

# Supplementary Information

## 1 Methods

### 1.1 Modeling the substrate insertion complex of RdRp

Since a tertiary elongation complex of RdRp (with a full length nsp12) was captured together with RNA strands (template and primer) and the RDV analog incorporated (post-catalysis or product state; PDB: 7BV2), presumably in the active-site closed state, we built an RDV-TP insertion model (pre-catalytic) of the CoV-2 RdRp directly using this tertiary complex, only replacing the incorporated RDV analog at the 3'-end of the RNA primer strand by a pre-catalytic RDV-TP. The three  $Mg^{2+}$  ions present in the product state are taken as initial positions for our model built for.

### 1.2 Missing Residues Added

PDBID:7btf is missing the following residues in the N-terminus domain: 1-30, 51-68, 75, 103-111, 895-906, and the following in the thumb sub-domain 920-932. Missing residues were completed using MODELLER 9.24 [1] with the apo structure PDBID:6M71 which is only missing residues 1-4.

### 1.3 Protonation (details)

Propka and pdb2pqr used to predict protonation states of Histidine residues:

- HID: 75 99 113 133 256 347 355 362 439 572 599 613 810 816 882 898
- HIE: 82 295 309 381 642 650 725 752 872 892

- Residue 295 and 642 are manually selected due to their orientation with Zn ion such that the proton is not oriented near the metal.

## 2 RDV-TP Force Field

### 2.1 Atom Types

In general the majority of atom types were kept the same as the adenosine where possible from the amber force field (see **Fig S1**) [2]. The swapped and additional atoms (nitrile functional group) used the following atom types for RDV: C9:CK, C7:CK, N4:na, C6':c1, and N6':n1. Where the lower case atom types are taken from the Generalized Amber Force Field (GAFF) [3].

### 2.2 Partial Charges

During Restrained Electrostatic Potential method (RESP [4]) the 3' 5' hydroxyl atomic charges are constrained [5] to O5' = -0.6223e, H5T = 0.4295e, O3' = -0.6541e, H3T = 0.4376e, partial atomic charges for the truncated Remdesivir are generated (see **Table S1**).

## 3 Constructing the initial binding complex and docking

### 3.1 Modeling the initial binding complex

The active-site open structure of the CoV-2 RdRp for NTP initial binding was obtained from the first determined cryo-EM structure (PDB: 7BTF). Then RDV-TP was placed to the active site of the open state structure (along with the RNA template and primer strand) by aligning the RdRp structure from the tertiary RDV-TP insertion complex (the closed one constructed above) with that of the open one, and then shifting the RDV-TP and RNA strands from the tertiary complex to the open state structure accordingly. Additionally, the

Mg<sup>2+</sup> ions in the initial binding state were kept similarly as in the insertion state. Followed, the modeled structural complex would be subject to MD simulation equilibration.

### 3.2 Substrate docking to the open form RdRp

In order to test whether the above constructed initial binding complex was reasonable, we also performed docking of RDV-TP (or RTP below) and ATP as ligands onto the open structural complex of RdRp (nsp12+nsp7+ns8 and RNA together as the receptor), using AutoDock Vina software (see **Fig S2**) [6]. The receptor complex was prepared by deletion of water molecules, addition of hydrogen molecules and by computing Kollman charges. The ligands (RTP and ATP) were prepared by computing Gasteiger charges. A grid box (x=40 Å, y=40 Å, z=40 Å) is specified around the active site for the search space on the receptor within which various positions of the ligand are to be considered. An energy range of 4 and exhaustiveness of 8 were assigned. Conformations with lowest binding free energetic scores are considered most stable or optimal.

We further conducted equilibrium MD simulations on the optimized docking complex of ATP. The results show that even upon the docking and equilibration, the stabilized configurations of ATP still converge to be very close to the initial modeled ones. Interestingly, even we chose another reference structure in docking (e.g. using the pre-insertion structure of T7 RNAP [7]), we could still obtain a docking configuration overlapping well with the insertion ATP. Hence, it justifies that the constructed ATP and RTP initial binding or the active-site open RdRp structural complexes are reasonable. A further comparison show that ATP binding configuration in our constructed open form RdRp complex is similar to that being captured in the PV RdRp.

