Probing remdesivir nucleotide analogue insertion to SARS-CoV-2 RNA dependent RNA polymerase in viral replication

Remdesivir is an antiviral compound designed as a nucleotide analogue to adenosine to target on the viral RNA dependent RNA polymerase (RdRp). In its active form, remdesivir triphosphate (RDV-TP) differs from ATP by only the addition of nitrile functional group to the sugar, and swap between 3 nitrogen/carbon in the adenine base. Nevertheless, the small modifications can lead to termination of RNA synthesis function of the viral RdRp. In this work we explore the SARS-CoV-2 RdRp structural dynamics upon the initial binding of RDV-TP (and ATP) to insertion, prior to catalysis. Our results show that upon initial binding, RDV-TP can form base stacking instead of base pairing with the template nucleotide, allowing it to take advantages of thermal fluctuations to 'squeeze' into the active site of the viral RdRp. In comparison, ATP binding and insertion are supported by traditional base pairing with the template. Our work also reveals delicate interactions between the incoming nucleotide (base,sugar, and triphosphate), in coordination with the template nucleotide, and key residues in the RdRp active site. These interactions and coordinations can be explored further with additional nucleotide analogues in comparison, and be considered for designing improved nucleotide analogue drugs with enhanced efficacy.

Journal Name

Probing remdesivir nucleotide analogue insertion to SARS-CoV-2 RNA dependent RNA polymerase in viral replication[†]

Moises Ernesto Romero,[a] Chunhong Long,[b] Daniel La Rocco,[c] Anusha Mysore Keerthi,[d] Dajun Xu,[e] and Jin Yu[*f]

Remdesivir (RDV) prodrug can be metabolized into a triphosphate form nucleotide analogue (RDV-TP) to bind and insert into the active site of viral RNA dependent RNA polymerase (RdRp) to further interfere with the viral genome replication. In this work, we computationally studied how RDV-TP binds and inserts to the SARS-CoV-2 RdRp active site, in comparison with natural nucleotide substrate adenosine triphosphate (ATP). To do that, we first constructed atomic structural models of an initial binding complex (active site open) and a substrate insertion complex (active site closed), based on high-resolution cryo-EM structures determined recently for SARS-CoV-2 RdRp or non-structural protein (nsp) 12, in complex with accessory protein factors nsp7 and nsp8. By conducting all-atom molecular dynamics simulation with umbrella sampling strategies on the nucleotide insertion between the open and closed state RdRp complexes, our studies show that RDV-TP can initially bind comparatively stabilized to the viral RdRp active site, as it primarily forms base stacking with the template Uracil nucleotide (nt +1), which under freely fluctuations supports a low free energy barrier of the RDV-TP insertion (~ 1.5 kcal/mol). In comparison, the corresponding natural substrate ATP binds initially to the RdRp active site in Watson-Crick base pairing with the template nt, and inserts into the active site with a medium low free energy barrier (~ 2.6 kcal/mol), when the fluctuations of the template nt are well quenched. The simulations also show that the initial base stacking of RDV-TP with the template can be specifically stabilized by motif C-S759, S682 (near motif B) with the base, and motif G-K500 with the template backbone. Although the RDV-TP insertion can be hindered by motif F-R555/R553 interaction with the triphosphate, the ATP insertion seems to be facilitated by such interactions. The inserted RDV-TP and ATP can be further distinguished by specific sugar interaction with motif B-T687 and motif A-D623, respectively.

1 Introduction

RNA dependent RNA polymerase (RdRp) is the core protein engine responsible for synthesizing genome in the replication/transcription machinery of RNA viruses, which represent a large class of human and animal pathogens to cause disease and pandemics[1,2]. Based on the template RNA strand, RdRp selec-tively recruits ribonucleotides one at a time to the active site and adds the nucleotide to the growing RNA chain upon catalyzing the phosphoryl-transfer reaction, which is then followed by product (pyrophosphate) release and the polymerase translocation. Due to its critical role in the viral RNA synthesis and highly conserved core structure, the viral RdRp serves a highly promising antiviral drug target for both nucleotide analogue and non-nucleoside inhibitors[3]. Remdesivir (or RDV), the only US-FDA proved drug (named VEKLURY) so far treating COVID-19[4], works as a prodrug that is metabolized into a nucleotide analogue to compete with natural nucleotide substrates of RdRp to be incorporated into viral RNA gnome to further terminate the RNA synthesis[5,6]. As a broad-spectrum anti-viral compound, RDV was developed originally for treatments of Ebola virus disease (EVD)[7], and then applied for infections of middle east and severe accurate respiratory syndrome coronavirus (MERS-CoV and SARS-CoV)[8], which are both close relatives to the currently emerged novel coronavirus

[a] Department of Chemistry, University of California, Irvine, CA 92697 USA

[b] School of Science, Chongqing University of Posts and Telecommunications, Chongqing, 400065, China

[c] Department of Physics, University of California, Berkeley, CA 94720 USA

[d] Department of Biotechnology, R.V College of Engineering, Bengaluru, 560059, India

[e] Department of Mathematics, University of California, Irvine, CA 92697 USA

[*f] Department of Physics and Astronomy, Department of Chemistry, NSF-Simons Center for Multiscale Cell Fate Research, University of California, Irvine, CA 92697, USA

*Email: jin.yu@uci.edu

† Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 00.0000/00000000.

(SARS-CoV-2) causing COVID-19. Recent in-vitro and in-vivo studies on RDV impacts to the viral RdRp function have confirmed the RDV analogue incorporation and inhibition during the viral RdRp replication, in particular, in SARS-CoV-2[9–12]. The existing evidences have consistently suggested that the active triphosphate form of RDV (RDV-TP) binds competitively with the natural substrate, i.e., adenosine triphosphate (or ATP), to the viral RdRp and the incorporation leads to a delayed chain termination[10]. Such an analogue incorporation and consequent chain termination indicate that RDV-TP can successfully evade from both nucleotide selectivity of the viral RdRp as well as the proofreading function from ExoN in coordination with RdRp in the coronavirus replication[5,13].

Nucleotide selectivity of the RdRp or polymerases in general serves as a primary fidelity control method in corresponding gene transcription or replication, i.e., during the template-based polymerase elongation[14–16]. The selectivity indeed proceeds throughout a full nucleotide addition cycle (NAC), consisting of nucleotide substrate initial binding, insertion to the active site, catalysis, product (or pyrophosphate) release, and together with the polymerase translocation[17]. To be successfully incorporated, the antiviral nucleotide analogue needs to pass almost every fidelity checkpoint in the polymerase NAC[18,19]. In coronaviruses with large genome sizes, proofreading conducted by an exonuclease (or ExoN) protein further improves the RNA synthesis fidelity[13]. Correspondingly, the nucleotide analogue drug need further evade from the ExoN proofreading to ultimately terminate the RdRp elongation. Although RDV succeeds as a nucleotide analogue drug to interfere with the CoV-2 RdRp function, as being demonstrated in vivo and in vitro, the underlying structural dynamics mechanisms on how that being achieved are still to be determined, and *in silio* approaches may particularly help. Recent modeling and computational efforts have been made to approach the underlying mechanisms of the RDV-TP binding and incorporation to the CoV-2 RdRp, from molecular docking[20] and binding free energy calculation upon the nucleotide initial association[21], to nucleotide addition together with potential ExoN proofreading activities[22]. Nevertheless, those studied structural systems were still made by constructing homology model of the SARS-CoV-2 RdRp according to a previously resolved structure of the SARS-CoV RdRp[23]. Upon very recent high-resolution cryo-EM structures being resolved on the SARS-CoV2 RdRp (the non-structural protein or nsp12), with and without incorporation of RDV[24,25], it becomes highly desirable to conduct all-atom modeling and molecular dynamics (MD) simulations directly on the CoV-2 RdRp structure, so that to probe how RDV succeeds at binding and inserting into the RdRp active site, despite of existing nucleotide selectivity of RdRp to be against non-cognate nucleotide species[16].

The high-resolution structures of SARS-CoV-2 RdRp or nsp12 were obtained in complex with accessory protein nsp7 and nsp8, which are supposed to assist processivity of the replication/transcription machinery along the viral RNA[26] (see **Fig 1A**). The core RdRp (residue 367-920, excluding the N-terminal NiRAN and interfacial region) adopts a handlike structure, consisting of fingers, palm, and thumb subdomains, similar to other single-subunit viral RNA polymerases (RNAPs) and family-A DNA

polymerases (DNAPs)[27–29]. There are seven highly conserved structural motifs shared by RdRps, located in the palm (A-E) and fingers (F-G) subdomains (**Fig 1B**). In general, when there is no substrate bound, the RdRp active site adopts an open conformation. A nucleotide substrate can bind to the active site in the open conformation, and inserts into the active site to reach a closed conformation, as the nucleotide is stabilized or to be ready for the catalytic reaction[30,31]. In the recently resolved SARS-CoV-2 RdRp structures, both the open and closed conformation state of the active site were captured, with the former in the absence of the substrate[25], and the latter captured with an RDV analogue already incorporated to the end of the synthesizing RNA chain (i.e., in post-catalytic or product state)[24]. In order to probe how a nucleotide or analogue binds and inserts to the RdRp active site, we accordingly constructed both an open (i.e. substrate initial binding) and a closed (substrate insertion) structural complex of the CoV-2 RdRp, based on the newly resolved structures (PDB: 7BTF[25] and 7BV2[24]) (see **Fig 1C** for a closed form). Note that in the single-subunit viral RNAPs or DNAPs, the nucleotide insertion, in accompany with the open to closed conformational transition (pre-chemistry transition or isomerization), usually happens slowly (e.g. milliseconds or above), i.e., to be rate limiting (or partially rate-limiting) in the NAC[32–34]. Such a slow nucleotide insertion step correspondingly plays a significant role in the nucleotide selection or fidelity control, for example, in the single-subunit viral T7 RNAP system studied recently[35–37]. To understand how RDV-TP can evade from the nucleotide selectivity of RdRp to be incorporated, it is therefore essential to probe how such a nucleotide analogue binds stably and inserts sufficiently fast or with low energy barrier into the active site, comparing to its natural substrate counterpart. Accordingly, in this work, we employed all-atom MD simulation to probe mainly the free energetics of the RDV-TP insertion into the CoV-2-RdRp active site, in comparison with the ATP insertion. To do that, umbrella sampling strategies were implemented connecting the initial substrate binding (active site open) and the insertion (active site closed) conformational states, in particular, by enforcing collective coordinates of atoms from structural motif A-G and the inserting NTP (excluding or including the template nucleotide or nt +1 with forcing). The simulations consequently reveal free energetics or potentials of mean force (PMFs) along the reaction coordinate of the RDV-TP and ATP insertion, demonstrating how local residues around the RdRp active site or NTP binding site coordinate with the nucleotide binding, insertion, and differentiation, comparing RDV-TP and ATP.

