biotechnol.* **2021**, *14* (1), 94–110. <https://doi.org/10.1111/1751-7915.13675>.
- (348) Russo, M.; Moccia, S.; Spagnuolo, C.; Tedesco, I.; Russo, G. L. Roles of Flavonoids against Coronavirus Infection. *Chem. Biol. Interact.* **2020**, *328*, 109211. <https://doi.org/10.1016/J.CBI.2020.109211>.
- (349) Abian, O.; Ortega-Alarcon, D.; Jimenez-Alesanco, A.; Ceballos-Laita, L.; Vega, S.; Reyburn, H. T.; Rizzuti, B.; Velazquez-Campoy, A. Structural Stability of SARS-CoV-2 3CLpro and Identification of Quercetin as an Inhibitor by Experimental Screening. *Int. J. Biol. Macromol.* **2020**, *164*, 1693–1703. <https://doi.org/10.1016/j.ijbiomac.2020.07.235>.
- (350) Choudhry, N.; Zhao, X.; Xu, D.; Zanin, M.; Chen, W.; Yang, Z.; Chen, J. Chinese Therapeutic Strategy for Fighting COVID-19 and Potential Small-Molecule Inhibitors against Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). *J. Med. Chem.* **2020**, *63* (22), 13205–13227. <https://doi.org/10.1021/acs.jmedchem.0c00626>.
- (351) Lee, C.; Lee, J. M.; Lee, N. R.; Kim, D. E.; Jeong, Y. J.; Chong, Y. Investigation of the Pharmacophore Space of Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) NTPase/Helicase by Dihydroxychromone Derivatives. *Bioorganic Med. Chem. Lett.* **2009**, *19* (16), 4538–4541. <https://doi.org/10.1016/j.bmcl.2009.07.009>.
- (352) Nguyen, T. T. H.; Woo, H. J.; Kang, H. K.; Nguyen, V. D.; Kim, Y. M.; Kim, D. W.; Ahn, S. A.; Xia, Y.; Kim, D. Flavonoid-Mediated Inhibition of SARS Coronavirus 3C-like Protease

Expressed in *Pichia Pastoris*. *Biotechnol. Lett.* **2012**, *34* (5), 831–838.
<https://doi.org/10.1007/s10529-011-0845-8>.

