

# UC Irvine

## UC Irvine Previously Published Works

### Title

Trapping a non-cognate nucleotide upon initial binding for replication fidelity control in SARS-CoV-2 RNA dependent RNA polymerase

### Permalink

<https://escholarship.org/uc/item/3q713774>

### Journal

Physical Chemistry Chemical Physics, 26(3)

### ISSN

0956-5000

### Authors

Romero, Moises E

McElhenney, Shannon J

Yu, Jin

### Publication Date

2024-01-17

### DOI

10.1039/d3cp04410f

### Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial-ShareAlike License, available at

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

Peer reviewed

# PCCP

Physical Chemistry Chemical Physics

Accepted Manuscript

This article can be cited before page numbers have been issued, to do this please use: M. E. Romero, S. J. McElhenney and J. Yu, *Phys. Chem. Chem. Phys.*, 2024, DOI: 10.1039/D3CP04410F.



This is an Accepted Manuscript, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this Accepted Manuscript with the edited and formatted Advance Article as soon as it is available.

You can find more information about Accepted Manuscripts in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this Accepted Manuscript or any consequences arising from the use of any information it contains.

Cite this: DOI: 00.0000/xxxxxxxxxx

Received Date  
Accepted Date

DOI: 00.0000/xxxxxxxxxx

# Trapping non-cognate nucleotide upon initial binding for replication fidelity control in SARS-CoV-2 RNA dependent RNA polymerase<sup>†</sup>

Moises E. Romero,<sup>a,b</sup> Shannon J. McElhenney,<sup>a,b</sup> and Jin Yu<sup>c</sup>

The RNA dependent RNA polymerase (RdRp) in SARS-CoV-2 is a highly conserved enzyme responsible for viral genome replication/transcription. To understand how the viral RdRp achieves fidelity control during such processes, here we investigate computationally natural non-cognate vs cognate nucleotide addition and selectivity during the viral RdRp elongation. We focus on nucleotide substrate initial binding (RdRp active site open) to insertion prechemically (active site closed) of the RdRp. Current studies have been first carried out using microsecond ensemble equilibrium all-atom molecular dynamics (MD) simulations. Due to slow conformational changes (from the open to closed) during nucleotide insertion and selection, enhanced or umbrella sampling methods have been further employed to calculate free energy profiles of the nucleotide insertion. Our studies find notable stability of noncognate dATP and GTP upon initial binding in the active-site open state. The results indicate that while natural cognate ATP and Remdesivir drug analogue (RDV-TP) are biased to be stabilized in the closed state to facilitate insertion, the natural non-cognate dATP and GTP can be well trapped in off-path initial binding configurations to be prevented from insertion or to be further rejected. Current work thus presents intrinsic nucleotide selectivity of SARS-CoV-2 RdRp for natural substrate fidelity control, which can be considered for antiviral drug design to overcome.

## 1 Introduction

The SARS-CoV-2 virus responsible for the COVID-19 pandemic continues to evolve<sup>1</sup> and pose a threat to human life<sup>2</sup>. While much of the drug development is besides successful vaccines has been focused on targeting the viral spike protein<sup>3–6</sup> or the main protease<sup>7–10</sup>, there are significant challenges remaining. The spike protein is known for its high variability<sup>1,11,12</sup> and the protease<sup>13</sup> can also mutate, becoming resistant to drugs, as seen in HCV<sup>14,15</sup>. In contrast, the core replication machinery of SARS-CoV-2, the RNA dependent RNA polymerase (RdRp) or nonstructural protein nsp12, is a highly conserved drug target<sup>16,17</sup>. Here we focus on studying SARS-CoV-2 RdRp to understand its elongation and fidelity control mechanism incorporating nucleotide substrates during viral replication for future antiviral drug development. Although nucleotide analogue compounds or inhibitors have been widely implemented for targeting on viral RdRps, the

underlying mechanism remain vague<sup>18–22</sup>. It is expected that successful nucleotide analogue drugs would overcome the RdRp fidelity control so that to inhibit the viral replication.

Upon the pandemic upheaval in 2020, a few high-resolution cryo-EM structures of SARS-CoV-2 RdRp were released immediately, including an apo form with the RdRp active site open<sup>23,24</sup> and a post-catalysis form with the active site closed<sup>24</sup>. The post-catalysis structure was bound with a nucleotide drug analogue from remdesivir (RDV)<sup>24</sup>. These structures were in complex with segments of the nsp7/nsp8 cofactors. Later, additional high-resolution structures were resolved with longer nsp8 being “sliding poles”<sup>25</sup>, as well as RdRp in conjunction with the nsp13 helicase enzyme<sup>26</sup>, and in both pre- and post-translocation states<sup>27</sup>. Further structures also illustrated RdRp backtracking<sup>28</sup> or stalling<sup>29</sup> due to the interaction with the drug analogue RDV. Similarly, there were structures obtained with favipiravir, another nucleotide analogue drug<sup>30,31</sup>. Overall, these structures adopt the post-catalysis state<sup>24</sup> in the nucleotide addition cycle

<sup>a</sup> Equal contribution.<sup>b</sup> Department of Chemistry, University of California, Irvine, CA 92697, USA.<sup>c</sup> Department of Physics and Astronomy, Department of Chemistry, NSF-Simmons Center for Multi-scale Cell Fate Research, University of California, Irvine, CA 92697, USA. E-mail: jin.yu@uci.edu<sup>†</sup> Electronic Supplementary Information (ESI) available: Supplementary figures and tables PDF included with submission.

(NAC) of the viral RdRp, leaving the initial nucleotide binding (active site open) and pre-catalytic insertion or substrate (closed) states unresolved. It was until very recently that the insertion state with ATP as a cognate nucleoside triphosphate (NTP) bound in the active site was resolved<sup>32</sup>. Currently, the initial binding or open state structure of RdRp remains unresolved, although such a structure has been identified for the Poliovirus<sup>33</sup> (PV) or Enterovirus<sup>34</sup> (EV) RdRp, which share structural similarities with SARS-CoV-2 RdRp.

Efforts to identify drug inhibitors for the SARS-CoV-2 RdRp have been made extensively<sup>18,35,36</sup>. Computational docking has often been employed, which predominantly focus on nucleotide immediate binding to an apo-form RdRp structure<sup>37,38</sup>, non-differentiating the active open or closed form, and initially largely overlooking the functional RdRp (nsp12) elongation complex composed additionally of nsp8, nsp7, and RNA strands<sup>39</sup>. Moreover, while atomic molecular dynamics (MD) studies have provided insights into interactions of nucleotide analogues with the viral RdRp, they often utilize directly the insertion state<sup>40–42</sup>, ignoring the initial nucleotide analogue binding stage that can be essential for nucleotide screening or selectivity upon entry. Meanwhile, single-molecule studies have offered a glimpse into the dynamics of the elongation cycle, revealing that RdRp can adopt fast, slow, or very slow catalytic pathways with variable rates contingent upon the kinetic pathway<sup>43</sup>. Additionally, information in regard to the RdRp translocation in the NAC has advanced through the cryo-EM studies, which demonstrated a structural rearrangement in nsp8 to accommodate the exiting RNA duplex<sup>27</sup>. Computational work has shown that the incorporation of RDV-TP into the primer strand results in a steric clash at the conserved motif-B of the RdRp leading to an unstable post-translocation state in comparison with pre-translocation, i.e., as a mechanism for antiviral analogue termination of elongation<sup>44</sup>. An alternative suggestion based on single-molecule studies<sup>43</sup> proposes that RdRp backtracks up to 30 nucleotides (nts) after RDV-TP incorporation, which can be interpreted as elongation termination in standard assays. Despite these efforts on quantitative studies of the RdRp NAC, a critical gap remains in understanding initial nucleotide substrate binding to insertion, which is fundamental for nucleotide selectivity and antiviral drug design, given the substrate screening and pre-chemistry inhibition as essential fidelity checkpoints in stepwise NAC<sup>45–47</sup>.

Accompanied with the nucleotide substrate binding to insertion, a substantial protein conformational transition or active-site re-arrangement happens, which is likely to be a rate limiting step in the NAC, as demonstrated in structurally similar single-subunit RNA or DNA polymerases (RNAPs or DNAPs)<sup>48–50</sup>. Such a rate-limiting pre-chemical step accordingly plays a highly essential role in the nucleotide substrate selectivity<sup>51</sup>. To quantify the process with energetics, we calculated the free energy profile or the potential of mean force (PMF) of the nucleotide insertion, recently for cognate substrate ATP and the corresponding nucleotide drug analogue RDV-TP<sup>52</sup>, and calculated the insertion PMFs of dATP and GTP. We found that both ATP and RDV-TP become significantly more stabilized in the insertion state than upon initial binding. In contrast to natural substrate ATP, RDV-TP

behaves differently in its interaction with the template nt. Cognate ATP forms the Watson-Crick (WC) base pairing with template uracil at the +1 position, from the initial binding to insertion. On the other hand, RDV-TP initially forms base stacking with the template nt at +1 upon binding. It then inserts into the active site, facing an energetic barrier that is marginally low or comparable to that of cognate ATP<sup>52</sup>.

Note that RDV-TP analogue differs from the cognate ATP by only a few atoms: with a 1' cyano group attached on the sugar C1' and 3 atomic replacements on the base. Interestingly, we have also noticed that one conserved motif-F essentially facilitates the insertion of cognate ATP via interactions with the ATP-triphosphate tail. While such phosphate interactions could potentially hinder the drug analogue RDV-TP insertion, it was subtly avoided upon sufficient thermal fluctuations from the template nt+1 that forms base stacking with RDV-TP<sup>52</sup>.

Also note that the intrinsic nucleotide selectivity serves for fidelity control in general for viral RdRp transcription/replication as well as in other RNA/DNA polymerase (RNAP/DNAP) systems<sup>20</sup>. The equilibrium free energetic disparity between Watson-Crick (WC) template-cognate NTP base pairing vs non-WC pairing (e.g. 2-3 kcal/mol) cannot explain the error rate down to 10<sup>-4</sup> to 10<sup>-6</sup> or lower. Indeed, such fidelity control is achieved via non-equilibrium elongation process dictated by the polymerase enzyme. The fundamental issues were initially addressed via kinetic proofreading mechanisms by Hopfield<sup>53</sup> and Ninio<sup>54</sup>. A recent kinetic framework addressing the stepwise nucleotide selectivity of RNAP via multiple kinetic checkpoints can be found from our previous works<sup>46,51</sup>.

In current work, we focus on characterizing the intrinsic nucleotide polymerase enzyme selectivity of SARS-CoV-2 RdRp, i.e., the selectivity or differentiation between natural cognate NTP (ATP here) and natural non-cognate NTP substrates. To do that, we examined the nucleotide insertion dynamics of non-cognate dATP and GTP, in comparison with cognate ATP and RDV-TP analogue, and calculated the insertion PMFs of dATP and GTP starting from initial binding stage. Since the polymerase NAC lasts over tens of milliseconds in general<sup>43</sup>, the rate-limiting transition accompanies the nucleotide binding to insertion (or active site from open to closed) of RdRp is expected to be on the millisecond timescale<sup>48–50</sup>. Therefore, such an insertion process cannot be sampled directly using equilibrium MD simulation that is limited by the sub-microsecond to microseconds timescale<sup>55,56</sup>. To obtain free energetics of such a dynamics process, we extended our previous methodology on employing umbrella sampling MD simulations to construct the PMFs of various NTPs from initial binding to insertion<sup>52,57</sup>. In such an implementation, harmonic forces are added to certain molecular configurations to ensure that high-energy or non-favored configurations are well sampled and the free energy profiles can be obtained after reweighting<sup>58,59</sup>. We first constructed atomic structural models of the RdRp-nsp7-nsp8-RNA complexes bound with the noncognate nucleotide (ncNTP) species, in both initial binding and insertion states. Subsequently, we performed all-atom equilibrium simulations at sub-microseconds in ensembles to characterize the respective initial binding and insertion states of the non-cognate

dATP/GTP bound RdRp complexes. Lastly, we obtained the PMFs of the dATP/GTP insertion processes using the umbrella sampling MD simulations, following initial insertion paths constructed on top of collective reaction coordinates (RCs), according to displacements of atomic coordinates from seven highly conserved structural motifs of RdRp (A to G) and incoming NTP (with and without the associating template nt). Our aim is to elucidate the intrinsic nucleotide selectivity of SARS-CoV-2 RdRp, which turns out to be primarily relying on 'trapping' the non-cognate nucleotide species upon entry or initial binding to certain configurations at the active site and preventing them from further insertion. It is expected that newly designed nucleotide drug compounds would be able to avoid such selectivity in order to be incorporated into the viral RNA chain for inhibition.

## 2 Methods

### 2.1 Modeling Details

#### 2.1.1 Constructing Open and Closed models for dATP and GTP

Based on the approach used in previous work on modeling SARS-CoV-2 RdRp<sup>52</sup>, we constructed models for initial binding (open) and insertion (closed) of various NTPs using high-resolution Cryo-EM structures of SARS-CoV-2 nsp12-nsp7-ns8-RNA complexes. The model for RDV-TP in the insertion state was created using a post-catalytic structure bound with RDV-TP (PDBid:7BV2) as a reference, which also includes catalytic Mg<sup>2+</sup> ions<sup>24</sup>. For the ATP insertion model, we positioned ATP over the RDV-TP in the insertion model. The RDV-TP initial binding model was created using an apo structure (PDBid:7BTF)<sup>23</sup> aligned with the RDV-TP insertion structure, with RNA, RDV-TP, and Mg<sup>2+</sup> ions copied over. The ATP initial binding model was built following the same process as described above. The dATP states were built using the ATP models as reference and removing the 2' OH group to form 3' dATP. Alternatively, GTP was aligned with the ATP models with subsequent geometry optimization to force wobble (WB) base pairing (see **Figure 2 lower right**)<sup>60</sup>. For complete details on preparation of RdRp model in association with RDV-TP (with force-field parameterization) and ATP, please see previous work<sup>52,61</sup>.