2 Computational Details

2.1 Building Open/Closed structures and RDV Force Field

High-resolution Cryo-EM structures for CoV-2-RdRp's elongation complex are available in a post-catalysis state with the RDV analogue incorporated (PDB:7BV2)[24]. Using this structure, RDV-TP and ATP were fitted into the active site to create the closed /substrate insertion complex (see **Supplementary Information** or **SI Methods** for details). Missing residues were added using MODELLER 9.24[38] and an apo nsp12 structure as a refer-
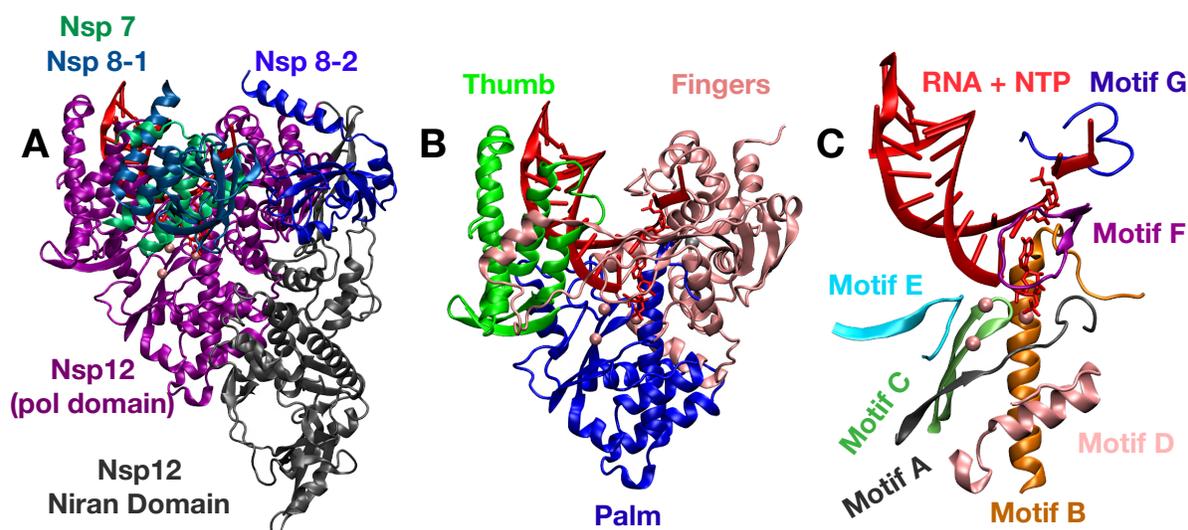
Fig. 1 SARS-CoV-2 RdRp elongation complex with an incorporated remdesivir (RDV) in the closed state (baed on PDB:7BV2 [24]). **A** The two main domains (N-terminus domain in grey, polymerase or pol domain in purple) of the RdRp along with the three cofactors (nsp8's in blue and nsp7 in green). **B** The pol domain consists three subdomains, the thumb (green), fingers (pink), and palm (blue). RNA (red) is shown along with incoming NTP and +1 template nt (red licorice). **C** Motifs A-G within the pol domain.

ence (PDB:7BTF) [25]. The open state was constructed from the apo nsp12 structure [25] along with additionally incorporated RNA strands and RDV-TP (or ATP) by fitting the above constructed RdRP closed structure with the apo RdRp structure.

Histidine protonation states were predicted using PDB2PQR [39] and PROPKA3 [40] followed by visual inspection. The two nsp8 N-terminals were cleaved and shorted by 11 residues to avoid instabilities (see **SI Methods**). A force field was generated for RDV, with partial charges calculated by following the formalism used in amber nucleic acid forcefields [41]. RDV-TP 3' and 5' terminals were truncated, and replaced with terminal hydroxyl groups (see **SI Methods** and **Fig S1**). A Hartree-Fock calculation at the level of HF/6-31G* was set to perform geometric optimization and a self-consistent calculation to obtain an electro-static potential for constrained charge fitting [42]. Using the two-stage Restrained Electrostatic Potential method [43], partial atomic charges for the RDV were generated(see **SI Table S1** for more details). Torsional parameters were taken from Parmbsc1 when applicable and the general Amber force field (GAFF) [44]. RDV force field parameters were constructed using antechamber [45]. Adaptive Poisson-Boltzmann Solver (APBS) mapping of the modeled substrate structures are provided in **SI Fig S2**. Docking of RDV-TP and ATP has also been conducted to the open state RdRp complex to compare and confirm with the constructed initial binding structural complex of RDV-TP or ATP(see **SI Methods Fig S3-4**).

2.2 Simulation Details

All MD simulations were performed using Gromacs 2019 package [46] with the Amber14sb protein force field [47] and Parmbsc1 nucleic acid parameters [48]. For the NTPs, triphosphate parameters calculated previously were used [49]. Each of the RdRp complexes was solvated with explicit TIP3P water [50] with a minimum distance from the complex to the wall set to 15Å, resulting in an average box size of 15.7nm x 15.7 nm x 15.7 nm (see **SI Fig S5**).

Sodium and chloride ions were added to neutralize the systems and make the salt concentration 100mM. Three magnesium ions were kept from the cryo-EM structures (though only two are supposed to be catalytically relevant) [24]. The full simulation systems contained on average about 382,000 atoms. For all simulations, the cut-off of van der Waals (vdw) and the short range electrostatic interactions were set to 10Å. Particle-mesh-Ewald (PME) method [51,52] was used to evaluate the long-range electrostatic interactions. Timestep was 2 fs and the neighbor list was updated every 10 steps. Temperature was kept at 310 K using the velocity re-scaling thermostat [53]. Pressure was kept at 1 bar using Berendsen barostat [54] during pre-equilibration and Parrinello-Rahman barostat [55] for production, targeted MD (TMD), and umbrella simulation runs. Each initial system was minimized for a maximum of 50000 steps using steepest-descent algorithm, followed by a 2-ns NVT MD simulation with all the heavy atoms in the system fully constrained. Next a 2-ns NPT simulation along with the same constraints was performed. Constraints were released in 1-ns intervals in the following order: RNA, nsp8/nsp-7, nsp12/NTP/metal ions. In total for each initial binding and insertion states, ten 100ns equilibration trajectories for ATP and RDV-TP systems were launched independently for a total of $4\mu s$ of simulation time. TMD and umbrella sampling simulations were accomplished by using gromacs patched with plumed 2.6.1 [56].

2.3 Determining the Reaction Coordinate and Calculating Free Energy

The open to close conformational change of the RdRp is expected to be on the order of milliseconds and therefore can not be captured by brute force MD. In order to calculate free energy, the umbrella sampling method was used. [57–59] To use such a method a reaction path needs to be specified and followed. In this study we used TMD to generate such a path between the open and closed states. TMD [60] implementation requires an initial and a

final reference structures to be specified which we continue using in the umbrella sampling simulations. In this work we implemented two slightly varied protocols by manipulating coordinates of two slightly varied atom sets: nsp12 motifs (motif A-G backbone atoms) and NTP (heavy atoms), with or without template +1 nt (heavy atoms). The corresponding RC is then constructed by aligning the structures to the reference structures via the fingers sub-domain and measuring the differences of RMSDs.

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

Where X is the coordinates for the above selected atom sets and $X_{\text{Open ref}}/X_{\text{Closed ref}}$ is for a chosen reference state (see **SI Methods**).

2.3.1 Target MD and Umbrella Sampling

Using the selected open and closed reference structures, the TMD is launched from each state to create paths (forward path started from the open to the closed reference structure, and the backward path started from the closed then to the open reference structure) that meet halfway on the RC (see **SI Fig S6** and **movies S1 & S2**). From the forward and backward TMD paths created between the open and closed states, structures are evenly (for every 0.1 Angstrom in the RC) selected to be used for umbrella sampling simulations. In the umbrella sampling simulations from the selected structures along the TMD paths, harmonic restraints are used along the RC. The force constants used in TMD are continually used in the umbrella sampling simulations (see **SI Table S2**). The biased histograms along the RC for each window were unbiased / re-weighted using the weighted histogram analysis method[61]. From the generated biased trajectories a short set of data is removed from the beginning of each for equilibration (10 ns for RTP simulations and 20 ns for ATP simulations as it takes longer time for ATP simulation systems to converge). Overlap for each set of windows was checked along the reaction coordinate (see **Fig S7**). The unbiased probabilities and then the free energy are also calculated using WHAM package[62], following equations:

$$P_i(RC) = \exp\left[\frac{-k(RC - RC^o)^2}{2k_BT}\right]P_i'(RC)$$

$$G(RC) = -k_BT \ln P_i(RC)$$
(2)

Where $P_i(RC)$ and $P_i'(RC)$ are the unbiased and biased probabilities sampled for the i-th window, respectively. The harmonic restraint potential is shown by $\frac{k}{2}(RC - RC^o)^2$ where RC^o is for the initial structure obtained from the TMD insertion path. Finally free energy profile G along the RC is calculated taking the logarithm of unbiased probabilities, which represent the PMF.