## 4 Selecting Reference Structures

The reference states used for the reaction coordinate or the implementation of TMD need to be close to equilibrium but not at equilibrium, since we want to sample both sides of equilibrium region along the RC, while the reference structures correspond to the two ends of the RC( $-RC_{\max}$ ,  $+RC_{\max}$ ), with  $RC_{\max} = \delta RMSD(X_{\text{Open ref}}, X_{\text{Closed ref}})$ . The reference

structures or states are selected using the first 50ns of the unrestrained NPT simulations, and the correspondingly defined RCs for the open and closed equilibrated structures need to satisfy the conditions below:

$$\begin{aligned} \delta RMSD(X_{\text{Open equi}}) &= RMSD(X_{\text{Open equi}}, X_{\text{Open ref}}) - \\ &\quad RMSD(X_{\text{Open equi}}, X_{\text{Closed ref}}) \\ \delta RMSD(r_{\text{Closed equi}}) &= RMSD(X_{\text{Closed equi}}, X_{\text{Open ref}}) - \\ &\quad RMSD(X_{\text{Closed equi}}, X_{\text{Closed ref}}) \end{aligned} \tag{1}$$

Where the requirement is:

$$\begin{aligned} -RC_{max} &< RC(X_{\text{Open equi}}) < 0 \\ 0 &< RC(X_{\text{Closed equi}}) < +RC_{max} \end{aligned} \tag{2}$$

Where the RC is  $\delta RMSD$  specified in equation (1). The force constants and other setting parameters used in the TMD and umbrella sampling simulations are listed in Table S2.

## 5 Constructing PMF's for ATP and RDV-TP insertion paths

ATP simulations take longer time to converge (less stabilized): in one case  $\sim 110-130\text{ns}$ , the other case,  $\sim 150-170\text{ns}$  (see **Fig. S13**); than RDV  $\sim 50\text{ns}$ . The longer convergence time ATP simulations take indicate more time required for equilibration than the RDV-TP simulations. Correspondingly, we removed more data for ATP (20ns) than RDV-TP (10ns) for constructing the potential of mean force (PMF).