Note that recent high-resolution characterizations of post-translocation (active-site open) and insertion state (active-site closed) structures of the SARS-CoV-2 RdRp complex became available (PDB:6YYT and 7CTT respectively).<sup>62,63</sup> We structurally aligned the currently modeled NTP initial binding (active site open) state and insertion (active site closed) states with the above two structures, respectively (shown in **SI Figure S1**), which showed high similarities in the active site. In addition, we also aligned the open and closed structures together in order to show subtle structural transitions between the two states (**SI Figure S1**). Due to contact between the fingers tip and thumb subdomain, the open-closed conformation transition is quite limited<sup>20,33</sup>.

#### 2.1.2 Simulation Parameters and Setup:

All MD simulations were performed using GPU-accelerated Gromacs 2021 software<sup>64</sup> with the following forcefields: Am-

ber14sb<sup>65</sup>, Parmbsc1<sup>66</sup>, triphosphate parameters developed previously<sup>67</sup>. Enhanced or umbrella sampling methods were performed using a Gromacs 2021 package patched with PLUMED<sup>68</sup>. Each complex was solvated with explicit TIP3P water<sup>69</sup> in a cubic box with a minimum distance from complex to the wall of 15 Å. Resulting in an average box dimensions of 15.7nm x 15.7 nm x 15.7 nm. The overall negative charge of the complex was neutralized and ions were added to create a salt concentration environment of 100mM. Full simulation systems were ~382,000 atoms in size. A cut-off of 10Å was used to treat short range electrostatics interactions and the Particle-mesh-Ewald (PME) algorithm to treat long range interactions<sup>70</sup>. The LINCS algorithm is used to constrain bonds to hydrogen atoms allowing the use of a 2 fs timestep when integrating the equations of motion<sup>64</sup>. Temperature was kept at 310K using the velocity rescaling thermostat. Pressure was kept at 1 bar using Berendsen barostat<sup>71</sup> during equilibration and Parrinello-Rahman barostat<sup>72</sup> for production, targeted MD (TMD), and umbrella simulation runs. Each system was minimized for a maximum of 50000 steps using steepest-descent algorithm, followed by a slow equilibration with restraints released (every 1 ns) going from NVT (canonical or constant volume and temperature) ensemble to NTP (constant pressure and temperature) as previously used<sup>52</sup>.

For each NTP initial binding and inserted states ten 100 ns equilibration trajectories were launched, with 10 ns removed from the start, to create 900 ns for each NTP state, totaling ~7.2 μs of simulation time for RDV-TP/ATP/dATP/GTP in both open and closed conformation states. The choice of running multiple comparatively short trajectories over a long one is to improve sampling efficiency in the conformation space of NTP bound structure. The generated equilibrium ensemble was then used for analysis and selection of references for free energy calculations.

### 2.2 Reaction coordinate (RC), launching NTP insertion path, and construction the Potential of Mean Force (PMF)

In order to enforce the milliseconds slow conformation transition to happen in all-atom MD simulations at sub-microseconds to measure the corresponding free energetic profile or PMF, one can implement umbrella sampling method<sup>73,74</sup> to enhance sampling, i.e., by imposing harmonic potentials to certain molecular configurations along a selected path or reaction coordinate (RC).

#### 2.2.1 Defining Collective RC(X)

The RC used in current umbrella sampling is the difference in root-mean-square-deviation (RMSD) (eq. 1) of a current frame (coordinates X) with respect to two reference structures, one in the open (NTP initial binding) and the other in the closed (NTP insertion) state, respectively.

$$RC(X) = \delta RMSD(X) = RMSD(X, X_{Open\ ref}) - RMSD(X, X_{Closed\ ref}) \quad (1)$$

The reference structures (Open/Closed ref) were selected using the the first half (50 ns) of the equilibrium trajectory (or quasi-equilibrated). In the case of ncNTP we visually inspected to select

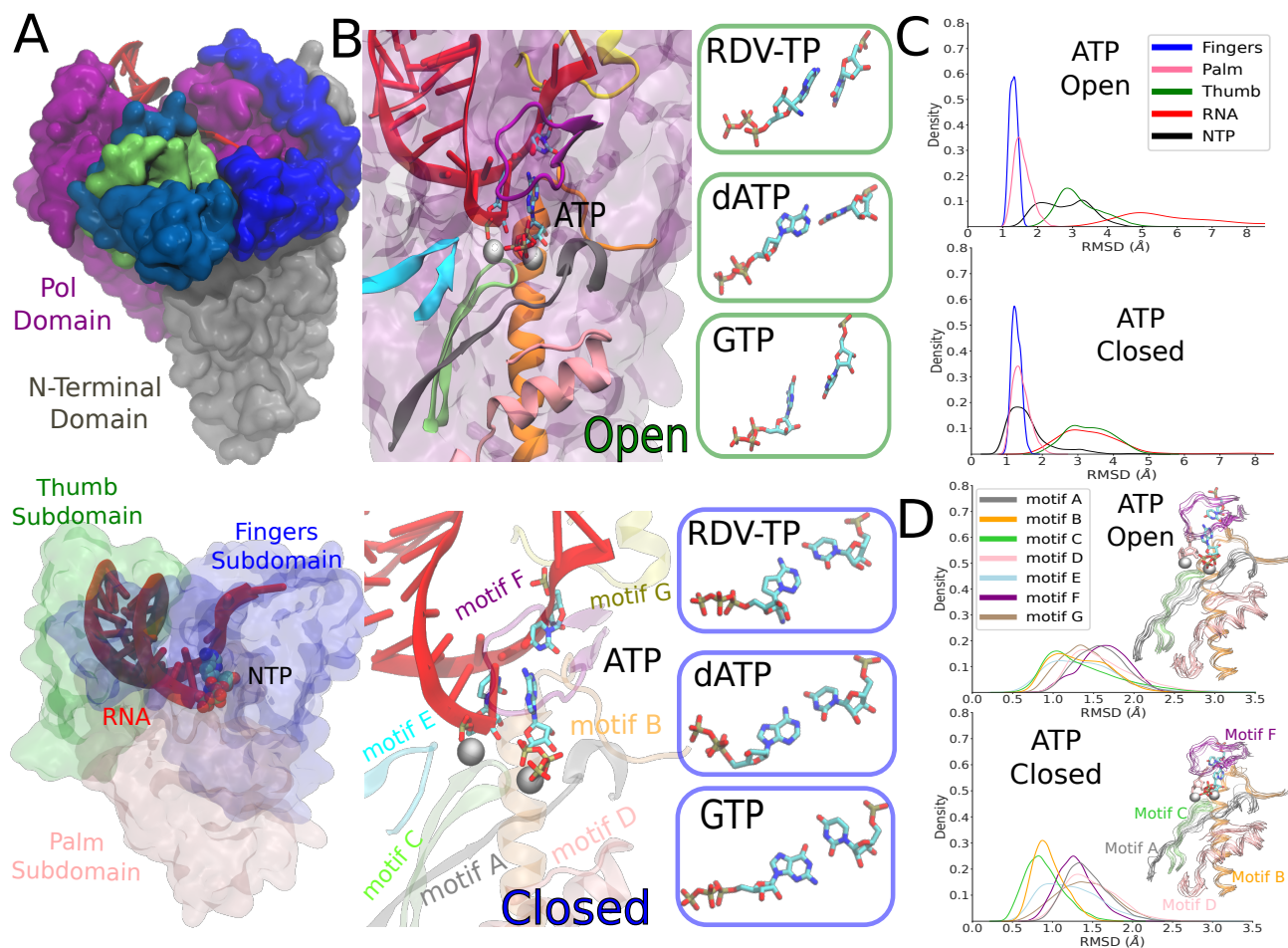


Fig. 1 The constructed structural models of the SARS-CoV-2 RdRp elongation complex in its initial NTP substrate binding (open) and then inserted (closed) states. A: The simulated elongation complex is depicted with the nsp12 pol domain (purple) + N-terminal domain (gray), two cofactor nsp8 (blue) and nsp7 (green), RNA (red) and an NTP in the active site (left). The three major subdomains within the pol domain: fingers in blue, palm in pink, and thumb in green. The RNA is shown in red as well as a NTP shown in space filling spheres (right). B: The modeled and simulation equilibrated ATP is shown as bound initially to an open active site, with the seven protein motifs highlighted in color. The boxes to the right show the initially bound RDV-TP, dATP, and GTP that were also modeled and equilibrated from the simulations (left). The modeled and equilibrated ATP is shown in the insertion or the active site closed state, with the seven structures motifs shown as well, and the boxes to the right displaying the modeled and equilibrated insertion configurations of RDV-TP, dATP, and GTP to the closed active site (right). C: The subdomain root mean squared deviation (RMSD) for simulating initial binding (left) and insertion (right) for ATP, see SI Figure S2 for the rest of the NTPs. D: The motif RMSD for initial binding (left) and insertion (right) for ATP, see SI Figure S3 & SI Table S1/S2 for the rest of the NTPs.

a base pair geometry between the incoming NTP and template nt+1 from the well sampled or representative regions **Figure 2**.

## 2.2.2 Launching NTP insertion path via TMD simulations

After selecting the reference structures, a path of the NTP insertion is generated for umbrella sampling using target molecular dynamics (TMD)<sup>75</sup> simulations. We create a forward path (open to closed ref) and a backward path (closed to open), applying force along the following RC or atomic coordinates ( $X$ ): nsp12 motifs (motif A-G backbone atoms), NTP (heavy atoms), and additionally tested two protocols<sup>52</sup>: i) include or with force on the template nt +1 ii) exclude or without force on the template nt +1.

## 2.2.3 Conducting umbrella sampling simulations

From the TMD paths created between the two reference structures (with only the first half of the respective paths included to avoid large deviations from the references), intermediate structures were evenly (every 0.1 Å along the RC) selected to be used in launching umbrella sampling simulations. The force constants used in the TMD were carried over for umbrella sampling simulations (see **SI Table S3**). The collected RC value histograms were then re-weighted using the weighted histogram analysis method<sup>76</sup> as implemented by the WHAM package<sup>77</sup>. For each set of trajectories making up the umbrella windows, 10-ns trajectory was removed from the start, followed by convergence check for every 10ns. For dATP and GTP, the convergence time ranged from 50ns-60ns (**SI Figure S4**), in the case of no force applied on the template nt +1. GTP and dATP used 24 and 21 windows,

respectively, for a total of  $\sim 1.2 \mu\text{s}$  of simulation time per PMF. Additionally, error bars are estimated using the bootstrapping error analysis method<sup>78</sup> implemented in the WHAM package. For a thorough description on utilizing the umbrella sampling method for constructing the PMFs for ATP and RDV-TP insertion see previous work<sup>52,74</sup>.

### 2.2.4 Testing stabilities of inserted dATP/GTP via SMD simulations

To additionally examine the insertion structures simulated from umbrella sampling, we used steered molecular dynamics (SMD) to check whether the stability of the inserted GTP/dATP show consistencies with the umbrella sampling or the PMFs constructed. The SMD was implemented controlling two center of mass (COM) distances defined as between the heavy atoms of the NTP and the  $C_\alpha$  atoms from residues within  $10 \text{ \AA}$  of the 3' end primer. The geometrical distance between the COMs of ATP in the open and closed configuration is as small as  $2 \text{ \AA}$ , so we only need to pull for up to  $1\text{-}2 \text{ \AA}$  overall, which is planned to be completed within hundreds of nanoseconds affordable. We chose the speed  $1 \text{ \AA}$  per  $100 \text{ ns}$  to balance a slow pulling speed and computational cost. The force constant is  $2.4 \text{ kcal/mol/\AA}^2$  above thermal fluctuations ( $1\text{-}2 K_B T$ ) but not too large ( $< 10 K_B T$ ) so that the introduced force can pull with bias but not strong.

## 2.3 Structural dynamics analyses

### 2.3.1 Calculations for equilibrium check

RMSD is measured for each RdRp complex on the subdomains (protein backbone), RNA (phosphate backbone), and NTP (heavy atoms) as on well as the the key motifs (from A to G) interacting with each NTP (see **Figure 1** and **SI Figure S2, S3**). From **SI Figure S5**, one can also see that system local equilibriums are reached after  $30\text{-}40 \text{ ns}$  of simulation. Both initial binding and insertion equilibrium ensemble trajectories were aligned via the finger's subdomain to their respective initial states after minimization for each NTP. Additionally, equilibration of representative hydrogen bonds was revealed by measuring between proton-acceptor distances (see **SI Figure S5**). Furthermore, catalytic  $\text{Mg}^{2+}$  ions were stabilized in their positioning with respect to coordinated phosphorus atoms from 3'-end primer and bound NTP at the active site (see **SI Figure S6**).

### 2.3.2 Examining NTP-template base pair geometry

Base pair geometry was measured between the incoming NTP and uracil template nt at +1 by calculating the base plane angle ( $C1'-C7-C5$  for NTP and  $C1'-C2-C5$  for template) and the distance between the center of mass (COM) of each base. A similar protocol was followed for measuring the geometry between each NTP and the 3' end primer: measuring the COM between bases and the base plane angle for ( $C1'-C7-C5$  for NTP and  $C1'-C2-C5$  for 3' end primer). Measurements were conducted using the MDAnalysis python package<sup>79</sup> and gromacs gangle module.

### 2.3.3 Calculating HB occupancy in NTP stabilization

Hydrogen bonds (HB) are measured using the gromacs 2021 module. A distance cutoff of  $\leq 3.5 \text{ \AA}$  between donor-acceptor in-

volved and hydrogen-donor-acceptor angle cutoff of  $\leq 30^\circ$  is used as the criterion to indicate a proper hydrogen bond. Unique HB with an occupancy greater than 20% (within a  $900 \text{ ns}$  combined equilibrium ensemble trajectory) are considered for analysis.