While constructing the PMF using WHAM, bootstrapping error analysis[63] is used to estimate errors. Bootstrapping re-samples RC_i in each window; from each bootstrapped trajectory $RC_{b,i}(t)$ a new histogram ($h_{b,i}(RC)$) is created. From the new histograms the PMF and $G_b(RC)$ are reconstructed, this process is repeated N times (N = 500 used in this study) generating N bootstrapped PMFs $G_{b,j}(RC)(j = 1, 2, ..., N)$. The uncertainty of a PMF is estimated by a standard deviation calculated by the N bootstrapped

PMF's.

$$\sigma_{PMF}(RC) = \left[(N-1)^{-1}\sum_{j=1}^{N}(G_{b,j}(RC) - \langle G_b(RC)\rangle)^2\right]^{\frac{1}{2}}$$
(3)

2.3.2 Hydrogen Bond Analysis around NTP

To examine the corresponding nucleotide insertion dynamics (with intermediate or transition state over-sampled in the umbrella sampling simulations), hydrogen bond (HB) analysis was performed on the trajectories sampled along the RC of the NTP insertion from open to closed. This was done by taking the last 10 ns of each window and combining them into a single trajectory for each simulation system. HBs were measured using the MDAnalysis[64] python package with a heavy atom donor-acceptor distance cutoff of 3.5 Å and angle cutoff of 140°. From this analysis, plots were created to indicate when a particular HB was present or not from open to close along the RC. Using a similar strategy plots for electrostatic interaction (with a distance cutoff 5 Å) of salt bridges for the NTP polyphosphate were also constructed (see **Fig S8**). Stacking was determined by measuring whether the two base rings are parallel and overlap well.(see **SI Methods** and **Fig S9**). The Mg^{2+} ions were also analyzed by measuring the distances between the Mg^{2+} ion and the NTP center of geometry (C.O.G.) (see **Fig S10**). The measurements show that only the two of the three Mg^{2+} ions (MgA and MgB) are comparatively stabilized near the bound NTP substrate, while the third one (MgC) stays a bit far, suggesting that the third Mg^{2+} does not play as much of a role in coordination as the two catalytically important Mg^{2+} ions.

3 Results

Upon modeling the active-site closed state complexes for the RDV-TP and ATP insertion, respectively, and then constructing the active-site open state complex to allow the substrate to bind initially (see **SI Methods**), we conducted equilibrium all-atom MD simulations for the closed and open complex systems, bound with RDV-TP or ATP (see **Fig 2**). Base pairing between RDV-TP or ATP with the +1 template nt (Uracil) can well be maintained in the closed state (see **Fig 2** and **Fig S11**). In the open state the base pairing between the RDV-TP or ATP with the template nt appears less or slightly less stabilized. Interestingly, base stacking configuration between RDV-TP and the template nt can be frequently captured (see **Fig S9**), in which the nt base usually stacks upstream relative to RDV-TP (see **Fig 2**). Then we performed TMD simulations to generate the nucleotide substrate (ATP and RDV-TP) insertion paths, and finally conducted a series of umbrella sampling simulations to obtain the nucleotide insertion PMFs for individual systems. The results show uniformly that the closed insertion state is more stabilized than the open initial binding state for each substrate, while the relative stability of the open states ($\Delta G^{OC} = G_{open} - G_{closed}$) and the insertion barriers ($\Delta h^{ins} = G^{Barrier} - G_{Open}$) vary for individual systems. We illustrate results on these systems below, for the ATP insertion, (i) excluding and (ii) including +1 template nt in the RC, initiated from the open state, with ATP base pairing with the template nt;
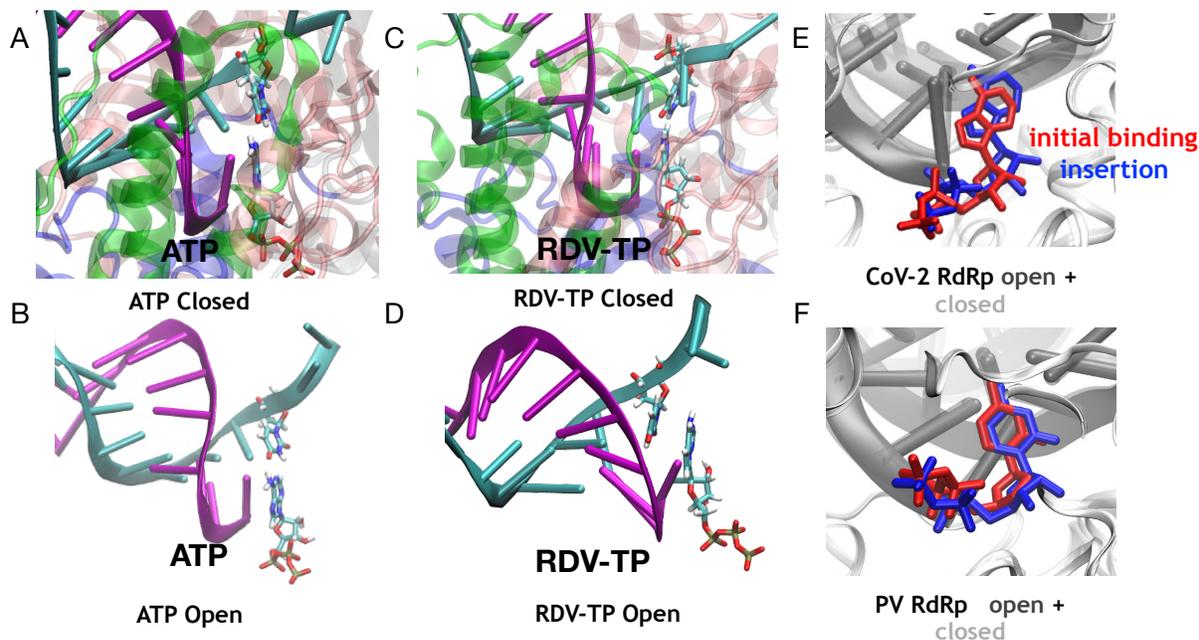
Fig. 2 Modeled insertion structural complexes of SARS-CoV-2 for RDV-TP and ATP. Left and Center: The active site views with inserted ATP and RDV-TP shown at the end of equilibrium simulations for the insertion (A & C) and initial binding (B & D) states. Right: The open and closed RdRp structures aligned (E), with ATP initial binding and inserted shown, respectively. The CoV-2 RdRp is shown in comparison with previously studied RdRp from Poliovirus (PV) (F) (PDBs: 3ola and 3ol7)[30].

for the RDV-TP insertion, initiated similarly from the (iii) RDV-TP base pairing with the template nt under forcing (i.e., included in the RC), and then from a varied initial configuration, i.e., RDV-TP stacking with the template nt, as the nt (iv) included and (iv) excluded in the RC (i.e., with and without forcing).

3.1 Insertion of ATP into the active site can be facilitated by base pairing with the stabilized template nt (+1)

Upon MD equilibration of the initial binding open-state RdRp complex with ATP (∼100 ns; see **SI Fig S12** for RMSD), we found that ATP shows primarily the base pairing initial binding configuration with the +1 template nt. The base pairing interactions seem to be much stabilized in the closed-state ATP insertion configuration (see **Fig 2A&B** and **SI Fig S11A&B**). By obtaining quasi-equilibrated reference structures from the open-state ATP binding and closed-state ATP insertion complexes, we performed the TMD simulation between these two reference structures and constructed the ATP insertion path for conducting the umbrella sampling simulations. The convergence of the PMFs for the ATP insertion requires about 100∼200 ns MD simulation for individual simulation window (see **Fig S13A&B**). In the first simulation system, ATP constantly forms base pairing with the +1 template nt in the initial binding or active-site open state. We conducted the umbrella sampling simulations by forcing atoms from motif A-G and ATP along the TMD insertion path. In this case, the +1 template nt is excluded from the RC, so it is subject only to thermal fluctuations but not the umbrella forcing or constraining. Under such conditions, we noticed that ATP can become highly destabilized by occasionally shifting its base far from the active site in the open state and during barrier crossing (see **Fig 3**). Overall,

the ATP insertion can still proceed toward the comparatively stabilized closed state, with ATP base pairing with the template nt much better than in the open state. Correspondingly, the open to closed free energy drop is obtained as ΔG^{OC} ∼4.8±0.3 kcal/mol and the ATP insertion barriers appears high as Δh^{ins} ∼5.0±0.3 kcal/mol. During insertion, one can see that motif F-K551 (R555) and K798 (near motif D C-term) constantly form HB interactions with the triphosphate of ATP throughout the process, along with motif F-K545 and the template nt; motif C-D760 form occasional HBs with the ATP sugar at open state to barrier crossing, but not into the closed state; motif B-N691 and motif A-D623 form no HBs with ATP sugar until the closed state or crossing the barrier, along with motif F-R553 with the ATP phosphate and motif G-K500 with the template backbone.