- (353) Shaldam, M. A.; Yahya, G.; Mohamed, N. H.; Abdel-Daim, M. M.; Al Naggar, Y. In Silico Screening of Potent Bioactive Compounds from Honeybee Products against COVID-19 Target Enzymes. *Environ. Sci. Pollut. Res.* **2021**, *28* (30), 40507–40514.
<https://doi.org/10.1007/s11356-021-14195-9>.
- (354) Yin, W.; Luan, X.; Li, Z.; Zhou, Z.; Wang, Q.; Gao, M.; Wang, X.; Zhou, F.; Shi, J.; You, E.; Liu, M.; Wang, Q.; Jiang, Y.; Jiang, H.; Xiao, G.; Zhang, L.; Yu, X.; Zhang, S.; Eric Xu, H. Structural Basis for Inhibition of the SARS-CoV-2 RNA Polymerase by Suramin. *Nat. Struct. Mol. Biol.* **2021**, *28* (3), 319–325. <https://doi.org/10.1038/s41594-021-00570-0>.
- (355) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J. Med. Chem.* **2004**, *47* (7), 1750–1759.
<https://doi.org/10.1021/jm030644s>.
- (356) Ortega, J. A.; Arencibia, J. M.; Minniti, E.; Byl, J. A. W.; Franco-Ulloa, S.; Borgogno, M.; Genna, V.; Summa, M.; Bertozzi, S. M.; Bertorelli, R.; Armirotti, A.; Minarini, A.; Sissi, C.; Osheroff, N.; De Vivo, M. Novel, Potent, and Druglike Tetrahydroquinazoline Inhibitor That Is Highly Selective for Human Topoisomerase II α over β . *J. Med. Chem.* **2020**, *63* (21), 12873–12886. <https://doi.org/10.1021/acs.jmedchem.0c00774>.
- (357) Arencibia, J. M.; Brindani, N.; Franco-Ulloa, S.; Nigro, M.; Kuriappan, J. A.; Ottonello, G.; Bertozzi, S. M.; Summa, M.; Girotto, S.; Bertorelli, R.; Armirotti, A.; De Vivo, M. Design, Synthesis, Dynamic Docking, Biochemical Characterization, and in Vivo Pharmacokinetics Studies of Novel Topoisomerase II Poisons with Promising Antiproliferative Activity. *J. Med. Chem.* **2020**, *63* (7), 3508–3521. <https://doi.org/10.1021/acs.jmedchem.9b01760>.
- (358) De Vivo, M.; Masetti, M.; Bottegoni, G.; Cavalli, A. Role of Molecular Dynamics and Related Methods in Drug Discovery. *J. Med. Chem.* **2016**, *59* (9), 4035–4061.
<https://doi.org/10.1021/ACS.JMEDCHEM.5B01684>.
- (359) Palermo, G.; Minniti, E.; Greco, M. L.; Riccardi, L.; Simoni, E.; Convertino, M.; Marchetti, C.; Rosini, M.; Sissi, C.; Minarini, A.; De Vivo, M. An Optimized Polyamine Moiety Boosts the Potency of Human Type II Topoisomerase Poisons as Quantified by Comparative Analysis Centered on the Clinical Candidate F14512. *Chem. Commun.* **2015**, *51* (76), 14310–14313. <https://doi.org/10.1039/c5cc05065k>.
- (360) Palermo, G.; Favia, A. D.; Convertino, M.; De Vivo, M. The Molecular Basis for Dual Fatty Acid Amide Hydrolase (FAAH)/Cyclooxygenase (COX) Inhibition. *ChemMedChem* **2016**, *11* (12), 1252–1258. <https://doi.org/10.1002/cmdc.201500507>.
- (361) Yan, L.; Ge, J.; Zheng, L.; Zhang, Y.; Gao, Y.; Wang, T.; Huang, Y.; Yang, Y.; Gao, S.; Li, M.; Liu, Z.; Wang, H.; Li, Y.; Chen, Y.; Guddat, L. W.; Wang, Q.; Rao, Z.; Lou, Z. Cryo-EM Structure of an Extended SARS-CoV-2 Replication and Transcription Complex Reveals an Intermediate State in Cap Synthesis. *Cell* **2021**, *184* (1), 184–193.e10.
<https://doi.org/10.1016/j.cell.2020.11.016>.
- (362) Yan, L.; Zhang, Y.; Ge, J.; Zheng, L.; Gao, Y.; Wang, T.; Jia, Z.; Wang, H.; Huang, Y.; Li, M.; Wang, Q.; Rao, Z.; Lou, Z. Architecture of a SARS-CoV-2 Mini Replication and Transcription Complex. *Nat. Commun.* **2020**, *11* (1), 1–6. <https://doi.org/10.1038/s41467-020-19770-1>.