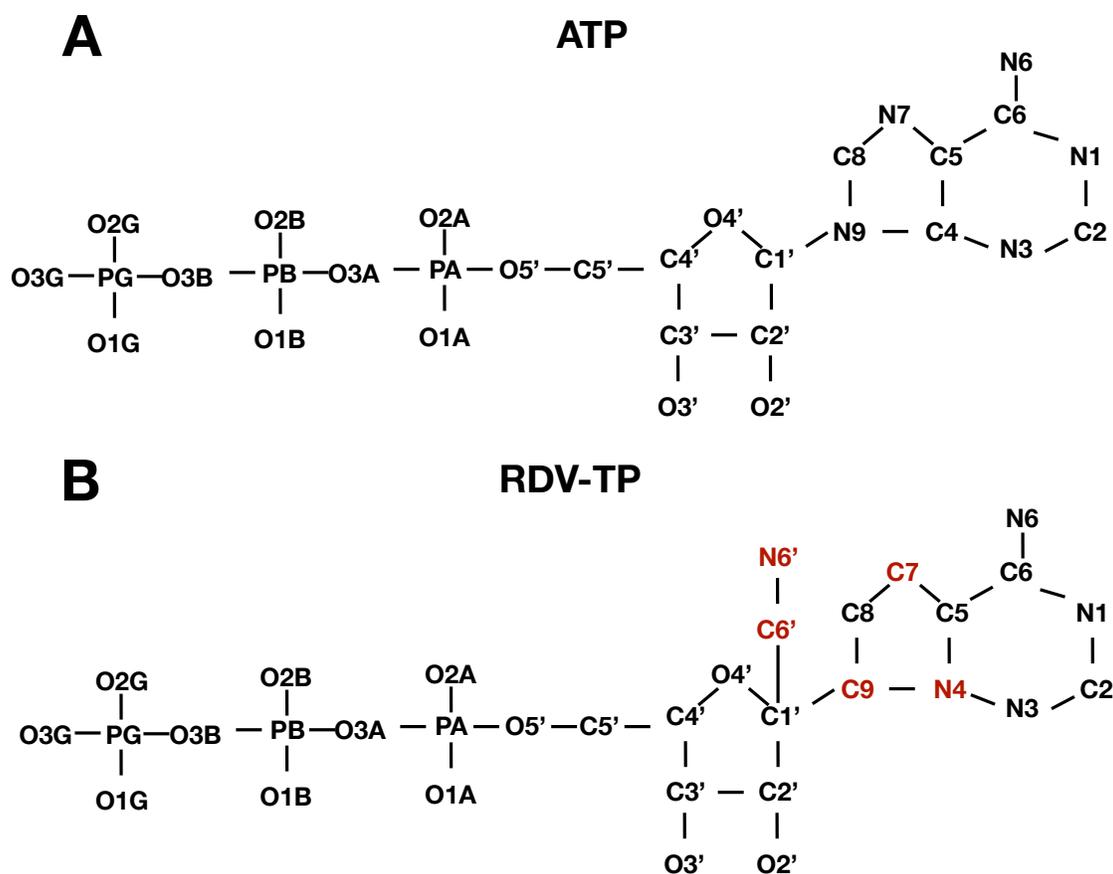


Figure S1: Comparing **A.** Adenosine Triphosphate (ATP) and **B.** Remdesivir Triphosphate (RDV-TP) heavy atom molecular structures. Hydrogens are omitted for a clear representation. Atoms colored in red highlight the differences in RDV from ATP. For partial charge calculation the RDV-TP is truncated at the O5' with the addition of a hydrogen H5T. H3T is the hydrogen atom bonded to the O3' oxygen (see **Table S1**).

Table S1: Summary of Partial Charges used for the RDV-TP force field compared with ATP. Charges are separated by section of the NTP.

Atom (PolyP)	ATP	RTP	Atom (Sugar)	ATP	RTP	Atom (Base)	ATP	RTP
O5'	-0.59870	-0.59870	C5'	0.05580	0.039981	N9	-0.02510	N.A.
PA	1.25320	1.25320	H5'1	0.06790	0.085276	C9	N.A.	-0.118619
O1A	-0.87990	-0.87990	H5'2	0.06790	0.085276	C8	0.20060	-0.228326
O2A	-0.87990	-0.87990	C4'	0.10650	0.083427	H8	0.15530	0.199975
O3A	-0.56890	-0.56890	H4'	0.11740	0.065203	N7	-0.60730	N.A.
PB	1.38520	1.38520	O4'	-0.35480	-0.332867	C7	N.A.	-0.259805
O1B	-0.88940	-0.88940	C1'	0.03940	0.130365	C5	0.05150	-0.394730
O2B	-0.88940	-0.88940	H1'	0.20070	N.A.	C6	0.70090	1.014028
O3B	-0.53220	-0.53220	C6'	N.A.	0.461023	N6	-0.90190	-1.042464
PG	1.26500	1.26500	N6'	N.A.	-0.505959	H61	0.41150	0.443695
O1G	-0.95260	-0.95260	C3'	0.20220	0.329872	H62	0.41150	0.443695
O2G	-0.95260	-0.95260	H3'	0.06150	0.076195	N1	-0.76150	-0.863108
O3G	-0.95260	-0.95260	C2'	0.06700	-0.074121	C2	0.58750	0.630021
			H2'1	0.09720	0.146103	H2	0.04730	0.076123
			O2'	-0.61390	-0.626760	N3	-0.69970	-0.744123
			HO'2	0.41860	0.462358	C4	0.30530	N.A.
			O3'	-0.65410	-0.65410	N4	N.A.	0.603011
			H3T	0.43760	0.437600			