Plots are created using python packages seaborn<sup>80</sup> and matplotlib<sup>81</sup>.

## 3 RESULTS

### 3.1 Equilibrium ensemble simulations: protein structural variations and distinctive dynamical responses to different NTPs upon initial binding and insertion

#### 3.1.1 RMSDs on subdomains and motifs

Upon modeling the initial binding (open) and insertion (closed) complexes for the four NTPs, we conducted equilibrium all-atom MD simulations of  $10 \times 100 \text{ ns}$  for each system. We began by measuring and comparing the RMSDs of the protein subdomains, RNA and NTP's using the respective energy minimized structures for insertion or initial binding as references (see **Figure 1** & **SI Figure S2**). For all four NTP binding, the fingers subdomain RMSD is centered around  $1.3 \text{ \AA}$ , and the palm subdomain remains similarly aligned with fingers, especially in the insertion state. The thumb subdomain has a comparatively wide distribution of RMSD in all cases, indicating conformational flexibility. It is interesting to notice that ATP in its initial binding or the active site open state fluctuates similarly as the flexible part of the complex (thumb), while in the ATP insertion or the active site closed state, it fluctuates much less and similarly as the stable fingers/palm subdomain (**Figure 1C**). Indeed all NTP display less variability in the insertion state, along with the fingers/palm subdomain (**SI Figure S2**). The cognate ATP and RDV-TP analogue both exhibit lower RMSDs than non-cognate dATP or GTP in the insertion state. For initial binding, the RMSDs of GTP along with the RNA scaffold are much larger than that for the other NTP binding. Additionally we compared the RMSDs of the seven conserved motifs (see **Figure 1D**, **SI Figure S3** and **SI Table S1/S2**). Motif A, D and F show generally large RMSDs, particularly, in the open state. Both motif B and C show low RMSDs in the insertion state. Motif-C demonstrates notable stability among all motifs for all inserted NTPs, likely due to it hosting the key catalytic residues (S759/D760/D761). The RMSDs of other structural motifs displayed significant variations upon initial binding of different NTPs. In the insertion state, motifs F and G show more distinctions dynamically than rest of motifs for different NTP substrates.

#### 3.1.2 NTP-template association geometry

Next, we examined the base association or pairing geometry between the initially binding/insertion NTP and template +1 nt (**Figure 2**). In the insertion equilibrium ensemble (active site closed), we observed that the NTP-template distance distribution predominately centers at  $\sim 6.5 \text{ \AA}$  and base plane angle  $\sim 30^\circ$  (see **Methods**), resulting in either the stable Watson-Crick (WC) base pairing (for ATP/RDV-TP/dATP-template) or wobble base pairing (WB) interactions (for GTP-template uracil). The probability of WC/WB base pairing is high for ATP (69%), RDV-TP (94%) and GTP (70%), comparatively low for dATP (47%). Indeed, dATP

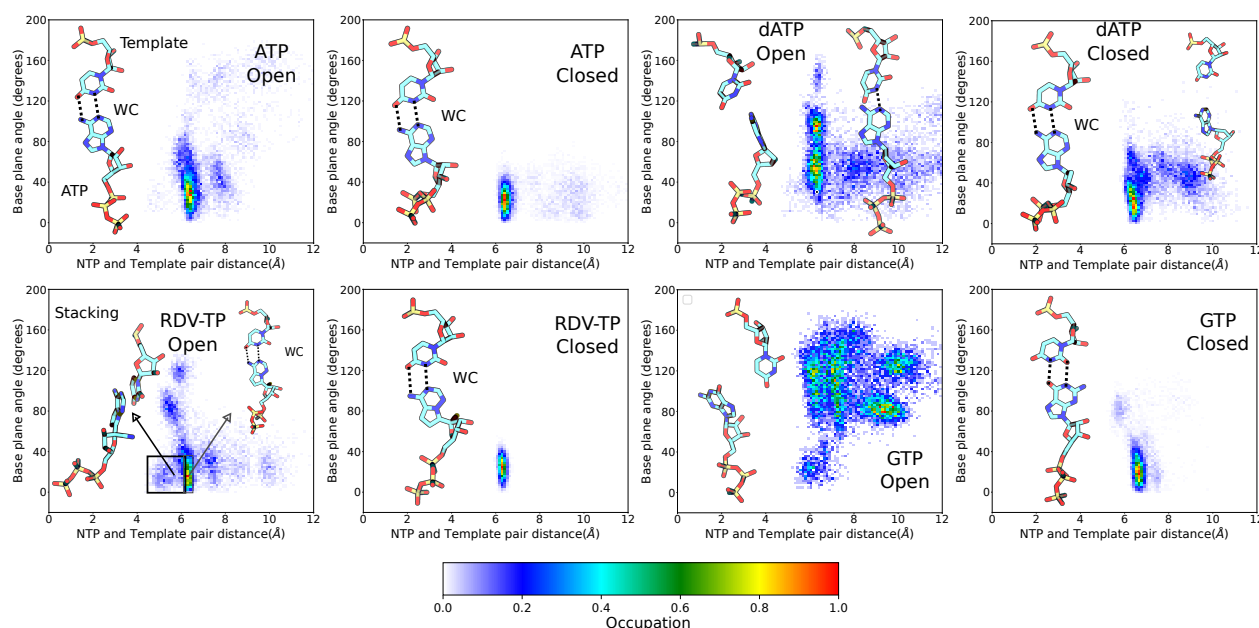


Fig. 2 NTP-template association geometries from ensemble equilibrium simulations. The geometric measures (see Methods) between the uracil template +1 nucleotide (nt) and individual incoming NTP are displayed, for ATP (upper left), RDV-TP (lower left), dATP (upper right), and GTP (lower right), upon initial binding (left) and insertion (right) for each NTP species. Licorice representations of the NTP and template nt show the dominant geometries for each simulation system. Dotted lines indicate the hydrogen bonds for the standard Watson-Crick (WC) or wobble base (WB) pairing (percentile denoted).

upon insertion displays more flexibility than other NTPs in association with the template nt (see **Figure 2 upper right**). In the NTP initial binding ensemble (active site open), the NTP-template association geometries vary significantly. Upon ATP initial binding, as a significant amount of WC population (48%) is identified, a comparable amount of un-paired or weakly paired (single HB) ATP-template uracil configurations are also present. Upon initial binding of RDV-TP, three configurations have been identified, with either the WC base pairing (36%) or base stacking (38%) being stabilized, and (26%) unpaired<sup>52</sup>. For non-cognate NTPs, the initially bound dATP marginally forms WC base pairing (12%) with the template nt, while GTP upon initial binding cannot form stabilized WB base pairing with the template uracil.

### 3.1.3 NTP-3' end primer association geometry

Additionally, we measured the geometry between the NTP and 3' end primer (**SI Figure S7**). From different NTP insertion ensembles, the NTP-3' end primer distribution centers closely around  $\sim 4\text{--}6$  Å and a base plane angle  $\sim 150^\circ$  to  $170^\circ$ , showing fair stability. In contrast, upon initial binding, the mismatch GTP associated with the 3' end primer in largely diverse geometries (distance spans from 4 to 13 Å and the angle varies from  $20^\circ$  to  $170^\circ$ ). The noncognate dATP upon initial binding also shows diverse geometries in association with the 3' end primer (distance spans from 4 to 9 Å and the angle varies from  $40^\circ$  to  $180^\circ$ ). For cognate ATP and RDV-TP, the initial binding geometries with respect to the 3' end primer are much more localized, i.e., the distance dominantly spans from 4 to 5 Å, with small populations in between 8–10 Å; the angle varies from around  $100^\circ$  to  $180^\circ$  below than the non-cognate species. In summary, in the active site closed state or

insertion equilibrium ensemble, various NTPs display much less variation of geometries with respect to the template +1 nt or the 3' end primer than those in the active site open state or initial binding equilibrium ensemble. The highly diverse configurations of initially bound NTP (particularly non-cognate NTP) suggest that detection of various species of incoming NTP starts well from the beginning, i.e., upon the initial NTP binding when the active site remains open.

### 3.1.4 HBs stabilizing NTP from equilibrium ensemble

Furthermore, we also measured the HB occupancies for various NTPs in respective associations with the protein, template +1 nt, and 3' end RNA primer (histogram statistics shown in **SI Figure S8**), for both the NTP initial binding and insertion equilibrium ensembles. We observed an increase in HB occupancies between the NTP triphosphate tail and protein from initial binding to the inserted ensembles for ATP and dATP, with more protein-sugar HBs for the inserted ATP (with D623 and N691) than for the inserted dATP. On the other hand, the inserted RDV-TP ends up with fewer protein-triphosphate HBs than the inserted ATP but much stronger HBs (with T687 and N691) in the protein-sugar association. Meanwhile, GTP insertion led to many but weak (low occupancy) HBs formed, indicative of some instability.

### 3.1.5 HBs stabilizing template/primer from equilibrium ensemble

Additionally (see **SI Figure S9**), in the open state, several protein residues (motif-F K551, K545, A558 and motif-B S682) form HBs with template +1 nt upon initial binding of GTP. While protein S501 (from motif G) forms HB with template +1 nt strongly



in the presence of ATP (71%) and RDV-TP (78%), or marginally upon dATP (40%) and not present in GTP. In the closed state, the S501-template HB persists and becomes highly stabilized for every NTP insertion state. These findings again reflect distinctive HB patterns formed around the active site upon association of various NTPs, from initial binding to insertion. Nevertheless, due to various populations of NTP binding configurations, especially in the non-cognate initial binding state, it is not clear which HB interactions contribute to stabilize or destabilize cognate vs ncNTPs for nucleotide selectivity.

### 3.2 Constructing the PMFs for noncognate GTP/dATP from initial binding to insertion

Upon conformational samplings from the equilibrium ensemble simulations, we noticed essential variations of protein structural motifs along with diverse NTP dynamics. Accordingly, we included structural motifs and NTP configurations into constructing a collective reaction coordinate, based on RMSD changes with respect to the open and closed state<sup>23,24</sup> structures (see **Methods**)<sup>52,57</sup>. Upon selecting appropriate reference states for each NTP initial binding/insertion system, we then proceeded to calculate the free energy profiles or PMFs of NTP insertion and to quantify the processes for NTP selectivity from initial binding (active site open) to the insertion (closed). The procedures of the PMF construction via umbrella sampling simulations can be found in **Methods**. The overlaps of umbrella sampling windows are provided in **SI Figure S10**. In current studies, we constructed the PMFs of the non-cognate GTP and dATP insertion, so that we can compare with the PMFs obtained previously for cognate ATP and RDV-TP<sup>52</sup>.

#### 3.2.1 Non-cognate GTP binding to insertion PMF

From the constructed PMF, GTP upon initial binding displays stabilization in comparison with the insertion state, with  $\Delta G = G_{\text{insertion}} - G_{\text{initial binding}} \sim 2$  kcal/mol ( $> 0$ ). This is in contrast with cognate ATP and RDV-TP analogue insertion, which are more stabilized in the insertion state, demonstrating  $\Delta G (< 0)$  values of  $-5.2$  kcal/mol and  $-2.7$  kcal/mol, respectively (**Figure 3**)<sup>52</sup>. Nevertheless, the initially bound GTP forms no WB base pairing with the template nt+1, whereas approximately 81% WB base pairing between GTP and the template is identified in the insertion state. Hence, there have to be other interactions to stabilize the initial binding GTP (to be addressed in next subsection). Additionally, the PMF calculations show that the non-cognate GTP is subject to an insertion barrier  $h^{\text{ins}} \sim 7$  kcal/mol from the initial binding state, much larger than that of ATP and RDV-TP ( $h^{\text{ins}} \sim 2.6$  kcal/mol and  $1.5$  kcal/mol, respectively; see **Figure 3**)<sup>52</sup>. Note that the convergence of the PMF for GTP was reached after 50 ns per window in the umbrella sampling simulation, following the protocol without (or with) enforcing on the uracil template +1 nt (**SI Figure S4**), which show similar results.

The placement of the GTP insertion PMF relative to that of ATP/RDV-TP was conducted according to alchemical simulations performed recently<sup>82</sup>. Given that the GTP-template WB base pairing geometries are comparatively stable in the insertion state (**Figure 3 lower right**), we placed the GTP insertion state  $\sim 3$

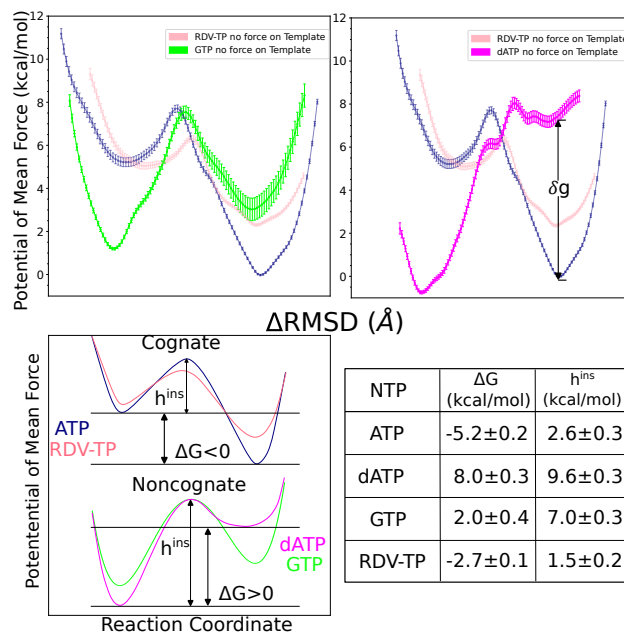


Fig. 3 The potentials of mean force (PMFs) calculated for NTP from initial binding (active site open) to the insertion (closed) state via umbrella sampling simulations. The difference of RMSDs with respect to open and closed reference structures  $RMSD(X, X_{\text{Open Ref}}) - RMSD(X, X_{\text{Closed Ref}})$ <sup>52</sup>, was used as the reaction coordinate in the PMF construction (see Methods, see umbrella windows in SI Figure S10). Top left shows the PMF for GTP (green) and top right for that of dATP (magenta), both in comparison with the PMFs obtained for cognate ATP (blue) and analogue RDV-TP (pink)<sup>52,82</sup>. Smoothing has been applied to each PMF for clarity (see original PMFs and converging tests in SI Figure S4). Note that the placement of the PMF of GTP relative to that of ATP/RDV-TP is according to the alchemical calculation in the closed state<sup>82</sup> while the placement of the PMF of dATP relative to that of ATP/RDV-TP is still uncertain, while the relative binding free energy at the closed state, denoted as  $\delta g$ , is estimated between 2-7 kcal/mol (see text). Bottom left shows schematics of the different PMF profiles for the cognate and ncNTP insertions. Bottom right shows the insertion free energy and barrier heights for the four PMFs shown on the top panel.

kcal/mol above that of the ATP insertion state, as the alchemical calculations indicate that  $\Delta\Delta G_{\text{binding}} = 2.95 \pm 0.66$  kcal/mol for the relative binding free energy of GTP with respect to ATP in the insertion state<sup>82</sup>.