Next, in order to stabilize the ATP insertion process, we included the +1 template nt in the RC (i.e., with the umbrella forcing) and constructed the second PMF (see **Fig 4**). Consequently, with the ATP:template nt base pairing is better stabilized. The ATP base deviated less frequently and not that far from the active site in the open to the barrier crossing state, and ATP base pairing with the template nt can recover sooner after barrier crossing. Notably, the ATP insertion barrier lowers to Δh^{ins} ∼2.6±0.3 kcal/mol, although the initial open state stability maintains similarly as in the first case (or slightly less stabilized: ΔG^{OC} ∼5.1±0.2 kcal/mol relative to the closed state). Hence, forcing on the template nt or quenching the fluctuations seems to facilitate the ATP insertion, likely by stabilizing the transition state with the ATP-template nt base paring. Such an operation can mimic the spontaneous ATP insertion process that happens sufficiently slowly (e.g. over milliseconds). Overall, the ATP local interactions with nearby amino
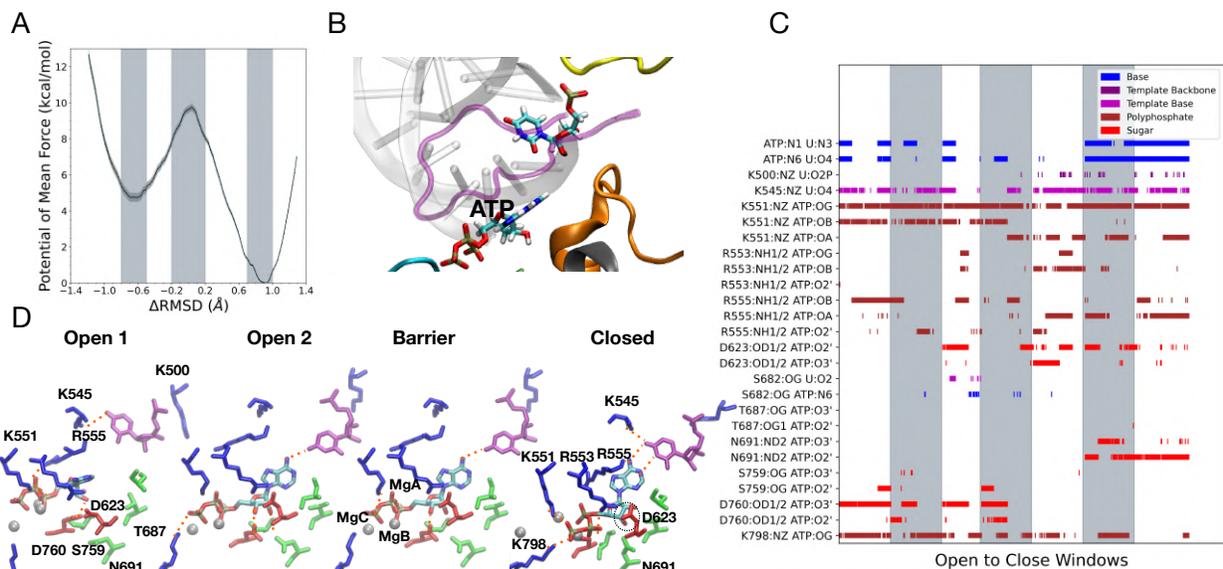
Fig. 3 ATP insertion from umbrella sampling MD simulation (without force on the +1 template nt in the RC). **A** PMF with barrier 5.0±0.3 kcal/mol and an initial binding stability of 4.8±0.3 kcal/mol. **B** Open conformation with ATP not forming hydrogen bonds with +1 template base. **C** Systematical HB patterns; the grey bars represent Open, Barrier, and Closed regions of the simulation windows (see **SI Fig S8A** for salt bridges). **D** Interaction snapshots from simulation windows: Two open states are shown due to the volatility of the open minima, ATP often flips out of plane from the +1 template base. As the barrier is crossed it begins to form consistent base pairing with the template. Dotted orange lines highlight essential HB interactions.
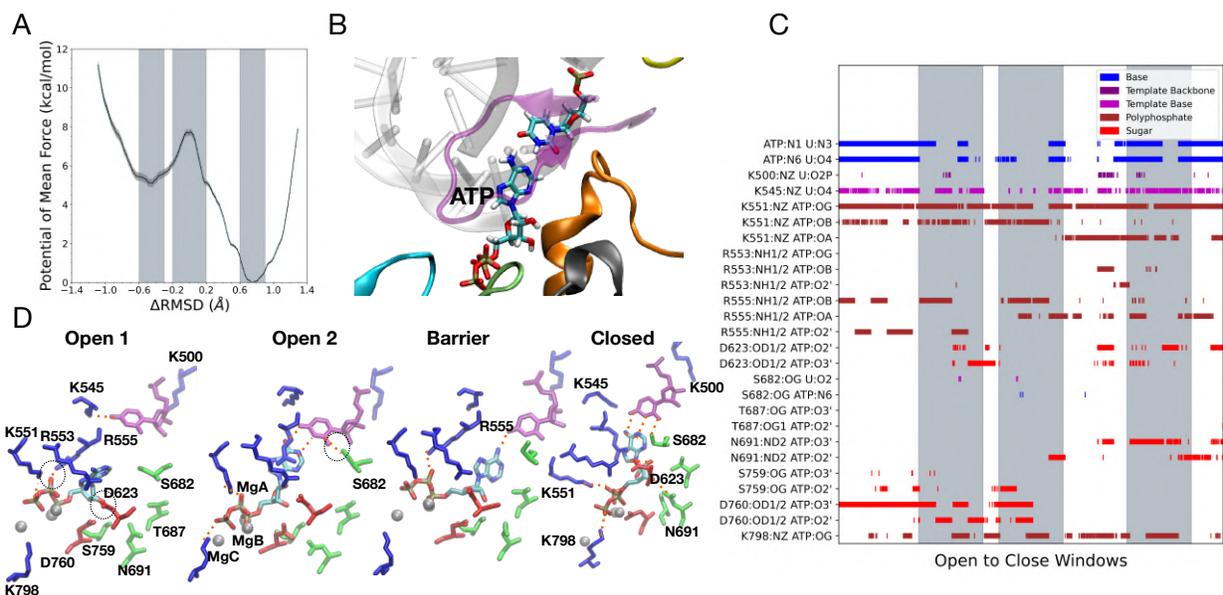

Fig. 4 ATP insertion from umbrella sampling MD simulation (with force on the +1 template nt in the RC). **A** PMF with barrier of 2.6 ± 0.3 kcal/mol and an initial binding stability of 5.1±0.2 kcal/mol. **B** Open conformation of ATP forming hydrogen bonds with +1 template base. **C** Systematical HB patterns; the grey bars represent Open, Barrier, and Closed regions of the simulation windows as shown in the PMF (see **SI Fig S8B** for salt bridges). **D** Interaction snapshots from simulation windows: Two open states are shown due to the volatility of the open minima. Although ATP still occasionally flips out of plane, it more consistently forms HB with the +1 template base. Dotted orange lines highlight essential HB interactions.

acids around the active site appear similarly in the two simulation systems, except that in the current template forced condition, the HBs from motif A-D623:sugar and motif G-K500: template formed a bit earlier in the open state, and motif F-R555 forms HBs with the ATP phosphates more often throughout the process. Hence, the D623-sugar, R555-phosphate, and the K500 template

nt interactions, along with the template forcing on stabilizing the ATP-template nt base pairing seem to contribute to the lowered ATP insertion barrier.

3.2 RDV-TP initial stacking with the +1 template nt is more stabilized than the base pairing

Upon MD equilibration of the open-state RdRp complex with RDV-TP ($\sim$ 100 ns; see **Fig S14** for RMSD), we found that RDV-TP shows primarily two unique open state binding configurations: one still with standard base pairing and the other with the RDV base stacking with the +1 template uracil base (see **Fig 2C&D**). We next constructed the PMF for the RDV-TP initially base pairing with the template nt (see **Fig 5**), applying force or constraint to the template (similarly as to the ATP insertion in **Fig 4**). Then we chose the varied initial binding configuration as the RDV forms base stacking with the template nt, keeping the force constraint on the template, and repeated the calculations (see **Fig 6**). Note that the convergences of the RDV-TP insertion energetics happen much faster ($\sim$50 ns; see **Fig S13C-E**) than that of the ATP system. The PMF of the RDV-TP insertion starting from the base pairing configuration shows that the insertion barrier is high ($\Delta h^{ins} \sim$5.4$\pm$0.3 kcal/mol), comparing to the ATP insertion barrier obtained in the similar conditions ($\Delta h^{ins} \sim$2.6$\pm$0.3 kcal/mol from **Fig 4**). The relative stability of the open initial binding state of RDV-TP to the closed insertion state is also measured ($\Delta G^{OC} \sim$4.5$\pm$0.3 kcal/mol), slightly more stabilized, relatively, than that in the corresponding ATP base pairing system ($\Delta G^{OC} \sim$5.1$\pm$0.2 kcal/mol from Fig 4). Now motif F-K551, R553&R555 form HBs less or more with the triphosphate of RDV-TP throughout the process, along with motif F-K545 with the template; motif C-S759 and D760 form HBs with the sugar at open state to barrier crossing, not afterwards or into the closed state; motif B-N691 and motif B-T687 barely forms HB with the sugar until the barrier crossing, along with motif G-K500 and the template. Overall, motif F-R553 and R555 form stronger interaction with the RDV-TP triphosphate than in the ATP insertion cases, while motif A-D623 barely forms HB with the RDV-TP sugar into the closed or insertion state (but with ATP sugar in the insertion state). In contrast, motif B-T687 forms HB with the RDV-TP sugar in the insertion state, while there is no HB interaction of it with ATP at all.