- (363) Hillen, H. S.; Kokic, G.; Farnung, L.; Dienemann, C.; Tegunov, D.; Cramer, P. Structure of Replicating SARS-CoV-2 Polymerase. *Nature* **2020**, *584* (7819), 154–156. <https://doi.org/10.1038/s41586-020-2368-8>.
- (364) Naydenova, K.; Muir, K. W.; Wu, L. F.; Zhang, Z.; Coscia, F.; Peet, M. J.; Castro-Hartmann, P.; Qian, P.; Sader, K.; Dent, K.; Kimanius, D.; Sutherland, J. D.; Löwe, J.; Barford, D.; Russo, C. J. Structure of the SARS-CoV-2 RNA-Dependent RNA Polymerase in the Presence of Favipiravir-RTP. *Proc. Natl. Acad. Sci. U. S. A.* **2021**, *118* (7). <https://doi.org/10.1073/pnas.2021946118>.
- (365) Yin, W.; Mao, C.; Luan, X.; Shen, D. D.; Shen, Q.; Su, H.; Wang, X.; Zhou, F.; Zhao, W.; Gao, M.; Chang, S.; Xie, Y. C.; Tian, G.; Jiang, H. W.; Tao, S. C.; Shen, J.; Jiang, Y.; Jiang, H.; Xu, Y.; Zhang, S.; Zhang, Y.; Xu, H. E. Structural Basis for Inhibition of the RNA-Dependent RNA Polymerase from SARS-CoV-2 by Remdesivir. *Science* (80-.). **2020**, *368* (6498), 1499–1504. <https://doi.org/10.1126/science.abc1560>.
- (366) Chen, J.; Malone, B.; Llewellyn, E.; Grasso, M.; Shelton, P. M. M.; Olinares, P. D. B.; Maruthi, K.; Eng, E. T.; Vatandaslar, H.; Chait, B. T.; Kapoor, T. M.; Darst, S. A.; Campbell, E. A. Structural Basis for Helicase-Polymerase Coupling in the SARS-CoV-2 Replication-Transcription Complex. *Cell* **2020**, *182* (6), 1560-1573.e13. <https://doi.org/10.1016/j.cell.2020.07.033>.
- (367) Kokic, G.; Hillen, H. S.; Tegunov, D.; Dienemann, C.; Seitz, F.; Schmitzova, J.; Farnung, L.; Siewert, A.; Höbartner, C.; Cramer, P. Mechanism of SARS-CoV-2 Polymerase Stalling by Remdesivir. *Nat. Commun.* **2021**, *12* (1), 1–7. <https://doi.org/10.1038/s41467-020-20542-0>.
- (368) Mishra, A.; Rathore, A. S. RNA Dependent RNA Polymerase (RdRp) as a Drug Target for SARS-CoV2. *J. Biomol. Struct. Dyn.* **2021**. <https://doi.org/10.1080/07391102.2021.1875886>.
- (369) Deshmukh, M. G.; Ippolito, J. A.; Zhang, C. H.; Stone, E. A.; Reilly, R. A.; Miller, S. J.; Jorgensen, W. L.; Anderson, K. S. Structure-Guided Design of a Peramppanel-Derived Pharmacophore Targeting the SARS-CoV-2 Main Protease. *Structure* **2021**, *29* (8), 823-833.e5. <https://doi.org/10.1016/j.str.2021.06.002>.
- (370) De Vivo, M.; Cavalli, A. Recent Advances in Dynamic Docking for Drug Discovery. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2017**, *7* (6), e1320. <https://doi.org/10.1002/wcms.1320>.
- (371) Ghahremanpour, M. M.; Tirado-Rives, J.; Deshmukh, M.; Ippolito, J. A.; Zhang, C. H.; Cabeza De Vaca, I.; Liosi, M. E.; Anderson, K. S.; Jorgensen, W. L. Identification of 14 Known Drugs as Inhibitors of the Main Protease of SARS-CoV-2. *ACS Med. Chem. Lett.* **2020**, *11* (12), 2526–2533. <https://doi.org/10.1021/acsmchemlett.0c00521>.
- (372) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004**, *47* (7), 1739–1749. <https://doi.org/10.1021/jm0306430>.
- (373) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. Ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99SB. *J. Chem. Theory Comput.* **2015**, *11* (8), 3696–3713. <https://doi.org/10.1021/acs.jctc.5b00255>.
- (374) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25* (9), 1157–1174.

<https://doi.org/10.1002/jcc.20035>.

- (375) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. *J. Phys. Chem.* **1993**, *97* (40), 10269–10280. <https://doi.org/10.1021/j100142a004>.
- (376) Joung, I. S.; Cheatham, T. E. Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem. B* **2008**, *112* (30), 9020–9041. <https://doi.org/10.1021/jp8001614>.
- (377) Hess, B. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4* (1), 116–122. <https://doi.org/10.1021/ct700200b>.
- (378) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126* (1), 014101. <https://doi.org/10.1063/1.2408420>.
- (379) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52* (12), 7182–7190. <https://doi.org/10.1063/1.328693>.
- (380) Piana, S.; Lindorff-Larsen, K.; Dirks, R. M.; Salmon, J. K.; Dror, R. O.; Shaw, D. E. Evaluating the Effects of Cutoffs and Treatment of Long-Range Electrostatics in Protein Folding Simulations. *PLoS One* **2012**, *7* (6), e39918. <https://doi.org/10.1371/journal.pone.0039918>.
- (381) Laidler, K. J.; King, M. C. The Development of Transition-State Theory. *J. Phys. Chem.* **1983**, *87* (15), 2657–2664. <https://doi.org/10.1021/j100238a002>.