#### 3.2.2 Non-cognate dATP binding to insertion PMF

The dATP insertion PMF demonstrates significantly more stability in the initial binding state, with  $\Delta G \sim 8.0$  kcal/mol comparing with the insertion state. Meanwhile, the insertion state of dATP can be hardly or only marginally stabilized. Besides, the non-cognate dATP is subject to an insertion barrier  $h^{\text{ins}} \sim 9.6$  kcal/mol, the highest among all NTP insertion cases examined (**Figure 3**). Similar to GTP, dATP does not form WC base pairing with the template nt+1 in the initial binding state but is capable of forming partial WC pairing with the template, though intermittently (64%), in the insertion state umbrella sampling windows. Therefore, there must also be some additional interactions stabilize or trapping dATP in the initial binding state, as reflected from the highly tilted PMF toward the initial binding state. Similar to GTP

case, the constructed PMF of dATP reached convergence after 50 ns per window of the umbrella sampling simulation, following the protocol without (or with) force applied on the Uracil template (SI Figure S4), which both show similar results.

The placement of the dATP insertion PMF relative to that of ATP/RDV-TP is, however, less certain. Given a wide range of association configurations between dATP and template even in the insertion state (Figure 2 upper right), one cannot directly use the alchemical calculation results, which were conducted around local configurational space (for stabilized ATP and slightly destabilized dATP) with limited sampling<sup>82</sup>. An estimation is nevertheless made here, based on relative binding free energy calculated locally between dATP and cognate ATP ( $\sim 2$  kcal/mol) along with that from the mmPBSA calculation ( $\sim 7$  kcal/mol)<sup>82</sup>, suggesting a range of free energetic values 2-7 kcal/mol in between the inserted dATP and ATP (shown as parameter  $\delta g$  in Figure 3).

### 3.2.3 SMD tests on stabilities of inserted non-cognate GTP/dATP

In order to further test the consistency of the PMF of insertion results for both GTP and dATP, we used SMD simulations to probe the stability of the insertion states for GTP (SI Figure S11) and dATP (SI Figure S12), respectively. To do that, the non-cognate GTP or dATP starting from the insertion state is pulled slightly away from the active site in the SMD simulations (see Methods), i.e., to start from the insertion state (active site closed) to the initial binding state (open) under the SMD force. In the case of GTP, it was robustly maintained within the insertion state without being able to cross the barrier (from closed to open) from three trials of the SMD simulations (one 500 ns, two 300 ns). In contrast, dATP was able to be readily pulled from the insertion state toward the initial binding state (closed to open) under the SMD force in all three simulations (300 ns each), demonstrating much lower stability or barrier from closed to open than that of GTP. These observations further support the instability of the dATP insertion state in comparison with the GTP insertion state, as being reflected from the PMFs constructed (see Figure 3).

### 3.2.4 Umbrella sampling on conformation subspace of ncNTP-template association

Next, we would proceed to examine additional interactions that stabilize or trap the non-cognate dATP or GTP in their initial binding state, i.e., as revealed from the PMFs constructed from the umbrella sampling simulation. However, before that, we want to examine whether the conformational space sampled between dATP/GTP and the template nt+1 near the initial binding (open) and insertion (closed) equilibrium in the umbrella sampling simulations overlap well with that from the equilibrium ensemble simulations. To do that, the NTP-template +1 nt base pairing geometries are compared between the equilibrium ensemble and the umbrella sampling simulations, from the latter three umbrella windows around the open/closed equilibrium were selected (40 ns RDV-TP/GTP/dATP and 160 ns ATP of simulation time per window). In Figure 4, one can see that the sampled geometries from the umbrella samplings are comparatively restricted, especially in the NTP initial binding state, but overlap well with the domi-

nant configurations sampled from the equilibrium ensemble simulations, in particular, in the insertion state. For ATP and GTP, the initial binding configurations from the umbrella samplings overlap well with some stabilized population from the equilibrium ensemble, though deviations from the equilibrium ensemble show in the umbrella sampling case, indicating potentially the forcing impacts from the umbrella sampling simulations. For RDV-TP initial binding, the stable configuration of the RDV-TP-template in base stacking is well sampled, as the umbrella sampling path was launched from the base stacking configurations which leads to a low insertion barrier<sup>52</sup>. For the dATP initial binding, the umbrella sampling simulation also well covers one stabilized population. Hence, the significantly stabilized configuration of GTP/dATP upon initial binding, detected from the umbrella sampling simulations, appears to be well located within a subspace in the equilibrium ensemble.

Since the initially bound non-cognate dATP or GTP could not be well stabilized by association with the template +1 nt (nor 3' end primer), it must be interactions from the RdRp protein along with the RNA scaffold around the active site that strongly hold the non-cognate dATP or GTP, which we examine and elaborate below.

### 3.3 Trapping noncognate dATP/GTP upon initial binding by persistent HB interactions from motif F/G/A, to NTP, template nt +1, or 3' end primer.

Though there were no WC or WB base pairing interactions observed for non-cognate dATP/GTP with the template uracil upon initial binding toward the open active site, some populations of dATP/GTP are strongly stabilized upon the initial binding according to their insertion free energetics or PMFs (shown in Figure 3). To gain understanding of this phenomena, we analyzed all HB interactions present among protein residues, NTP, and RNA strands (template and primer), around the active site. Given the variations amongst NTPs, we simplified the analyses by summing up overall HB populations exceeding 20% of occupancy (during the umbrella simulation 40ns/window) amongst the protein (residues within 10 Å of the active site center), template +1 nt, 3' end of the primer, and the initially bound NTP; those HBs were further grouped according to interactions with the NTP on polyphosphate, sugar, or base (see Figure 5). The cumulative HB measure is then calibrated over that of the cognate ATP. Accordingly, those HB components showing small deviations from the ATP level are similar in strength with that in the ATP binding system. Those deviating largely from the ATP bar are notable HBs that differentiate noncognate GTP/dATP from ATP.

As a result, by reading HB components shown in Figure 5, one can see that GTP differs from ATP in initial binding by protein HBs formed with its phosphates and sugar, and protein HBs with template nt+1, similar to ATP by protein HBs with its base, and almost no HB formed with 3'-end primer as in ATP binding (SI Figure S14). In comparison, one sees that dATP differs from ATP mainly by protein HBs with its base and 3'-end primer, and 3'-end primer HB with its sugar; while dATP is similar to ATP on protein HB with its sugar and with template nt.

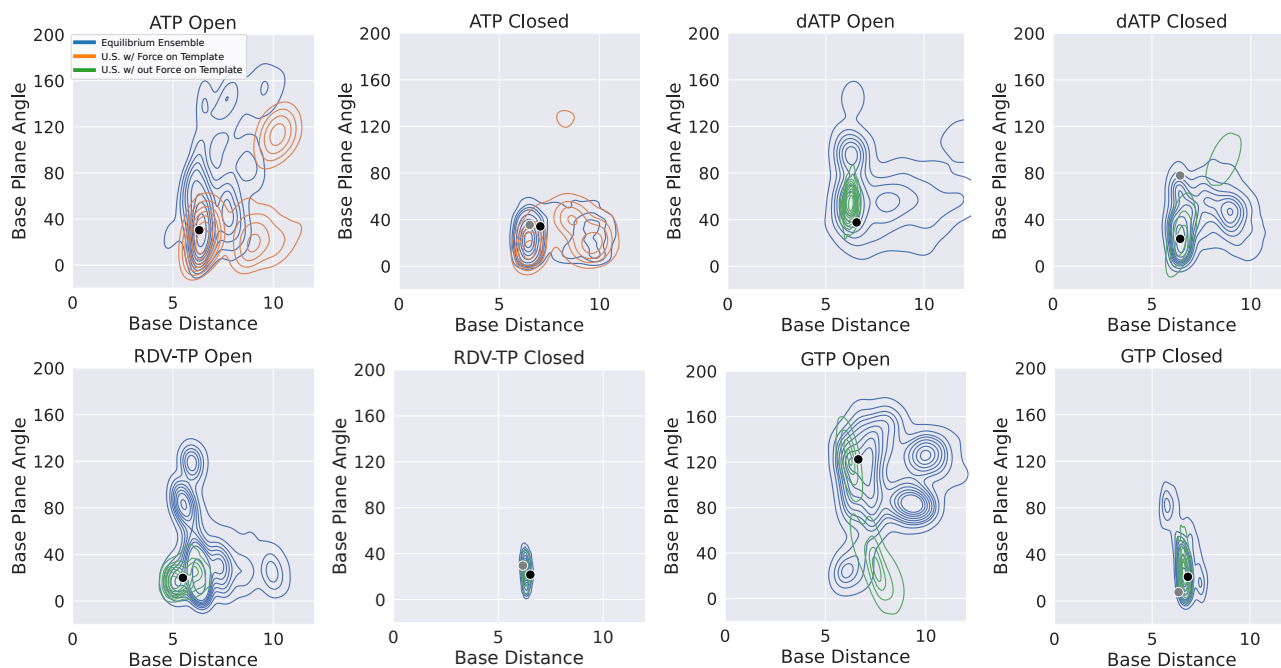


Fig. 4 Comparing NTP-template association geometry distributions obtained from the umbrella sampling simulations (for the PMF calculations) and that from ensemble equilibrium simulations. A kernel density estimate has been used to visualize the data. Each simulation system (open and closed, for ATP, RDV-TP, dATP and GTP as in Figure 2), the equilibrium ensemble distribution is shown (blue), along with that obtained from the umbrella sampling (w/ force on template +1 nt in orange; w/out force on the template +1 nt in green; see Methods and SI for further details). The black dot indicates the reference state used to generate the initial paths (see Methods) for the umbrella sampling, and grey dot the reference state used in the alchemical calculations<sup>82</sup>, for the full dataset see SI Figure S13.

Notably, one finds that GTP initial binding is stabilized predominantly by HBs (and salt bridges in the case of positively charged residues LYS/ARG) formed between its polyphosphate group and the protein residues in motifs (mainly motif F and A). In addition, the protein residues (motif F and G) stabilize particularly the template nt +1, due to the absence of WC or WB base pairing between the initially bound GTP and the template nt. Meanwhile, there is lack of protein HB association with the 3' end primer around the initially bound GTP. However, this association is present around the initially bound RDV-TP (see SI Figure S14), and the association appears even stronger around the non-cognate dATP initially bound.

In the case of dATP initial binding, though it does not form WC base pairing with the template nt+1, the adenine base is nonetheless stabilized via protein HB (again from motif-F). Intriguingly, despite dATP's lack of a 2' OH group, it still maintains a strong HB via the sugar and the 3' end primer. As mentioned above, the 3' end primer around the initially bound dATP displays the strongest HB interactions among all the NTP's with the protein.

Below, we show the individual HB interactions structurally around GTP and dATP and compare them with those in the case of cognate RDV-TP or ATP binding, respectively (Figure 6 and Figure 7). One can find statistics of HBs formed between NTP (polyphosphate, sugar, or base) and protein residues or template nt +1/3' end primer, and between protein residues and template nt+1 or the 3' end primer (SI Figure S14).

### 3.3.1 GTP binding stabilized by protein HBs with its phosphates and sugar along with template stabilization

Notably, we have found that upon initial binding, GTP exhibits very strong HB or salt bridge interactions between motif-F K551/R553 (together with motif-A K621) and the polyphosphate (see Figure 6A, SI Table S4). We attribute such interactions to hinder the insertion of GTP, or say, the phosphate-K551/R553 interactions contribute significantly to the barrier of GTP insertion. In the cognate initial binding, the ATP sugar forms a HB with motif-C D760. In contrast, GTP initial binding is mainly stabilized by HB between sugar and motif-A D623 (from umbrella sampling simulations). Instead of WB pairing with the non-cognate GTP upon initial binding, the template nt+1 base forms HBs with motif-F K545 and A558 (see Figure 6B, SI Table S4). Furthermore, the template nt +1 backbone forms HB with motif-G K511, as opposed to motif-G S501 seen in the other NTP binding cases. Overall, the protein motifs F and G seem to well stabilize the template nt +1 upon the base mismatched GTP binding. We attach stereo views for Figure 6A & B in SI Figure S15.