More interesting results come from comparing RDV-TP insertion energetics and interactions simulated at the varied conditions. In **Fig 6**, we show the PMF from RDV-TP initially stacking with the +1 template nt, with forcing still implemented. Although the insertion barrier ($\Delta h^{ins} \sim$5.2$\pm$0.3 kcal/mol) remains similarly high as the above case (from **Fig 5**), the relative stability of the initial open state to the final insertion or closed state changes (to $\Delta G^{OC} \sim$2.6$\pm$0.2 kcal/mol), indicating that the initial stacking configuration of RDV-TP is more stabilized (about -3 k_BT) than the initial base pairing configuration with the template nt). By comparing the HB patterns (**Fig 5C** and **Fig 6C**), one finds that the stabilizing interactions to the base stacking configuration at the open state mainly come from motif A-D623 and motif C-S759 with the sugar, S682 interaction with the RDV-TP base, K798 (near motif D) along with motif F-K551 interaction with the phosphate, as well as motif G-K500 interaction with the template nt. The motif F R555/R553 interaction with the RDV-TP triphosphate weaken from the initial base pairing to the stacking configuration.

3.3 RDV-TP insertion to the active site is facilitated by thermal fluctuations and absence of motif F-R553/R555 interaction with the polyphosphate

Since the above results show that the RDV-TP initial stacking with the +1 template nt is more stabilized than the base pairing configuration, we further explored the RDV-TP insertion barrier by removing the forcing on the +1 template nt (i.e. being excluded from the RC). Notably, the insertion now is greatly facilitated by allowing sufficient fluctuations on the template , such that the insertion barrier becomes lowest ($\Delta h^{ins} \sim$1.5$\pm$0.2 kcal/mol; see **Fig 7A**). Meanwhile, the relative stability of the open binding state to the closed insertion state of RDV-TP maintains ($\Delta G^{OC} \sim$2.7$\pm$0.1 kcal/mol), as in the above system from **Fig 6**.

It appears that thermal fluctuations on the template nt can actually support the RDV base stacking with the template nt along with "shaking" the motif F-R553/R555 interaction off triphosphate before transition toward the insertion configuration, in which RDV-TP can form very stabilized base pairing interactions with the template nt.

Additional close inspections on the RDV-TP local interactions with nearby residues show that the majority of HB and SB interactions are similar between the cases without and with forcing on the +1 template nt (**Fig 7C** and **Fig 6C**). Interestingly, one can identify that both HB and SB interactions from R555 and R553 (located on the motif F) with the triphosphate of RDV-TP, which are formed for the RDV-TP initial binding in the former stacking case with template forcing (see **Fig 6C**), but become absent in the current case (without forcing on the template nt, **Fig 7C**). Otherwise, the local HB/SB interactions with RDV-TP are highly similar for the two systems (**Fig 6** and **Fig 7**), both initiated from the RDV-TP base stacking with the template nt binding configuration. Hence, in the RDV-TP insertion, the presence of the template forcing (or reduced fluctuations) along with the R555 (and R553) interaction with the triphosphate seems to hinder the RDV-TP insertion, which appears to be opposite to the trend in the ATP insertion (i.e., stronger R555/R553-ATP phosphate interaction in the open state under template forcing condition leads to a lowered ATP insertion barrier).

4 Discussion

In this work we modeled and simulated insertion of the triphosphate form nucleotide analogue drug remdesivir (RDV-TP) into the SARS-CoV-2 RdRp active site, in comparison with natural nucleotide substrate ATP. Our work is based on high-resolution cryo-EM structures solved for the SARS-CoV-2 nsp12 in complex with cofactors nsp7 and nsp8 [24,25], modeled in an active-site open form (PDB: 7BTF) for the nucleotide initial binding, and in an active-site closed form (PDB: 7BV2) for the stabilized nucleotide insertion, prior to catalytic addition of the nucleotide to the synthesizing RNA chain. The viral RdRp or nsp12 in the coronavirus species works with other non-structural proteins (nsp7 to nsp16) for viral genome replication and transcription [65,66], with nsp7 and nsp8 the cofactors to assist the replication machinery stability
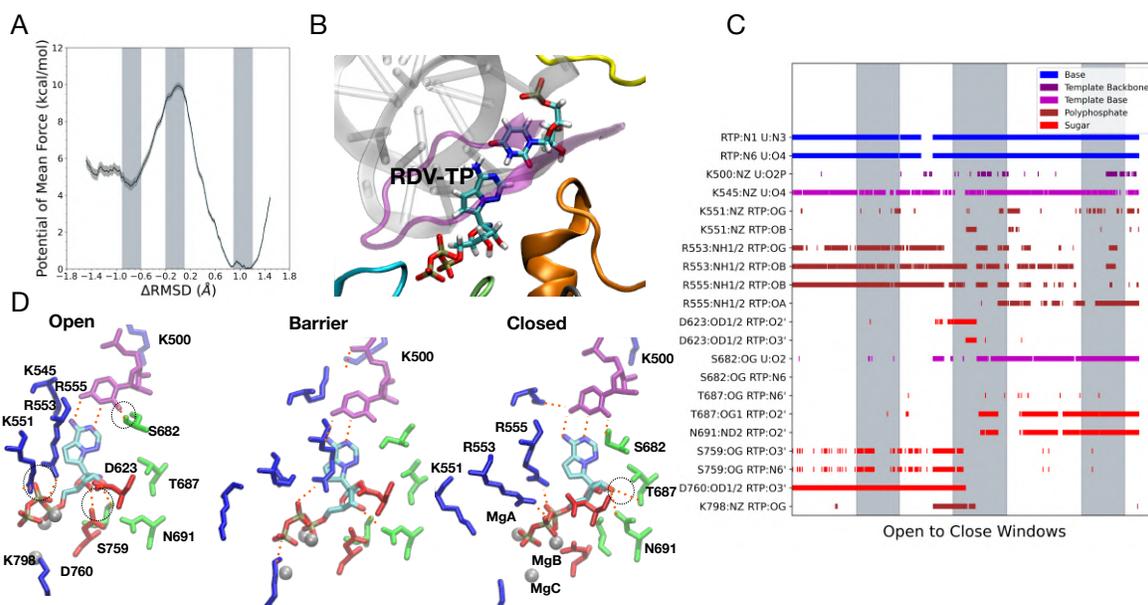
Fig. 5 RDV-TP insertion with the open state forming good base pairing from umbrella sampling MD simulation (with force on the +1 template nt in the RC). **A** PMF with barrier of 5.4 ± 0.3 kcal/mol and an initial binding stability of 2.6±0.3 kcal/mol. **B** Open conformation of RDV-TP forming hydrogen bonds with +1 template base. **C** Systematical HB patterns; the grey bars represent Open, Barrier, and Closed regions of the simulation windows as shown in the PMF (see **SI Fig S8C** for salt bridges). **D** Interaction snapshots from simulation windows: Throughout the open state stable HB form with the RDV and +1 template base. Dotted orange lines highlight essential HB interactions.
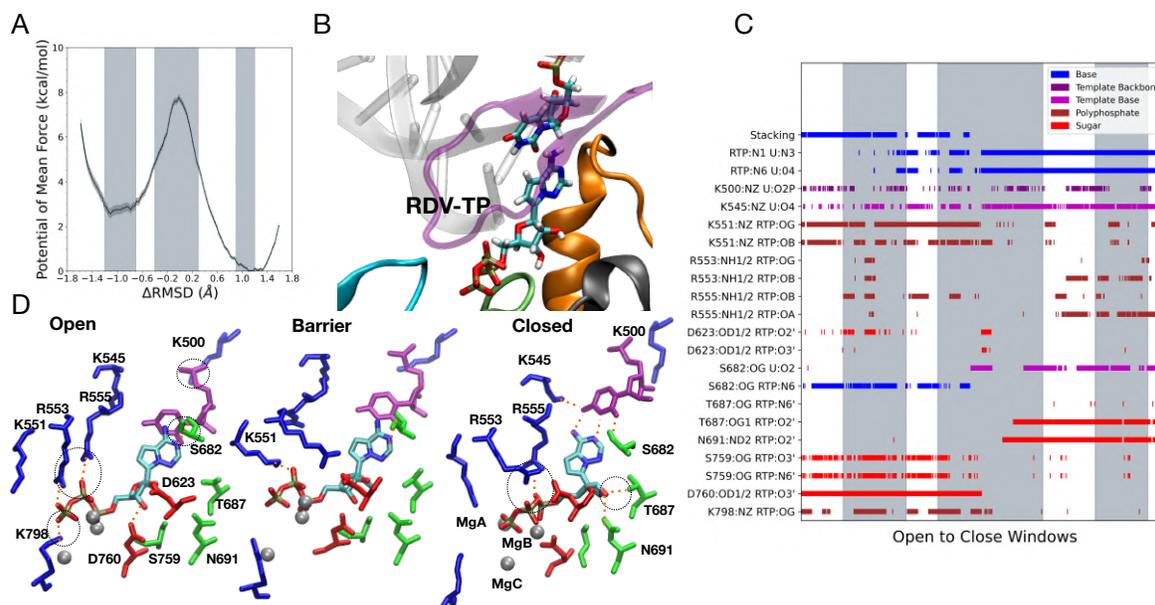

Fig. 6 RDV-TP insertion with the open state forming base stacking with the +1 template base from umbrella sampling MD simulation (with force on the +1 template nt in the RC). **A** PMF with barrier of 5.2 ± 0.3 kcal/mol and an initial binding stability of 2.6±0.2 kcal/mol. **B** Open conformation of RDV-TP forming base stacking with +1 template base. **C** Systematical HB patterns; the grey bars represent Open, Barrier, and Closed regions of the simulation windows as shown in the PMF (see **SI Fig S8D** for salt bridges). **D** Interaction snapshots from simulation windows: Throughout the open state base stacking forms resulting in a more stable minima. Dotted orange lines highlight essential HB interactions.