In the case of RDV-TP initial binding via base stacking with the template nt +1 (in the absence of force on the template nt), however, only one HB is observed on the polyphosphate from motif-F K551 (see Figure 6C, SI Table S4), so that the RDV-TP won't be hindered by the phosphate interaction for its insertion<sup>52</sup>. In addition, the base stacking configuration of RDV-TP with the template allows for a unique HB to form between motif-B S682 and the RDV-TP base, while the template nt +1 has fewer HBs with

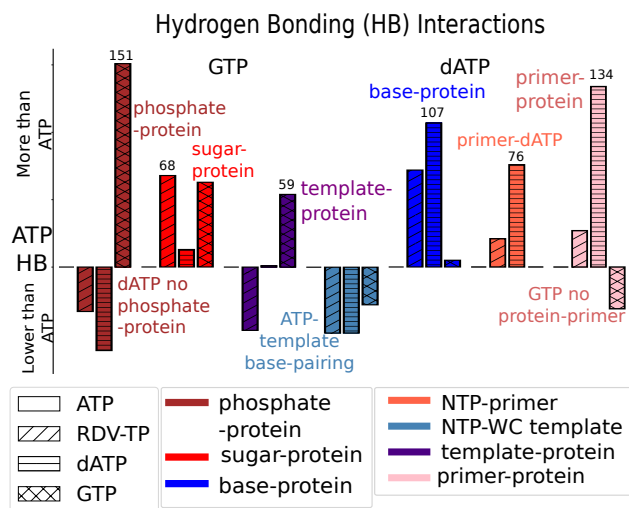


Fig. 5 Summary of hydrogen bonding (HB) interactions that stabilize non-cognate GTP/dATP or surroundings upon initial binding from umbrella sampling simulations. Four interacting partners are considered: protein, incoming NTP (ATP, RDV-TP, dATP, and GTP), uracil template nucleotide +1, and the 3'-end primer. The HBs with occupancy >20% in the simulations were identified among these four interaction partners. The relative HB occupancy levels are shown for the initial binding dATP and GTP (along with RDV-TP) with respect to that of cognate ATP as reference (with bar; see SI Figure S14 for full HB statistics for all the simulation systems).

motif-F than that upon GTP initial binding (Figure 6D, SI Table S4).

### 3.3.2 dATP binding stabilized by protein HBs with its based and with 3'-end primer stabilization

In comparison, as from current umbrella sampling simulations, dATP upon initial binding forms a single HB with template nt +1 at a very high occupancy of 95%. In addition, a HB is uniquely established between the dATP base and motif-F T556 at an occupancy of 87% (Figure 7A, SI Table S4). The dATP initial binding also forms two persistent HBs between the sugar and motif-C D760 and the 3' end primer, respectively. The template nt +1 is further stabilized by interactions with motif-F K545 and S501G, as seen with the cognate initial binding (Figure 7B, SI Table S4). Importantly, 3' end primer forms the most persistent HBs with motif-F K545/R555 and motif-C S759 in the case of dATP initial binding. We attach stereo views for Figure 7A & B in SI Figure S16.

In the case of ATP initial binding, its base is stabilized by forming occasional WC base pairing. Note that the initially bound ATP formed two HBs with template nt +1, with occupancies of 50% and 44%, respectively. The ATP sugar forms a consistent single HB with motif-C D760. Stable associations also form between motif-F K551/R553/R555 and the ATP polyphosphate (Figure 7C, SI Table S4). Such interactions were suggested to facilitate the cognate ATP insertion instead<sup>52</sup>. The template nt +1 forms stable HB with motif-F K545 and motif-G S501, similar to dATP initial binding. In addition, the 3' end primer forms only a single transient HB with motif-C S759 (Figure 7D, SI Table S4), weaker

than that is present between the 3' end and motif-C/F upon the dATP initial binding (Figure 7B, SI Table S4).

## 4 DISCUSSION

In current work, we have focused on computationally probing from initial binding to the insertion and selectivity mechanisms of noncognate natural nucleotides to SARS-CoV-2 RdRp. Note that for viral RdRp core structures, the fingers subdomain has its tip touching the top of thumb subdomain to encircle the active site, so that it cannot support large conformational changes upon an incoming nucleotide binding and insertion, during which nucleotide selection takes place<sup>20,33,34</sup>. The insertion step process involves subtle conformational change of the RdRp pol domain (Figure 1), leading to essentially an active site open or nucleotide initial binding state to the active site closed or insertion state, with coordination of seven highly conserved structural motifs. In all NTP incorporation systems simulated, the fingers subdomain displays similar conformational flexibility as the palm subdomain, which moves closer to the finger's subdomain in the insertion state than in the initial NTP binding state (SI Figure S2). In the insertion state of cognate ATP or RDV-TP, motif-B and C have similar conformational flexibility (via RMSD) demonstrated in the equilibrium simulation, while this feature is absent in noncognate GTP and dATP (SI Figure S3). Overall, the motifs respond differently for each incoming NTP studied. The equilibrium ensemble simulations showed generically that the insertion state sampled a restricted subspace between the NTP and template nt +1 over that of the initial binding state, which indeed accommodates a wide range of configurations (Figure 2). Additionally, RNA template/primer nucleotides or protein residues around the active site forms more HBs with the NTP in the insertion state than in the initial binding state (SI Figure S8).

Due to time scale limit of equilibrium sampling, we probed the NTP insertion dynamics and calculated the corresponding insertion energetics using the umbrella sampling methods. The energetic profile or insertion PMF was constructed along a collective RC according to a difference of RMSDs between the modeled intermediate structure of the RdRp and the open and closed reference states, respectively. The essential atomic coordinates included those of backbone atoms from seven highly conserved structural motifs (A to G), which are crucial for recruiting nucleotide substrates with selectivity and supporting catalysis<sup>83</sup>, and heavy atoms on incoming NTP along with (or without) the template nt +1. While such a choice on enforcing on the template nt or not played some significant role in the PMF construction of cognate ATP and analogue RDV-TP (see SI Figure S4)<sup>52</sup>, it made little difference in the non-cognate dATP or GTP results, e.g., as observed from the base pairing geometry measured between NTP and template nt +1 (Figure 4 & SI Figure S13). The insertion barriers were not affected much by the choices for dATP or GTP either (SI Figure S4). In contrast with the insertion PMFs of cognate ATP and RDV-TP analogue that bias toward a more stabilized insertion state, we have found intriguingly that the initial binding states for ncNTPs (dATP and GTP currently) can be much more stabilized than their insertion state. In addition, the insertion barrier of ncNTP becomes very high (up to 7-10 kcal/mol), also in

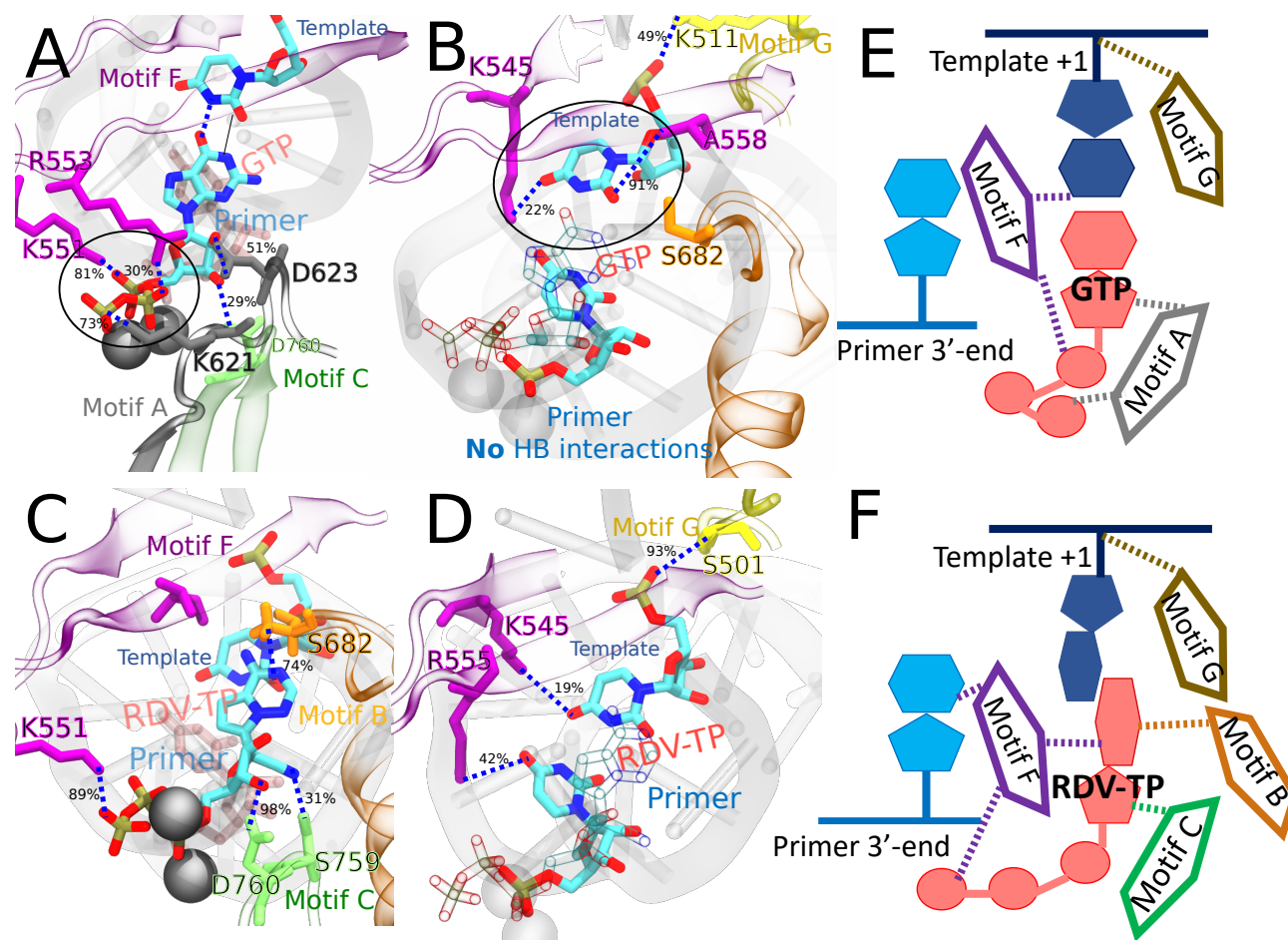


Fig. 6 Comparing key interactions that stabilize the non-cognate GTP (and surroundings) and drug analogue RDV-TP in the initial binding system. The conserved protein motifs are shown in cartoon representation while interacting residues from these motifs are shown in licorice using the same color. A/C: Incoming GTP/RDV-TP and template nucleotide are shown in licorice colored by atom name. Hydrogen bonds (HBs) formed between GTP (or RDV-TP) and template nucleotide uracil/protein/3' end RNA primer. Black circle highlights the essential interactions involved in 'trapping' non-cognate GTP phosphate in the initially bound state (A), which are absent for RDV-TP phosphate (C). HBs formed between the protein and template for stabilization, without proper WB base pairing for GTP-template (B) and with template base stacking in RDV-TP (D). For detailed HB occupancy plots see SI Figure S14. E: Schematic and cartoon of key motifs stabilizing GTP. F: Schematic and cartoon of key motifs stabilizing RDV-TP. For stereoscopic views of A/B see SI Figure S15.

contrast with the marginally low insertion barriers of cognate ATP and RDV-TP ( $\sim 2$  kcal/mol) identified previously<sup>52</sup>. Such free energetic calculations and structural dynamics examinations reveal intrinsic nucleotide selectivity of SARS-CoV-2 RdRp, i.e., to inhibit the insertion of ncNTPs to the active site by trapping the ncNTPs *off-path* upon initial binding to the peripheral of the RdRp active site.

#### 4.1 Free energetics favor insertion of cognate NTPs but disfavor insertion of non-cognate NTPs

In previous work<sup>52</sup>, we calculated the insertion PMFs for ATP and RDV-TP, respectively. While the RDV-TP initial binding could form WC base pairing with the template nt+1, a more stabilized conformation was found for RDV-TP in base stacking with the template nt + 1. Besides, the RDV-TP insertion barrier would become high ( $h^{ins} \sim 5$  kcal/mol) when the enforcing in the umbrella sampling simulation included the template nt + 1 (SI Fig-

ure S17A). The striking feature was due to the enhanced HB (or salt-bridge from positively charged LYS/ARG) interactions between the motif-F residues (K551/R553/R555) and the RDV-TP polyphosphate (SI Figure S17B), upon the enforcing on the template nt. Removing forcing on the template nt+1, i.e., allowing sufficient thermal fluctuations on the template, however, the motif-F interactions with the polyphosphate reduce, and the insertion barrier lowers to a marginal value ( $h^{ins} \sim 1.5$  kcal/mol).

Upon the cognate ATP binding, including the template nt + 1 for enforcing in the umbrella sampling simulations nevertheless supports an insertion barrier  $h^{ins}$  as low as  $\sim 2.6$  kcal/mol (SI Figure S17A). The well controlled template nt+1 facilitated WC base pairing and supported enhanced HB interactions between the motif-F K551/R555 and polyphosphate (SI Figure S17C). Consequently, it was suggested that the motif-F interactions with the phosphate facilitate insertion of the cognate ATP, while the motif F-phosphate interactions also appear to be important for NTP insertion in the PV RdRp<sup>84</sup>. Regardless of the exact proto-

col, the insertion state was always more stable than the initial binding state for the cognate ATP and analogue RDV-TP, both of which are actively biased or recruited into the closed active site for catalytic incorporation to the 3'-end of primer, experiencing marginally high barriers of insertion to the active site<sup>52</sup>.

In contrast, upon the initial binding of ncNTP (dATP or GTP), a large configuration space of the ncNTP with respect to the template+1 nt and 3'-end primer was identified, and the PMF tilts toward the initial binding state, i.e., biased or energetically stabilized upon initial binding of certain ncNTP configurations (**Figure 3 & SI Figure S4**). Additionally, the insertion barriers become very large for noncognate GTP and dATP ( $\Delta G^{\ddagger} \sim 7.0$  and 9.6 kcal/mol, respectively). The stability bias toward the initial binding state and tremendously large barrier of insertion seem to trap the ncNTP upon initial binding at certain configurations (or the *off-path*), which would likely lead to dissociation of the ncNTPs from the active site or from the RdRp in the end.