and processivity along the viral genome, and with nsp13 [67,68] and nsp14 [69] functioning as helicase and exonuclease, respectively. In the simulation of the nsp12-nsp7-nsp8 complex along with RNA strands, we found that shortening of the nsp8 N-terminal (e.g. to start from residue M67) is necessary to stabilize the simulation complex in all-atom explicit water condition. It is however noted

that the two copies of nsp8 can extend very long as 'sliding poles' on a protruding exiting RNA duplex, as being captured from another high-resolution cryo-EM complex of nsp12-nsp7-nsp8 [26]. In modeling of an initial binding complex of the nucleotide or analogue, we placed ATP or RDV-TP to the open active site of CoV-2 RdRp, according to RdRp structural alignments between the prod-
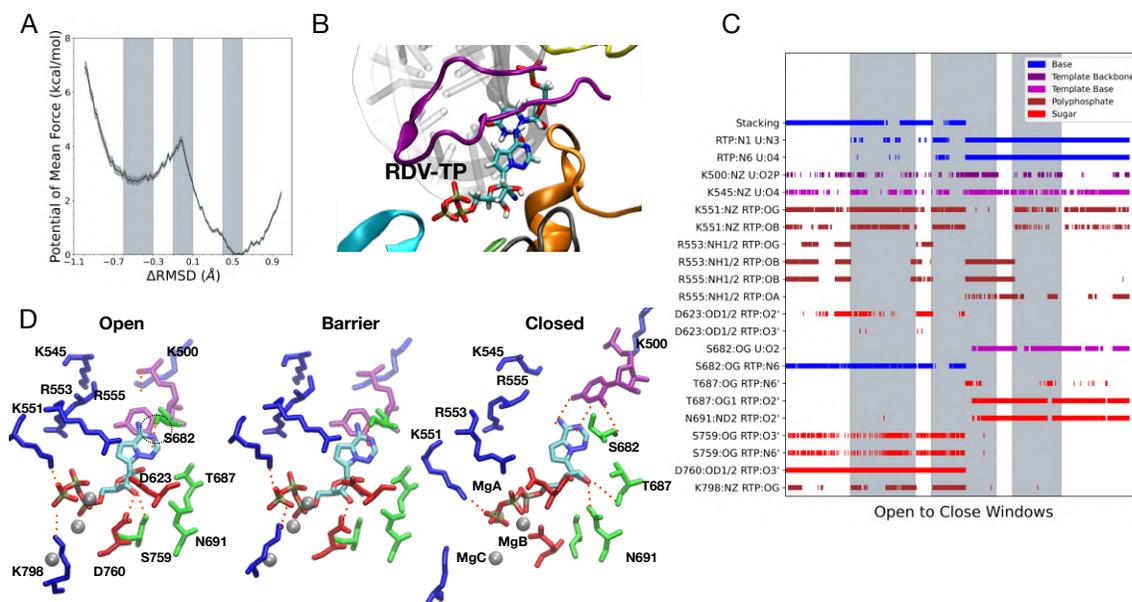
Fig. 7 RDV-TP insertion with the open state forming base stacking with the +1 template base from umbrella sampling MD simulation (with no force on the +1 template nt in the RC). **A** PMF with barrier of 1.5 ± 0.2 kcal/mol and an initial binding stability of 2.7±0.1 kcal/mol. **B** Open conformation of RDV-TP forming base stacking with +1 template base. **C** Systematical HB patterns; the grey bars represent Open, Barrier, and Closed regions of the simulation windows as shown in the PMF (see **SI Fig S8E** for salt bridges). **D** Interaction snapshots from simulation windows: Throughout the open state base stacking forms resulting in a more stable minima. Dotted orange lines highlight essential HB interactions.

uct complex (closed form) of RDV-TP and the open one. Accordingly, the positioning of RDV-TP or ATP appear similar between the open and closed structures. Molecular docking and simulation equilibration confirmed such an initial nucleotide binding configuration is dominant (see **SI Fig S3&S4**), which also shows similarly to that being captured in the poliovirus (PV) RdRp[30]. Hence, for the RDV-TP and ATP insertion probed in this work, we focus mainly on subtle local interactions around the active site of the viral RdRp as for the incoming nucleotide being recruited, interrogated, and re-positioned to allow chemical addition. Meanwhile, we note that the open and closed forms of the viral RdRp structure still involve collective movements of the highly conserved motifs (A to G) which we manipulate as a whole in the umbrella sampling simulations, to ensure the concerted nucleotide insertion. Note that motifs A to E are located in the palm subdomain hosting the active site, with motif C mainly responsible for catalysis, and motif A,B, and D for nucleotide binding and selection; motif F-G from in the fingers subdomain also impacts on the incoming nucleotide entry as well as the +1 template nt for the Watson-Crick (or WC) base pairing or fidelity check[16,70].

Correspondingly, we conducted first the equilibrium MD simulations, which show that upon the initial binding, ATP frequently forms the WC base pairing with the template nt but with notable fluctuations; in contrast, RDV-TP primarily forms base stacking with the template nt, squeezing the template base to upstream most of time. Although RDV-TP has also been sampled in base paring with the template uracil base, such a base stacking configuration appears more stable. In the closed RdRp or insertion state, RDV-TP anyhow forms highly stabilized base pairing with the template nt, with even lower fluctuations than ATP for natural base pairing. APBS mapping zoomed into the closed active site of CoV-

2 RdRp shows notable differences between the local electrostatic environment around the inserted RDV-TP and ATP (see **Fig S2**), in particular around the sugar region, where an extra cyano group is attached to RDV-TP, with T687 and N691 associated nearby. In order to see how exactly RDV-TP and ATP insert into the active site from the initial binding state, as the open active site closes, we then performed the TMD and umbrella sampling simulations connecting the open and closed RdRp complex structures, with slightly varied initial and collective coordinate forcing conditions.

The purpose of running the TMD simulations was to construct feasible dynamical paths of the nucleotide insertion to be used in the umbrella sampling simulations for the PMF construction, upon that the structural dynamics (with enhanced sampling in the transition state or barrier region) and energetics (or free energy profiles) of the insertion processes reveal and can be further compared. Our simulations first confirm that the nucleotide inserted or the closed form of the CoV-2 RdRp is indeed much more stabilized than the open form for nucleotide initial binding (about -3 to -5 kcal/mol), for RDV-TP or ATP. While the base pairing configurations of the initial binding ATP and RDV-TP are similarly stabilized (∼5 kcal/mol) relative to the corresponding closed insertion state, such an initial binding configuration is only dominant to ATP but not RDV-TP. Essentially, our calculations show that RDV-TP primarily forms base stacking with the +1 template nt rather than base pairing upon initial binding. Comparison between the RDV-TP insertion simulations conducted with varied initial binding configurations (stacking and base pairing) shows that motif A-D623 may stabilize the RDV-TP base stacking over the base pairing in the open state, by forming HBs with the sugar; in addition, motif C-S759 specifically forms HB with the RDV-TP sugar; S682 (near motif B) forms highly notable HB contact with

Nucleotide species	Open state (initial binding)	Insertion barrier region	Closed state (insertion)
RDV-TP Initial binding	Initially base stacking with template $\Delta G^{oc} \sim$ 2.7 kcal/mol Base stacking stabilized by: ▪ Motif A-D623 and Motif C-S759 with sugar; ▪ S682 (near Motif B) with base; ▪ Motif F-K551 and K798 with phosphate; ▪ Motif G-K500 with template +1 backbone	Low barrier (no template forcing or with sufficient fluctuations) $\Delta h^{ins} \sim$1.5 kcal/mol High (with template forcing or low fluctuations) $\Delta h^{ins} \sim$5.2 kcal/mol Hindered by: ▪ Motif F-R555/R553 with phosphate from open or initial binding	Distinguished by: ▪ Motif B-T687 with sugar ▪ Highly stabilized base-pairing with template
	Initially base pairing with template $\Delta G^{oc} \sim$ 4.5 kcal/mol	High barrier (with template forcing or low fluctuations) $\Delta h^{ins} \sim$5.4 kcal/mol	
ATP Initial binding	Initially base pairing with template $\Delta G^{oc} \sim$ 5.1 kcal/mol Insertion facilitated by: ▪ Motif F-R555 with phosphate ▪ Motif A-D623with sugar	Medium Low barrier (with template forcing or low fluctuations) $\Delta h^{ins} \sim$2.6 kcal/mol	
	Initially base pairing with template $\Delta G^{oc} \sim$ 4.8 kcal/mol	High (no template forcing or high fluctuations) $\Delta h^{ins} \sim$5.0 kcal/mol	Distinguished by ▪ Motif A-D623 with sugar ▪ K798 with γ-phosphate ▪ Stabilized base pairing with template
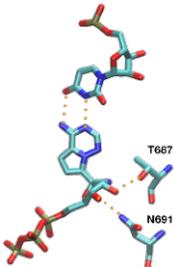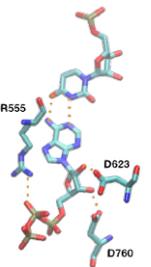
Table 1 Summary of essential hydrogen bonding (and salt-bridge) interactions and energetics during RDV-TP and ATP insertion into SARS-CoV-2 RdRp

the RDV-TP base, only in the base stacking configuration; motif F-K551 and K798 near the C-terminal of motif D stabilize the base stacking configuration by forming HB (or SB) interactions with the RDV-TP triphosphate,(also happen for ATP initial binding); motif G-K500 also forms HB with the +1 template backbone. Interestingly, as motif C-S759 does not form HB to ATP sugar upon initial binding, it forms HBs with the 3'-end of the primer RNA nt sugar in that case (see **Fig S15**). Such interactions persist into the insertion states of both ATP and RDV-TP system. Additionally, in the case of RDV-TP base stacking with the template +1 nt, motif F-R555 also forms HB with the 3'-end of the RNA nt sugar (see **Fig S15**). Consequently, the 3'-end of primer RNA nt cannot be involved with base stacking with RDV-TP, while R555 interaction with the phosphate of RDV-TP may also be prevented (see analyses later). The overall stabilization leads to ∼-2 kcal/mol (or ∼-3 kBT) relative initial binding free energy between the RDV-TP stacking and base pairing configuration. A docking stabilization

energetics (∼-0.6 kcal/mol) between the RDV-TP and ATP was reported to a homology modeled CoV-2 RdRp[20], and a similar energetic score revealed from our own docking trials (using the open form RdRp complex with RNA strands, see **SI Fig S3&S4**). Hence, it seems that the initial binding of RDV-TP to the CoV-2 RdRp can be about -2 to -3 kcal/mol more stabilized than ATP. An alchemical MD simulation for relative binding free energy calculation have presented a comparable stabilization energetics between RDV-TP and ATP (∼-2.8 kcal/mol) upon binding to the RdRp active site[21]. Nevertheless, the alchemical calculation was conducted in the absence of RNA, so it is unable to be compared in regard to the template RNA configuration. The computational results so far consistently point out that RDV-TP can bind to the CoV-2 RdRp active site (in an open form) more favorably than the natural nucleotide substrate ATP.