Our current discoveries on such stabilization of the non-cognate substrate upon initial binding may seem counterintuitive. Commonly, high binding affinity of a ligand substrate to the receptor protein indicates a preference of the receptor to the substrate<sup>85,86</sup>. Such an idea prevails in the drug design which aims at identifying high-affinity binders. In the current viral RdRp system, however, the NAC proceeds with two pre-chemical steps: substrate initial binding and insertion. As shown in current study, a high substrate binding affinity at the first step may also contribute to high insertion barrier for the second step, which slows down the NAC or an enzymatic cycle. Hence, the ncNTP stabilization or trapping upon initial binding becomes an intriguing but effective strategy to deter or inhibit the non-cognate substrate from further incorporation. On the other hand, antiviral drug design targeting on the viral RdRp would develop nucleotide analog drug compounds that are capable of avoiding or escaping from such a trapping mechanism of nucleotide selectivity upon initial binding.

#### 4.2 Key residues from conserved motifs detect and stabilize the non-cognate dATP/GTP and its surroundings at the initial binding off-path state

Current free energy calculations reveal that the noncognate GTP/dATP is more stabilized upon initial binding to the RdRp active site than in the insertion state, in contrast with cognate ATP/RDV-TP to be more stabilized in the insertion state. To explain notable stabilized configurations sampled for the ncNTP initial binding state from the umbrella sampling simulations, we identified a variety of HB interactions around the active binding site formed among NTP (base, sugar and phosphate), RNA strands (template and primer), and the conserved protein motifs (**Figure 5**). To well explain the trapping mechanism, we can also compare current system with a previously studied RdRp from Enterovirus or EV<sup>34</sup>, which is structurally similar to SARS CoV-2 RdRp. In EV RdRp, NTP insertion to the active site is suggested via several steps: first the base recognition, next the ribose sugar recognition, and then followed by the active site open to closed conformational transition, accompanied by the palm subdomain

(motifs A,B,C,D,E) closing. Below we connect current observations of NTP initial binding to those suggested steps.

In the case of GTP initial binding, the template nt +1 (uracil) fails to interact with the mismatch GTP in the absence of WC or WB base pairing. The template nt +1 and GTP cannot be mutually stabilized, hence the protein motif-F residues (K545/A558) respond by stabilizing the template base. Additionally, motif-G K511 forms HB with the template RNA backbone, instead of S501 in the ATP/RDV-TP system. Given the non-stabilized GTP base, the sugar is next checked by the protein via motif-A D623, which is prominent upon GTP initial binding (in the umbrella sampling ensemble). The D623 interaction brings motif-A closer to GTP and allows a unique HB (or salt bridge) from K621 to the polyphosphate. Substantial HBs (or salt bridges) with the polyphosphate come additionally from motif-F K551/R553, similarly as seen in RDV-TP (with force on template nt +1). In both cases of initial binding (GTP and RDV-TP with forcing on the template nt+1), the insertion barriers are high. Such observations support the previously proposed mechanism that the Lys/Arg interactions with phosphates inhibit the ncNTP insertion but facilitate cognate NTP insertion<sup>52</sup>, or say, the protein-NTP phosphate interactions play a significant role in nucleotide selectivity or fidelity control. Recent NMR experiments on the structurally similar PV RdRp suggest that interactions from charged residues in motif-F are an important fidelity checkpoint, as they allow the triphosphate to rearrange prior to catalysis<sup>87</sup>.

In the case of dATP binding, obviously, it fails on the sugar recognition. The sugar is unable to anchor in the active site due to missing the 2' OH functional group. Instead, the 3' OH group forms HB with motif-C D760 (**Figure 7A**), similar to ATP and RDV-TP. On the other hand, since dATP has the same base as the cognate ATP, it is capable of forming stable WC base pairing with the template nt, but it does not. Indeed, dATP base adopts a tilted conformation, which supports only one persistent HB w/ template nt +1 (**Figure 7A**). The dATP base is then further stabilized by motif-F T556. The missing WC base pairing between dATP and the template nt +1 is supplemented by another HB formed between the dATP base and motif-F K545 (**Figure 7B**). In addition to the HB formed between the dATP sugar and D760 from motif C, further stabilization of dATP comes from an HB formed between the dATP sugar and 3'-end primer, and another HB formed between the 3' end primer and motif-C S759 (**Figure 7B**). Hence, it appears that the 3'-end primer plays an important role in stabilizing dATP with a network of HB interactions upon initial binding or say, *off-path*.

Although the non-cognate dATP initial binding stability appears perplexing, prior crystal structure studies on the structurally similar PV RdRp have used dCTP to stall and resolve the RdRp structure in the open state<sup>33</sup>, supporting a stable binding configuration of dNTP binding to the viral RdRp. In addition, experimental work on the nsp14 exonuclease enzyme of SARS-CoV-2 has shown that for excision of an incorrect nt, the nt needs to have a proper RNA sugar<sup>88</sup>, indicating the 2' OH group of the RNA ribose as a critical component for RdRp template recognition and elongation. Furthermore, recent experimental studies have shown that an elongation complex soaked in solution with

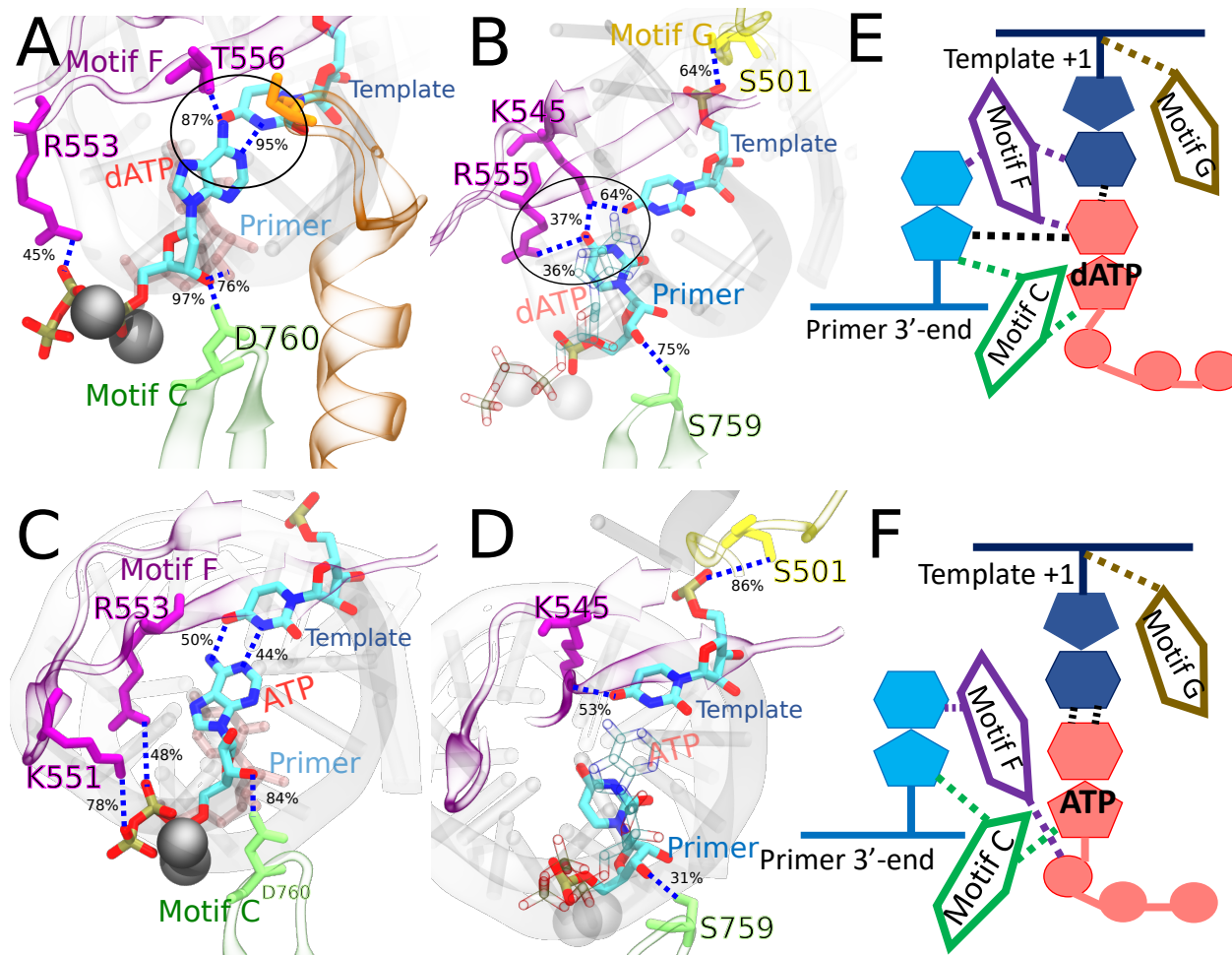


Fig. 7 Comparing the key interactions that stabilize the non-cognate dATP (and surroundings) and cognate ATP in the initial binding system. The conserved protein motifs are shown in cartoon representation while interacting residues from these motifs are shown in licorice using the same color. A/C: Incoming dATP/ATP and template nucleotide are shown in licorice colored by atom name. Hydrogen bonds (HBs) formed between dATP (or ATP) and template nucleotide uracil/protein/3'-end RNA primer. Black circle highlights the strongest interactions involved in 'trapping' non-cognate dATP in the initially bound state. B/D: Incoming dATP/ATP is shown in transparent representation for clarity, 3'-end primer and template nucleotide are shown in licorice colored by atom name. HBs formed between the protein and template nucleotide uracil / 3'-end RNA primer for stabilization, in the absence of proper WC base pairing in dATP (B). HBs formed between the protein and template nucleotide uracil / 3'-end RNA primer for stabilization, with proper WC base pairing in ATP (D). For detailed HB occupancy plots see SI Figure S14. E: Schematic and cartoon of key motifs stabilizing dATP. F: Schematic and cartoon of key motifs stabilizing ATP. For stereoscopic views of A/B see SI Figure S16.

dNTPs had no catalytic activity<sup>89</sup>. Michaelis–Menten kinetics also showed the SARS-CoV-2 RdRp selectivity of ATP over dATP is  $\sim 1000$  ( $\frac{V_{max}}{K_m} = 0.02$  in dATP and 23 in ATP)<sup>90</sup>. Similar trends were observed in PV *in vitro* biochemical studies with  $\sim 117$  discrimination factor ( $\frac{k_{pol}}{K_m}$ )<sub>CTP</sub> / ( $\frac{k_{pol}}{K_m}$ )<sub>dCTP</sub><sup>91</sup>. Other computational works also tested the design of inhibitors with ribose sugar modifications or removal of the OH function groups entirely<sup>40,41</sup>.

In general, the nucleotide selection in accompany can be achieved by initial NTP-template base pairing and then the ribose recognition. A cluster of residues responsible for the ribose hydroxyl recognition have been suggested also for the PV RdRp, in which motif A and D move toward the active site upon closing to assist nucleotide insertion and selection<sup>33</sup>. It was also found that mutations inhibiting the open-closed transition likely improve the fidelity control<sup>16,17</sup>. In hepatitis C virus (HCV), the RdRp or NS5B has its motif B in addition to motif A for positioning ri-

bose of incoming NTP, using an extensive hydrogen bonding network from motif A and B to recognize 2'-hydroxyl of the incoming nucleotide<sup>92</sup>. NS5B has been suggested to contain a nucleotide mediated excision mechanism in order to excise the mismatched nucleotide at the 30-end of the nascent RNA *in vitro*<sup>93,94</sup>. Interestingly, motif B involved in translocation of the HCV RdRp is also linked to fidelity control, as mutations in motif B could not only slow down translocation, but also increase the time of residency of a mismatch in the active site to allow it to be excised<sup>93</sup>. Motif G may also affect the NTP recognition and selection as it influences the orientation of the template strand/nucleotide or nt<sup>93</sup>. Motif F can be linked to fidelity control as well since it guides the nucleotide into the tunnel as well as to direct the pyrophosphate out of the tunnel after catalysis, which can be rate limiting in NS5B to improve the fidelity<sup>95</sup>. In **SI Table S5**, we provide a summary of motifs A-G and sequence alignments for various

RdRps in comparison (from SARS-CoV-2, PV, EV, and HCV).

## 5 Conclusion

To conclude, we have employed all-atom MD simulations in equilibrium ensembles and umbrella sampling methods to demonstrate intrinsic nucleotide selectivity of SARSCoV-2 RdRp, comparing non-cognate dATP/GTP with cognate ATP/RDV-TP analog on structural dynamics and free energetics from binding to insertion to the RdRp active site. Our studies show that the non-cognate dATP/GTP is well stabilized or trapped upon initial binding to certain off-path configurations, as the highly conserved structural motifs F/G/A/C of the viral RdRp form HBs with the ncNTP, RNA template nt, and/or 3'-end RNA primer. Intrinsically, it is not the polymerase enzyme that determines a right/cognate or wrong/non-cognate nucleotide substrate in the template-based polymerization. The cognate or non-cognate NTP species are determined by the RNA template nt primarily via the WC base pairing, and additionally by the 3'-end primer via base stacking etc. With incoming ncNTP that is incapable of stabilizing the template counterpart or the 3'-end primer, the RdRp structural motifs sense such instability, and then takes over to select against the ncNTP. Presumably, ncNTP can be rejected in case of low binding affinity, i.e., upon initial binding to RdRp. However, it appears that in SARS-CoV-2 RdRp, the ncNTP can be particularly stabilized, say *off-path*, upon initial binding to certain configurations, so that to be prevented or inhibited from further insertion to the active site. Such mechanism of nucleotide selection seems to be well supported by the two-step binding and insertion processes, *pre-chemically*, in the single-subunit viral polymerase enzymes. Partial off-path initial binding and inhibition for insertion of ncNTPs were also suggested computationally for single-subunit T7 RNAP<sup>57,96</sup>. Such nucleotide substrate selection mechanisms on trapping non-cognate species upon initial binding to prevent further incorporation would particularly draw attention from future antiviral drug design, so that to develop drug compounds that are capable of escaping from such trapping mechanism by the viral RdRp.

Technically, we have employed the umbrella sampling simulations to characterize the slow NTP insertion dynamics that is accompanied by the open to closed conformational changes around the RdRp active site. As the slow pre-chemical conformational transition likely takes place over milliseconds, it becomes indispensable to enhance computational sampling while limiting artifacts to be introduced. In the simulation, we well manipulated collective atomic coordinates from all structural motifs along with key players of NTP/template. Nevertheless, how to identify the most essential coordinates is a continual challenging issue<sup>97,98</sup>. Additionally, accelerated simulations<sup>99</sup> may be conducted to sample a sufficient amount of conformational transitions, i.e., from open to closed. For multiple NTP species incorporation, enhanced computational sampling would become even more demanding, considering that multiple reaction paths exist<sup>100</sup>. Further exploration and validation of our current studies would require substantial experimental studies. Recent high-resolution characterizations of SARS-CoV-2 RdRp complexes have brought about additionally the post translocation and pre-catalytic states for im-

proved modeling<sup>62,63</sup>. Resolving high-resolution structures of the SARS-CoV-2 RdRp complexes with the stabilized *off-path* initial binding configurations of non-cognate dATP or GTP, as being proposed in current computational work, would be highly anticipated.

## 6 ACKNOWLEDGEMENTS

Part of this work was supported by National Science Foundation (NSF) Award #2028935. This research used resources of the Oak Ridge Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC05-00OR22725. Additional computational resources were provided by UCI HPC3 - High Performance Community Computing Cluster. JY has also been supported by the CMCF of UCI via NSF DMS 1763272 and the Simons Foundation grant #594598 and start-up funding from UCI.

## 7 Data Availability

Input files available with provided zip. Trajectories available on request.

## Notes and references

- 1 W. T. Harvey, A. M. Carabelli, B. Jackson, R. K. Gupta, E. C. Thomson, E. M. Harrison, C. Ludden, R. Reeve, A. Rambaut, S. J. Peacock and D. L. Robertson, *Nature Reviews Microbiology* 2021 19:7, 2021, **19**, 409–424.
- 2 Y. Araf, F. Akter, Y. d. Tang, R. Fatemi, M. S. A. Parvez, C. Zheng and M. G. Hossain, *Journal of Medical Virology*, 2022, **94**, 1825–1832.
- 3 P. Pandey, J. S. Rane, A. Chatterjee, A. Kumar, R. Khan, A. Prakash and S. Ray, *Journal of Biomolecular Structure and Dynamics*, 2020, **39**, 1–11.
- 4 A. C. Papageorgiou and I. Mohsin, *Cells* 2020, Vol. 9, Page 2343, 2020, **9**, 2343.
- 5 A. M. Almehti, G. Khoder, A. S. Alchakee, A. T. Alsayyid, N. H. Sarg and S. S. Soliman, *Infection*, 2021, **49**, 855–876.
- 6 R. Singh, V. K. Bhardwaj, J. Sharma, D. Kumar and R. Purohit, *Computers in Biology and Medicine*, 2021, **136**, 104631.
- 7 S. Ullrich and C. Nitsche, *Bioorganic and Medicinal Chemistry Letters*, 2020, **30**, 127377.
- 8 Z. Jin, H. Wang, Y. Duan and H. Yang, *Biochemical and Biophysical Research Communications*, 2021, **538**, 63–71.
- 9 H. Tarannum, K. Rashmi and S. Nandi, *Current Drug Targets*, 2021, **23**, 802–817.
- 10 V. K. Bhardwaj, R. Singh, J. Sharma, V. Rajendran, R. Purohit and S. Kumar, *Journal of Biomolecular Structure & Dynamics*, 2020, 1–10.
- 11 N. Magazine, T. Zhang, Y. Wu, M. C. McGee, G. Veggiani and W. Huang, *Viruses* 2022, Vol. 14, Page 640, 2022, **14**, 640.
- 12 V. Papanikolaou, A. Chrysovergis, V. Ragos, E. Tsiambas, S. Katsinis, A. Manoli, S. Papouliakos, D. Roukas, S. Mastronikolis, D. Peschos, A. Batistatou, E. Kyrodimos and N. Mastronikolis, *Gene*, 2022, **814**, 146134.
- 13 J. A. Mótyán, M. Mahdi, G. Hoffka and J. Tózsér, *Interna-*



- tional Journal of Molecular Sciences* 2022, Vol. 23, Page 3507, 2022, **23**, 3507.
- 14 S. Iketani, H. Mohri, B. Culbertson, S. J. Hong, Y. Duan, M. I. Luck, M. K. Annavajhala, Y. Guo, Z. Sheng, A. C. Uhlemann, S. P. Goff, Y. Sabo, H. Yang, A. Chavez and D. D. Ho, *Nature* 2022 613:7944, 2022, **613**, 558–564.
  - 15 Y. Hu, E. M. Lewandowski, H. Tan, X. Zhang, R. T. Morgan, X. Zhang, L. M. C. Jacobs, S. G. Butler, M. V. Gongora, J. Choy, X. Deng, Y. Chen and J. Wang, *bioRxiv*, 2022, 2022.06.28.497978.
  - 16 D. G. Ahn, J. K. Choi, D. R. Taylor and J. W. Oh, *Archives of Virology*, 2012, **157**, 2095–2104.
  - 17 M. Sevajol, L. Subissi, E. Decroly, B. Canard and I. Imbert, *Virus Research*, 2014, **194**, 90–99.
  - 18 W. Zhu, C. Z. Chen, K. Gorshkov, M. Xu, D. C. Lo and W. Zheng, *SLAS Discovery*, 2020, **25**, 1141–1151.
  - 19 P. Gong, *Frontiers in Molecular Biosciences*, 2022, **8**, 822218.
  - 20 C. Long, M. E. Romero, D. La Rocco and J. Yu, *Computational and Structural Biotechnology Journal*, 2021, **19**, 3339–3348.
  - 21 L. P. Jordheim, D. Durantel, F. Zoulim and C. Dumontet, *Nature Reviews Drug Discovery*, 2013, **12**, 447–464.
  - 22 M. Gaetani, P. Sabatier, A. A. Saei, C. M. Beusch, Z. Yang, S. L. Lundström and R. A. Zubarev, *Journal of Proteome Research*, 2019, **18**, 4027–4037.
  - 23 Y. Gao, L. Yan, Y. Huang, F. Liu, Y. Zhao, L. Cao, T. Wang, Q. Sun, Z. Ming, L. Zhang, J. Ge, L. Zheng, Y. Zhang, H. Wang, Y. Zhu, C. Zhu, T. Hu, T. Hua, B. Zhang, X. Yang, J. Li, H. Yang, Z. Liu, W. Xu, L. W. Guddat, Q. Wang, Z. Lou and Z. Rao, *Science*, 2020, **368**, 779–782.
  - 24 W. Yin, C. Mao, X. Luan, D.-D. Shen, Q. Shen, H. Su, X. Wang, F. Zhou, W. Zhao, M. Gao, S. Chang, Y.-C. Xie, G. Tian, H.-W. Jiang, S.-C. Tao, J. Shen, Y. Jiang, H. Jiang, Y. Xu, S. Zhang, Y. Zhang and H. E. Xu, *Science*, 2020, **368**, 1499–1504.
  - 25 H. S. Hillen, G. Kocic, L. Farnung, C. Dienemann, D. Tegunov and P. Cramer, *Nature*, 2020, **584**, 154–156.
  - 26 J. Chen, B. Malone, E. Llewellyn, M. Grasso, P. M. Shelton, P. D. B. Olinares, K. Maruthi, E. T. Eng, H. Vatandaslar, B. T. Chait, T. M. Kapoor, S. A. Darst and E. A. Campbell, *Cell*, 2020, **182**, 1560–1573.
  - 27 Q. Wang, J. Wu, H. Wang, Y. Gao, Q. Liu, A. Mu, W. Ji, L. Yan, Y. Zhu, C. Zhu, X. Fang, X. Yang, Y. Huang, H. Gao, F. Liu, J. Ge, Q. Sun, X. Yang, W. Xu, Z. Liu, H. Yang, Z. Lou, B. Jiang, L. W. Guddat, P. Gong and Z. Rao, *Cell*, 2020, **182**, 417–428.
  - 28 B. Malone, J. Chen, Q. Wang, E. Llewellyn, Y. J. Choi, P. D. B. Olinares, X. Cao, C. Hernandez, E. T. Eng, B. T. Chait, D. E. Shaw, R. Landick, S. A. Darst and E. A. Campbell, *Proceedings of the National Academy of Sciences of the United States of America*, 2021, **118**, 2102516118.
  - 29 J. P. Bravo, T. L. Dangerfield, D. W. Taylor and K. A. Johnson, *Molecular Cell*, 2021, **81**, 1548–1552.
  - 30 Q. Peng, R. Peng, B. Yuan, M. Wang, J. Zhao, L. Fu, J. Qi and Y. Shi, *Innovation*, 2021, **2**, 100080.
  - 31 K. Naydenova, K. W. Muir, L. F. Wu, Z. Zhang, F. Coscia, M. J. Peet, P. Castro-Hartmann, P. Qian, K. Sader, K. Dent, D. Kimanius, J. D. Sutherland, J. Löwe, D. Barford and C. J. Russo, *Proceedings of the National Academy of Sciences of the United States of America*, 2021, **118**, e2021946118.
  - 32 B. F. Malone, J. K. Perry, P. D. B. Olinares, H. W. Lee, J. Chen, T. C. Appleby, J. Y. Feng, J. P. Bilello, H. Ng, J. Sotiris, M. Ebrahim, E. Y. Chua, J. H. Mendez, E. T. Eng, R. Landick, M. Götte, B. T. Chait, E. A. Campbell and S. A. Darst, *Nature*, 2023, **614**, 781–787.
  - 33 P. Gong and O. B. Peersen, *Proceedings of the National Academy of Sciences of the United States of America*, 2010, **107**, 22505–22510.
  - 34 B. Shu and P. Gong, *Proceedings of the National Academy of Sciences*, 2016, **113**, E4005–E4014.
  - 35 J. H. Beigel, K. M. Tomashek, L. E. Dodd, A. K. Mehta, B. S. Zingman, A. C. Kalil, E. Hohmann, H. Y. Chu, A. Luetkemeyer, S. Kline, D. Lopez de Castilla, R. W. Finberg, K. Dierberg, V. Tapson, L. Hsieh, T. F. Patterson, R. Paredes, D. A. Sweeney, W. R. Short, G. Touloumi, D. C. Lye, N. Ohmagari, M.-d. Oh, G. M. Ruiz-Palacios, T. Benfield, G. Fätkenheuer, M. G. Kortepeter, R. L. Atmar, C. B. Creech, J. Lundgren, A. G. Babiker, S. Pett, J. D. Neaton, T. H. Burgess, T. Bonnett, M. Green, M. Makowski, A. Osinusi, S. Nayak and H. C. Lane, *New England Journal of Medicine*, 2020, **383**, 1813–1826.
  - 36 T. K. Warren, R. Jordan, M. K. Lo, A. S. Ray, R. L. Mackman, V. Soloveva, D. Siegel, M. Perron, R. Bannister, H. C. Hui, N. Larson, R. Strickley, J. Wells, K. S. Stuthman, S. A. Van Tongeren, N. L. Garza, G. Donnelly, A. C. Shurtleff, C. J. Retterer, D. Gharaibeh, R. Zamani, T. Kenny, B. P. Eaton, E. Grimes, L. S. Welch, L. Gomba, C. L. Wilhelmsen, D. K. Nichols, J. E. Nuss, E. R. Nagle, J. R. Kugelman, G. Palacios, E. Doerffler, S. Neville, E. Carra, M. O. Clarke, L. Zhang, W. Lew, B. Ross, Q. Wang, K. Chun, L. Wolfe, D. Babusis, Y. Park, K. M. Stray, I. Trancheva, J. Y. Feng, O. Barauskas, Y. Xu, P. Wong, M. R. Braun, M. Flint, L. K. McMullan, S. S. Chen, R. Fearn, S. Swaminathan, D. L. Mayers, C. F. Spiropoulou, W. A. Lee, S. T. Nichol, T. Cihlar and S. Bavari, *Nature*, 2016, **531**, 381–385.
  - 37 S. Ahmed, R. Mahtarin, S. Islam, A. Das, S. Al Mamun, A. Samina, A. Ali, S. Das, A. Al Mamun and S. Ahmed, *Journal of Biomolecular Structure and Dynamics*, 2022, **40**, 11111–11124.
  - 38 S. Koulgi, V. Jani, V. N. Mallikarjunachari Uppuladinne, U. Sonavane and R. Joshi, *Journal of Biomolecular Structure and Dynamics*, 2022, **40**, 7230–7244.
  - 39 R. Singh, V. K. Bhardwaj and R. Purohit, *Computers in Biology and Medicine*, 2021, **139**, 104965.
  - 40 A. Parise, G. Ciardullo, M. Prejanò, A. De La Lande and T. Marino, *Journal of Chemical Information and Modeling*, 2022, **62**, 4916–4927.
  - 41 Y. Li, D. Zhang, X. Gao, X. Wang and L. Zhang, *Journal of Physical Chemistry Letters*, 2022, **13**, 4111–4118.
  - 42 M. Giannetti, C. Mazzuca, G. Ripani and A. Palleschi,