Nevertheless, the initial binding to RdRp only provides an initial nucleotide association and selection checkpoint to the nu-

cleotide addition cycle or NAC. The followed insertion of the nucleotide to the active site becomes a next and likely the most important checkpoint in the NAC, in particular, for the single-subunit handlike RNA or DNA polymerases (RNAPs or DNAPs). In several such polymerase species, the nucleotide insertion is rate-limiting (or partially rate-limiting) [34,71,72], thus being critical for nucleotide selection [35]. Comparing to phage T7 RNAP we studied previously [36,37], for which a substantial fingers subdomain rotation happens with respect to the palm subdomain (from open to closed) during the nucleotide insertion, the viral RdRp conformational changes in accompany with the nucleotide insertion are mainly the active site distortions (from open to closed) [30], though remote residues on the structural motifs (A-G) can be more or less involved in the process. From the TMD simulations enforcing the CoV-2 RdRp from open to closed (see **SI Fig S6**), we found that the motif A and D close similarly as that in PV RdRp [31]. Interestingly, the inserting ATP or RDV-TP has the base easily repositioned toward the closed configuration in the TMD simulation, but has the triphosphate moiety hardly reaching to the targeted closed configuration (see **SI movies S1 & S2**). Hence, repositioning of the triphosphate during the nucleotide insertion appears to link to events of crossing the free energy barrier. In current umbrella sampling simulations, the ATP or RDV-TP insertion barrier indeed depends on the relative template nt configuration or fluctuations, as well as local residue interactions with the triphosphate. In the ATP insertion, a comparatively low energetic barrier ($\sim$2.6 kcal/mol) shows when the template nt is enforced or constrained to maintain stabilized base paring with ATP as if in the long-time unperturbed nucleotide insertion; the motif F-R555 interaction with the ATP phosphates along with motif A-D623 interaction with the sugar at the open state seems to facilitate the further ATP insertion. In comparison, for RDV-TP, the insertion barrier can be even lower ($\sim$1.5 kcal/mol) when it is inserted without enforcing the template nt, so that the initial base stacking between RDV-TP and template nt can proceed freely to easily transit to the base pairing configuration into the closed insertion state. Contrary to the ATP insertion, motif F-R555/R553 close interactions (hydrogen bonding and salt-bridge) with the RDV-TP triphosphate in the open state appears to impede the RDV-TP insertion, which happens as the template nt is enforced in the simulation, no matter which initial configuration RDV-TP starts with the template (base pairing or stacking). Current simulations comparing RDV-TP and ATP thus suggest that the nucleotide insertion is coordinated by +1 template nt as well as the 3'-end of primer RNA nt with some notable interactions on the nucleotide upon initial binding, in which the triphosphate stabilization and re-positioning appear to be essential. It should be pointed out that the triphosphate reorientation of the incoming nucleotide had been suggested for the PV RdRp fidelity control [73]. Additionally, it is interesting to notice that motif F-R555 structurally corresponds to R174 from PV RdRp and R158 in HCV RdRp [16], as well as to Y639 from T7 RNAP that is key to nucleotide selectivity and polymerase translocation [36]. Overall, the ATP insertion seems to be facilitated by an insertion path with quenched fluctuations on the +1 template nt for stabilized base pairing, while the RDV-TP insertion dominated by the template base-stacking pop-

ulations, is supported by freely fluctuating template nt, leading to transition to the highly stabilized base pairing configurations, with an insertion free energy barrier as low as $\sim$1.5 kcal/mol or $\sim$2-3k_BT, marginally above thermal fluctuations.

Both the inserted ATP and RDV-TP can be then further stabilized well in the active site by the base paring interaction with the template nt. Though we haven't yet conducted energetic calculations to evaluate the relative stability between the RDV-TP and ATP in the insertion state, the equilibrium simulations of the insertion complexes of the two species suggest that the RDV-TP can be similarly or even more stabilized than ATP in the closed insertion state. There are also specific interactions that can well distinguish the natural nucleotide substrate from the nucleotide analogue in the insertion state: motif A-D623 forms specific HB contact with the ATP sugar but not with the inserted RDV-TP; K798 near motif D also closely interacts with the ATP gamma-phosphate into the insertion state, but not closely with that of RDV-TP; in contrast, motif B-T687 specifically forms HB with the RDV-TP sugar but not with that of ATP. The overall results (see **Table 1**) thus suggest that binding/insertion of RDV-TP can be more facilitated than the natural substrate ATP to the active site of SARS-CoV-2 RdRp, seemingly consistent with in vitro measurements of the Michaelis-Menten constant K_m obtained smaller for RDV-TP than for ATP, respectively [10,11]. If the nucleotide insertion is a single rate-limiting step (i.e., as in T7 RNAP [34]), then V_{max} should also be significantly larger for RDV-TP than that for ATP, due to a lowest insertion barrier of the RDV-TP. However, the *in vitro* measurements of V_{max} are similar for RDV-TP and ATP [10]. Hence, other rate-limiting steps than the pre-chemical NTP insertion can exist in the NAC of the CoV-2 RdRp, e.g., the chemical catalysis, which may happen a bit slower for RDV-TP than ATP, so that overall the maximum elongation rates become similar for the two nucleotide species. More close examinations of stepwise kinetics of SARS-CoV-2 RdRp are therefore expected, ideally for both cognate and non-cognate nucleotide species, so that substantial information on the complete NAC as well as nucleotide selectivity could reveal. Note that following a successful RDV-TP incorporation to the end of viral RNA chain, additional nucleotide insertion still appears viable until the addition of the nucleotide downstream +3 to the incorporated RDV analog. Such mechanism has been suggested as a delayed chain termination of the nucleotide analogue [10], which arises likely due to aberrant impacts of incorporated analogue on the synthesizing RNA chain in association with the viral RdRp, together with failure of ExoN cleavage or proofreading to the nucleotide analogue.

5 Conclusions

Via modeling and all-atom MD simulation, we found that remdesivir nucleotide analogue can bind to the open active site of SARS-CoV-2 RdRp via base stacking with the +1 template nt. Such a stacking configuration appears to be more stabilized, relative to the insertion state, than the Watson-Crick base pairing configuration formed between ATP and the template uracil base. Umbrella sampling simulations further show that the remdesivir analogue stacking with the fluctuating template then inserts or transits to form high-stabilized base pairing with the template as the active

site closes. The corresponding insertion barrier for remdesivir analogue can be even lower than that of a low-energetic path of the ATP insertion, during which the template forms stabilized base pairing with ATP. Additionally, our analyses on hydrogen bonding and salt-bridge interactions during the nucleotide or analogue insertion show that (i) the initial remdesivir base stacking can be particularly stabilized by motif A-D623 along with motif C-S759 with sugar, S682 with base, and motif G-K500 with the template, motif F-K551 and K798 with phosphate, as well as motif F-R555 with 3'-end primer; (ii) insertion of remdesivir analogue can be facilitated by thermal fluctuations but hindered by motif F-R555/R553 interaction with the triphosphate, while insertion of ATP is made easier by lowering fluctuations and taking advantage of the R555/R553 interaction with the triphosphate; (iii) the inserted remdesivir analogue and ATP are distinguished by specific sugar interaction via motif B-T687 and motif-A D623, respectively. Such findings also reveal potential SARS-CoV-2 RdRp fidelity control via particular residue interactions with the nucleotide substrate sugar, base, and triphosphate moieties, along with +1 template coordination.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

Notes and references

1 R. Carrasco-Hernandez, R. Jácome, Y. L. Vidal and S. P. de León, *ILAR Journal*, 2017, **58**, 343–358.

2 P. Ahlquist, A. O. Noueiry, W.-M. Lee, D. B. Kushner and B. T. Dye, *JOURNAL OF VIROLOGY*, 2003, **77**, 8181–8186.

3 L. P. Jordheim, D. Durantel, F. Zoulim and C. Dumontet, *Nature Publishing Group*, 2013, **12**, 447.

4 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.

5 M. L. Agostini, E. L. Andres, A. C. Sims, R. L. Graham, T. P. Sheahan, X. Lu, E. C. Smith, J. B. Case, J. Y. Feng, R. Jordan, A. S. Ray, T. Cihlar, D. Siegel, R. L. Mackman, M. O. Clarke, R. S. Baric and M. R. Denison, *mBio*, 2018, **9**, 1–15.

6 A. Shannon, N. T. T. Le, B. Selisko, C. Eydoux, K. Alvarez, J. C. Guillemot, E. Decroly, O. Peersen, F. Ferron and B. Canard, *Antiviral Research*, 2020, **178**, 104793–104808.