- Molecules*, 2023, **28**, 191.
- 43 S. C. Bera, M. Seifert, R. N. Kirchdoerfer, P. van Nies, Y. Wubulikasimu, S. Quack, F. S. Papini, J. J. Arnold, B. Carnard, C. E. Cameron, M. Depken and D. Dulin, *Cell Reports*, 2021, **36**, 109650.
- 44 X. Luo, T. Xu, X. Gao and L. Zhang, *Chinese Journal of Chemical Physics*, 2022, **35**, 407–412.
- 45 Y. Yuzenkova, A. Bochkareva, V. R. Tadigotla, M. Roghanian, S. Zorov, K. Severinov and N. Zenkin, *BMC Biology*, 2010, **8**, 54.
- 46 J. Yu, *Computational and Mathematical Biophysics*, 2014, **2**, 141–160.
- 47 K. A. Johnson, *Annual Review of Biochemistry*, 1993, **62**, 685–713.
- 48 J. Huang, L. G. Brieba and R. Sousa, *Biochemistry*, 2000, **39**, 11571–11580.
- 49 T. Schlick, K. Arora, W. A. Beard and S. H. Wilson, *Theoretical Chemistry Accounts*, 2012, **131**, 1287.
- 50 V. S. Anand and S. S. Patel, *Journal of Biological Chemistry*, 2006, **281**, 35677–35685.
- 51 C. Long and J. Yu, *Entropy*, 2018, **20**, 306.
- 52 M. E. Romero, C. Long, D. La Rocco, A. M. Keerthi, D. Xu and J. Yu, *Molecular Systems Design and Engineering*, 2021, **6**, 888–902.
- 53 J. J. Hopfield, *Proceedings of the National Academy of Sciences of the United States of America*, 1974, **71**, 4135–4139.
- 54 J. Ninio, *Biochimie*, 1975, **57**, 587–595.
- 55 J. L. Klepeis, K. Lindorff-Larsen, R. O. Dror and D. E. Shaw, *Current Opinion in Structural Biology*, 2009, **19**, 120–127.
- 56 R. Lazim, D. Suh and S. Choi, *International Journal of Molecular Sciences*, 2020, **21**, 1–20.
- 57 C. Long, E. Chao, L. T. Da and J. Yu, *Nucleic Acids Research*, 2019, **47**, 4721–4735.
- 58 G. M. Torrie and J. P. Valleau, *Journal of Computational Physics*, 1977, **23**, 187–199.
- 59 J. Kästner, *WIREs Computational Molecular Science*, 2011, **1**, 932–942.
- 60 R. Kierzek, M. E. Burkard and D. H. Turner, *Biochemistry*, 1999, **38**, 14214–14223.
- 61 L. Zhang, D. Zhang, X. Wang, C. Yuan, Y. Li, X. Jia, X. Gao, H. L. Yen, P. P. H. Cheung and X. Huang, *Physical Chemistry Chemical Physics*, 2021, **23**, 5852–5863.
- 62 Q. Peng, R. Peng, B. Yuan, M. Wang, J. Zhao, L. Fu, J. Qi and Y. Shi, *The Innovation*, 2021, **2**, 100080.
- 63 B. F. Malone, J. K. Perry, P. D. B. Olinares, H. W. Lee, J. Chen, T. C. Appleby, J. Y. Feng, J. P. Bilello, H. Ng, J. Sotiris, M. Ebrahim, E. Y. D. Chua, J. H. Mendez, E. T. Eng, R. Landick, M. Götte, B. T. Chait, E. A. Campbell and S. A. Darst, *Nature*, 2023, **614**, 781–787.
- 64 M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindah, *SoftwareX*, 2015, **1-2**, 19–25.
- 65 J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser and C. Simmerling, *Journal of Chemical Theory and Computation*, 2015, **11**, 3696–3713.
- 66 I. Ivani, P. D. Dans, A. Noy, A. Pérez, I. Faustino, A. Hospital, J. Walther, P. Andrio, R. Goñi, A. Balaceanu, G. Portella, F. Battistini, J. L. Gelpí, C. González, M. Vendruscolo, C. A. Laughton, S. A. Harris, D. A. Case and M. Orozco, *Nature Methods*, 2015, **13**, 55–58.
- 67 K. L. Meagher, L. T. Redman and H. A. Carlson, *Journal of Computational Chemistry*, 2003, **24**, 1016–1025.
- 68 M. Bonomi, G. Bussi, C. Camilloni, G. A. Tribello, P. Banáš, A. Barducci, M. Bernetti, P. G. Bolhuis, S. Bottaro, D. Branduardi, R. Capelli, P. Carloni, M. Ceriotti, A. Cesari, H. Chen, W. Chen, F. Colizzi, S. De, M. De La Pierre, D. Donadio, V. Drobot, B. Ensing, A. L. Ferguson, M. Filizola, J. S. Fraser, H. Fu, P. Gasparotto, F. L. Gervasio, F. Giberti, A. Gil-Ley, T. Giorgino, G. T. Heller, G. M. Hocky, M. Iannuzzi, M. Invernizzi, K. E. Jelfs, A. Jussupow, E. Kirilin, A. Laio, V. Limongelli, K. Lindorff-Larsen, T. Löhner, F. Marinelli, L. Martinsamos, M. Masetti, R. Meyer, A. Michaelides, C. Molteni, T. Morishita, M. Nava, C. Paissoni, E. Papaleo, M. Parrinello, J. Pfendtner, P. Piaggi, G. M. Piccini, A. Pietropaolo, F. Pietrucci, S. Pipolo, D. Provasi, D. Quigley, P. Raiteri, S. Raniolo, J. Rydzewski, M. Salvalaglio, G. C. Sosso, V. Spivok, J. Šponer, D. W. Swenson, P. Tiwary, O. Valsson, M. Vendruscolo, G. A. Voth and A. White, *Nature Methods*, 2019, **16**, 670–673.
- 69 D. J. Price and C. L. Brooks, *Journal of Chemical Physics*, 2004, **121**, 10096–10103.
- 70 U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *The Journal of Chemical Physics*, 1995, **103**, 8577–8593.
- 71 H. J. Berendsen, J. P. Postma, W. F. Van Gunsteren, A. Dinola and J. R. Haak, *The Journal of Chemical Physics*, 1984, **81**, 3684–3690.
- 72 M. Parrinello and A. Rahman, *Journal of Applied Physics*, 1981, **52**, 7182–7190.
- 73 G. M. Torrie and J. P. Valleau, *Journal of Computational Physics*, 1977, **23**, 187–199.
- 74 J. Kästner, *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2011, **1**, 932–942.
- 75 J. Schlitter, M. Engels and P. Krüger, *Journal of Molecular Graphics*, 1994, **12**, 84–89.
- 76 S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen and P. A. Kollman, *Journal of Computational Chemistry*, 1992, **13**, 1011–1021.
- 77 A. Grossfield, *WHAM - Grossfield Lab*, [http://membrane.urmc.rochester.edu/?page\\_id=126](http://membrane.urmc.rochester.edu/?page_id=126).
- 78 B. Efron and R. Tibshirani, *An Introduction to the Bootstrap*, Chapman and Hall/CRC, 1994.
- 79 N. Michaud-Agrawal, E. J. Denning, T. B. Woolf and O. Beckstein, *Journal of Computational Chemistry*, 2011, **32**, 2319–2327.
- 80 M. L. Waskom, *Journal of Open Source Software*, 2021, **6**, 3021.
- 81 J. D. Hunter, *Computing in Science and Engineering*, 2007, **9**, 90–95.

- 82 Chunhong Long, M. Ernesto Romero, Liqiang Dai and Jin Yu, *Physical Chemistry Chemical Physics*, 2023, **25**, 13508–13520.
- 83 O. Poch, I. Sauvaget, M. Delarue and N. Tordo, *The EMBO Journal*, 1989, **8**, 3867–3874.
- 84 J. Shi, J. M. Perryman, X. Yang, X. Liu, D. M. Musser, A. K. Boehr, I. M. Moustafa, J. J. Arnold, C. E. Cameron and D. D. Boehr, *Biochemistry*, 2019, **58**, 3735–3743.
- 85 H. Alonso, A. A. Bliznyuk and J. E. Gready, *Medicinal Research Reviews*, 2006, **26**, 531–568.
- 86 J. Schimunek, P. Seidl, K. Elez, T. Hempel, T. Le, F. Noé, S. Olsson, L. Raich, R. Winter, H. Gokcan, F. Gusev, E. M. Gutkin, O. Isayev, M. G. Kurnikova, C. H. Narangoda, R. Zubatyuk, I. P. Bosko, K. V. Furs, A. D. Karpenko, Y. V. Kornoushenko, M. Shuldau, A. Yushkevich, M. B. Benabderrahmane, P. Bousquet-Melou, R. Bureau, B. Charton, B. C. Cirou, G. Gil, W. J. Allen, S. Sirimulla, S. Watowich, N. A. Antonopoulos, N. E. Epitropakis, A. K. Krasoulis, V. P. Pitsikalis, S. T. Theodorakis, I. Kozlovskii, A. Maliutin, A. Medvedev, P. Popov, M. Zaretskii, H. Eghbal-Zadeh, C. Halmich, S. Hochreiter, A. Mayr, P. Ruch, M. Widrich, F. Berenger, A. Kumar, Y. Yamanishi, K. Y. J. Zhang, E. Bengio, Y. Bengio, M. J. Jain, M. Korablyov, C.-H. Liu, G. Marcou, E. Glaab, K. Barnsley, S. M. Iyengar, M. J. Ondrechen, V. J. Haupt, F. Kaiser, M. Schroeder, L. Pugliese, S. Albani, C. Athanasiou, A. Beccari, P. Carloni, G. D'Arrigo, E. Gianquinto, J. Goßen, A. Hanke, B. P. Joseph, D. B. Kokh, S. Kovachka, C. Manelfi, G. Mukherjee, A. Muñoz-Chicharro, F. Musiani, A. Nunes-Alves, G. Paiardi, G. Rossetti, S. K. Sadiq, F. Spyraakis, C. Talarico, A. Tsengenes, R. C. Wade, C. Copeland, J. Gaiser, D. R. Olson, A. Roy, V. Venkatraman, T. J. Wheeler, H. Arthanari, K. Blaschitz, M. Cespugli, V. Durmaz, K. Fackeldey, P. D. Fischer, C. Gorgulla, C. Gruber, K. Gruber, M. Hetmann, J. E. Kinney, K. M. Padmanabha Das, S. Pandita, A. Singh, G. Steinkellner, G. Tesseyre, G. Wagner, Z.-F. Wang, R. J. Yust, D. S. Druzhilovskiy, D. A. Filimonov, P. V. Pogodin, V. Poroikov, A. V. Rudik, L. A. Stolbov, A. V. Veselovsky, M. De Rosa, G. De Simone, M. R. Gullotta, J. Lombino, N. Mekni, U. Perricone, A. Casini, A. Embree, D. B. Gordon, D. Lei, K. Pratt, C. A. Voigt, K.-Y. Chen, Y. Jacob, T. Krischuns, P. Lafaye, A. Zettor, M. L. Rodríguez, K. M. White, D. Fearon, F. Von Delft, M. A. Walsh, D. Horvath, C. L. Brooks III, B. Falsafi, B. Ford, A. García-Sastre, S. Yup Lee, N. Naffakh, A. Varnek, G. Klambauer and T. M. Hermans, *Molecular Informatics*, n/a, e202300262.
- 87 X. Yang, X. Liu, D. M. Musser, I. M. Moustafa, J. J. Arnold, C. E. Cameron and D. D. Boehr, *Journal of Biological Chemistry*, 2017, **292**, 3810–3826.
- 88 A. N. Jones, A. Mourão, A. Czarna, A. Matsuda, R. Fino, K. Pyrc, M. Sattler and G. M. Popowicz, *Scientific Reports*, 2022, **12**, 9593.
- 89 I. Petushkov, D. Esyunina and A. Kulbachinskiy, *FEBS Journal*, 2023, **290**, 80–92.
- 90 C. J. Gordon, E. P. Tchesnokov, E. Woolner, J. K. Perry, J. Y. Feng, D. P. Porter and M. Götte, *Journal of Biological Chemistry*, 2020, **295**, 6785–6797.
- 91 G. Campagnola, S. McDonald, S. Beaucourt, M. Vignuzzi and O. B. Peersen, *Journal of Virology*, 2015, **89**, 275–286.
- 92 T. C. Appleby, J. K. Perry, E. Murakami, O. Barauskas, J. Feng, A. Cho, D. Fox, D. R. Wetmore, M. E. McGrath, A. S. Ray, M. J. Sofia, S. Swaminathan and T. E. Edwards, *Science*, 2015, **347**, 771–775.
- 93 B. Selisko, N. Papageorgiou, F. Ferron and B. Canard, *Viruses*, 2018, **10**, 59.
- 94 Z. Jin, V. Leveque, H. Ma, K. A. Johnson and K. Klumpp, *Proceedings of the National Academy of Sciences of the United States of America*, 2013, **110**, E348–E357.
- 95 B. Villalba and K. A. Johnson, *The Journal of Biological Chemistry*, 2020, **295**, 16436–16444.
- 96 C. Long, E. Chao, L. T. Da and J. Yu, *Computational and Structural Biotechnology Journal*, 2019, **17**, 638–644.
- 97 S. Bhakat, *RSC Advances*, 2022, **12**, 25010–25024.
- 98 L. Bonati, E. Trizio, A. Rizzi and M. Parrinello, *bioRxiv*, 2023.
- 99 J. Wang, P. R. Arantes, A. Bhattarai, R. V. Hsu, S. Pawnikar, Y. M. Huang, G. Palermo and Y. Miao, *WIREs Computational Molecular Science*, 2021, **11**, e1521.
- 100 L. T. Chong, A. S. Saglam and D. M. Zuckerman, *Path-sampling strategies for simulating rare events in biomolecular systems*, 2017.