7 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. Fearns, S. Swaminathan, D. L. Mayers, C. F. Spiropoulou, W. A. Lee, S. T. Nichol, T. Cihlar and S. Bavari, *Nature*, 2016, **531**, 381–385.

8 T. P. Sheahan, A. C. Sims, R. L. Graham, V. D. Menachery, L. E. Gralinski, J. B. Case, S. R. Leist, K. Pyrc, J. Y. Feng, I. Trantcheva, R. Bannister, Y. Park, D. Babusis, M. O. Clarke, R. L. MacKman, J. E. Spahn, C. A. Palmiotti, D. Siegel, A. S. Ray, T. Cihlar, R. Jordan, M. R. Denison and R. S. Baric, *Science Translational Medicine*, 2017, **9**, 3653.

9 K. T. Choy, A. Y. L. Wong, P. Kaewpreedee, S. F. Sia, D. Chen, K. P. Y. Hui, D. K. W. Chu, M. C. W. Chan, P. P. H. Cheung, X. Huang, M. Peiris and H. L. Yen, *Antiviral Research*, 2020, **178**, 104786.

10 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.

11 T. L. Dangerfield, N. Z. Huang and K. A. Johnson, *iScience*, 2020, **23**, 101849.

12 A. J. Pruijssers, A. S. George, A. Schäfer, S. R. Leist, L. E. Gralinksi, K. H. Dinnon, B. L. Yount, M. L. Agostini, L. J. Stevens, J. D. Chappell, X. Lu, T. M. Hughes, K. Gully, D. R. Martinez, A. J. Brown, R. L. Graham, J. K. Perry, V. Du Pont, J. Pitts, B. Ma, D. Babusis, E. Murakami, J. Y. Feng, J. P. Bilello, D. P. Porter, T. Cihlar, R. S. Baric, M. R. Denison and T. P. Sheahan, *Cell Reports*, 2020, **32**, 107940.

13 F. Robson, K. S. Khan, T. K. Le, C. Paris, S. Demirbag, P. Barfuss, P. Rocchi and W. L. Ng, *Molecular Cell*, 2020, **79**, 710–727.

14 J. F. Sydow and P. Cramer, *Current Opinion in Structural Biology*, 2009, **19**, 732–739.

15 C. E. Cameron, I. M. Moustafa and J. J. Arnold, *Enzymes*, Academic Press, 2016, vol. 39, pp. 293–323.

16 C. Long, M. E. Romero, D. La Rocco and J. Yu, *Computational and Structural Biotechnology Journal*, 2021, **19**, 3339–3348.

17 F. Brueckner, J. Ortiz and P. Cramer, *Current Opinion in Structural Biology*, 2009, **19**, 294–299.

18 Y. Yuzenkova, A. Bochkareva, V. R. Tadigotla, M. Roghanian, S. Zorov, K. Severinov and N. Zenkin, *BMC Biology*, 2010, **8**, 54.

19 J. Yu, *Computational and Mathematical Biophysics*, 2014, **2**, 141–160.

20 A. A. Elfiky, *Life Sciences*, 2020, **253**, 117592.

21 L. Zhang and R. Zhou, *Journal of Physical Chemistry B*, 2020, **124**, 6955–6962.

22 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.

23 R. N. Kirchdoerfer and A. B. Ward, *Nature Communications*, 2019, **10**, 1–9.

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 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 (New York, N.Y.)*, 2020.

26 H. S. Hillen, G. Kokic, L. Farnung, C. Dienemann, D. Tegunov and P. Cramer, *Nature*, 2020, **584**, 154–156.

27 N. Cermakian, T. M. Ikeda, P. Miramontes, B. F. Lang, M. W. Gray and R. Cedergren, *Journal of Molecular Evolution*, 1997, **45**, 671–681.

28 J. Filée, P. Forterre, T. Sen-Lin and J. Laurent, *Journal of Molecular Evolution*, 2002, **54**, 763–773.

29 S. T. de Farias, A. P. dos Santos, T. G. Rêgo and M. V. José, *Frontiers in Genetics*, 2017, **8**, 1–7.

30 P. Gong and O. B. Peersen, *Proceedings of the National Academy of Sciences of the United States of America*, 2010, **107**, 22505–22510.

31 G. Campagnola, S. McDonald, S. Beaucourt, M. Vignuzzi and O. B. Peersen, *Journal of Virology*, 2015, **89**, 275–286.

32 K. A. Johnson, *Conformational coupling in DNA polymerase fidelity*, 1993.

33 T. Schlick, K. Arora, W. A. Beard and S. H. Wilson, *Theoretical Chemistry Accounts*, 2012, **131**, 1–8.

34 V. S. Anand and S. S. Patel, *Journal of Biological Chemistry*, 2006, **281**, 35677–35685.

35 C. Long and J. Yu, *Entropy*, 2018, **20**, 306.

36 C. Long, E. Chao, L. T. Da and J. Yu, *Computational and Structural Biotechnology Journal*, 2019, **17**, 638–644.

37 C. Long, E. Chao, L. T. Da and J. Yu, *Nucleic Acids Research*, 2019, **47**, 4721–4735.

38 A. Šali and T. L. Blundell, *Journal of Molecular Biology*, 1993, **234**, 779–815.

39 T. J. Dolinsky, J. E. Nielsen, J. A. McCammon and N. A. Baker, *Nucleic Acids Research*, 2004, **32**, 665–666.

40 M. H. Olsson, C. R. SØndergaard, M. Rostkowski and J. H.

Jensen, *Journal of Chemical Theory and Computation*, 2011, **7**, 525–537.

41 C. I. Bayly, K. M. Merz, D. M. Ferguson, W. D. Cornell, T. Fox, J. W. Caldwell, P. A. Kollman, P. Cieplak, I. R. Gould and D. C. Spellmeyer, *Journal of the American Chemical Society*, 1995, **117**, 5179–5197.

42 R. Ditchfield, W. J. Hehre and J. A. Pople, *The Journal of Chemical Physics*, 1971, **54**, 720–723.

43 J. Wang, P. Cieplak and P. A. Kollman, *Journal of Computational Chemistry*, 2000, **21**, 1049–1074.

44 J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *Journal of Computational Chemistry*, 2004, **25**, 1157–1174.

45 J. Wang, W. Wang, P. A. Kollman and D. A. Case, *Journal of Molecular Graphics and Modelling*, 2006, **25**, 247–260.

46 M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindah, *SoftwareX*, 2015, **1-2**, 19–25.

47 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.

48 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.

49 K. L. Meagher, L. T. Redman and H. A. Carlson, *Journal of Computational Chemistry*, 2003, **24**, 1016–1025.

50 D. J. Price and C. L. Brooks, *Journal of Chemical Physics*, 2004, **121**, 10096–10103.

51 T. Darden, D. York and L. Pedersen, *The Journal of Chemical Physics*, 1993, **98**, 10089–10092.

52 U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *The Journal of Chemical Physics*, 1995, **103**, 8577–8593.

53 G. Bussi, D. Donadio and M. Parrinello, *Journal of Chemical Physics*, 2007, **126**, 014101.

54 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.

55 M. Parrinello and A. Rahman, *Journal of Applied Physics*, 1981, **52**, 7182–7190.

56 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. Löhr, F. Marinelli, L. Martin-Samos, M. Masetti, R. Meyer, A. Michaelides, C. Molteni, T. Morishita, M. Nava, C. Paissoni, E. Papaleo, M. Parrinello, J. Pfaendtner, 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. Spiwok, J. Šponer, D. W. Swenson, P. Tiwary, O. Valsson, M. Vendruscolo, G. A. Voth and A. White, *Nature Methods*, 2019, **16**, 670–673.

57 G. M. Torrie and J. P. Valleau, *Journal of Computational Physics*, 1977, **23**, 187–199.

58 J. Kästner, *Umbrella sampling*, 2011.

59 W. You, Z. Tang and C.-E. A. Chang, *J Chem Theory Comput*, 2019, **15**, 2433–2443.

60 J. Schlitter, M. Engels and P. Krüger, *Journal of Molecular Graphics*, 1994, **12**, 84–89.

61 S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen and P. A. Kollman, *Journal of Computational Chemistry*, 1992, **13**, 1011–1021.

62 A. Grossfield, *WHAM - Grossfield Lab*, `http://membrane.urmc.rochester.edu/?page_id=126`.

63 B. Efron and R. Tibshirani, *An Introduction to the Bootstrap*, Chapman and Hall/CRC, 1994.

64 N. Michaud-Agrawal, E. J. Denning, T. B. Woolf and O. Beckstein, *Journal of Computational Chemistry*, 2011, **32**, 2319–2327.

65 P. V'kovski, A. Kratzel, S. Steiner, H. Stalder and V. Thiel, *Nature Reviews Microbiology*, 2021, **19**, 155–170.

66 M. Romano, A. Ruggiero, F. Squeglia, G. Maga and R. Berisio, *Cells*, 2020, **9**, 1267.

67 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.

68 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.

69 S. Lin, H. Chen, Z. Chen, F. Yang, F. Ye, Y. Zheng, J. Yang, X. Lin, H. Sun, L. Wang, A. Wen, H. Dong, Q. Xiao, D. Deng, Y. Cao and G. Lu, *Nucleic Acids Research*, 2021, **49**, 5382–5392.

70 D. D. Boehr, J. J. Arnold, I. M. Moustafa and C. E. Cameron, *C. E. Cameron*, 2014, pp. 309–333.

71 K. A. Johnson, *Biochimica et Biophysica Acta - Proteins and Proteomics*, 2010, **1804**, 1041–1048.

72 E. D. Smidansky, J. J. Arnold, S. L. Reynolds and C. E. Cameron, *Biochemistry*, 2011, **50**, 5016–5032.

73 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.