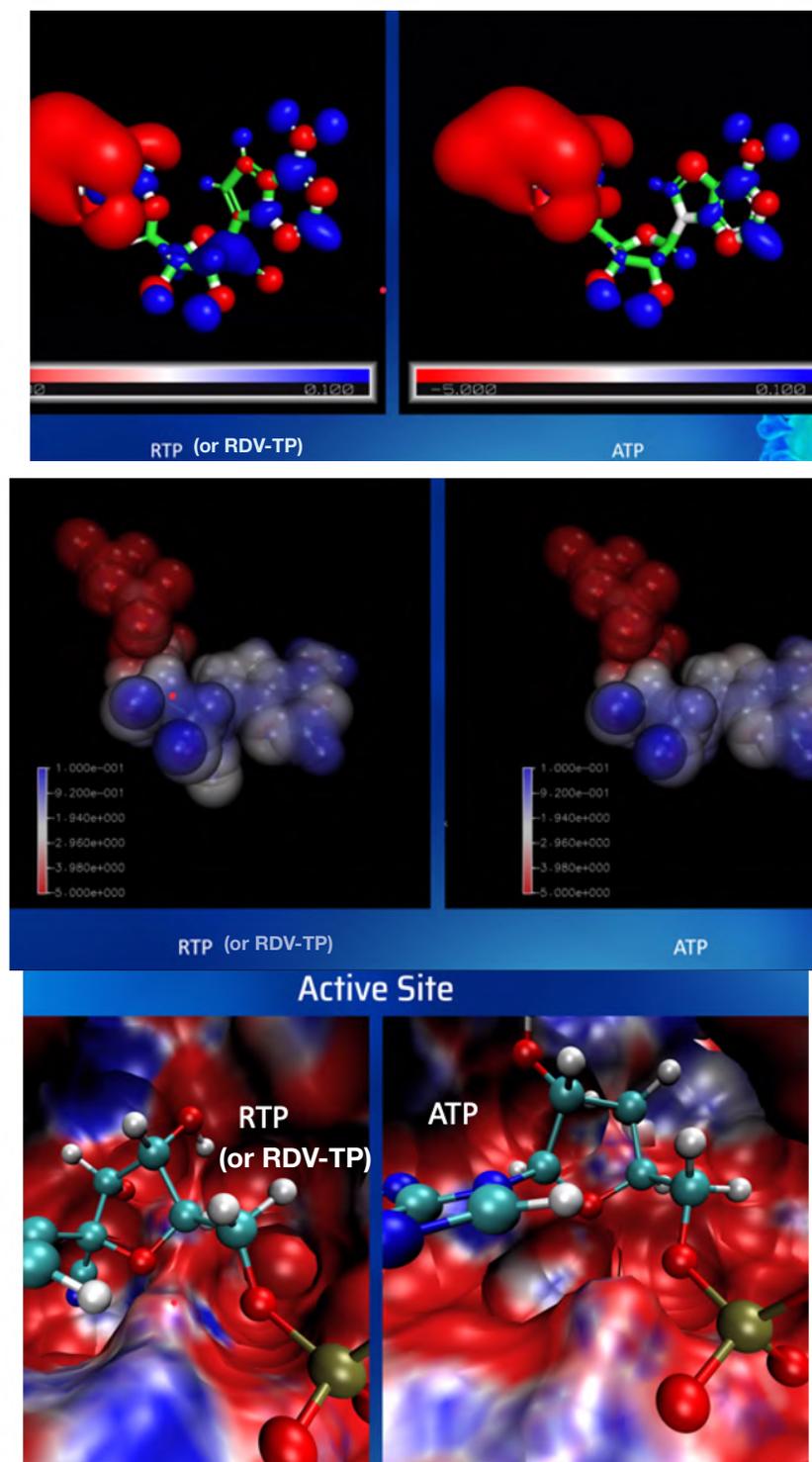


Figure S2: The electrostatic potential generated by solving the Poisson-Boltzmann equation using the APBS solver [8]. The potential (top) due to partial charges assigned for RDV-TP (or RTP) in comparison with that of ATP (middle). The potentials around the active site of the CoV-2 RdRp (PDB: 7BV2) with RDV-TP and ATP inserted (bottom), respectively.

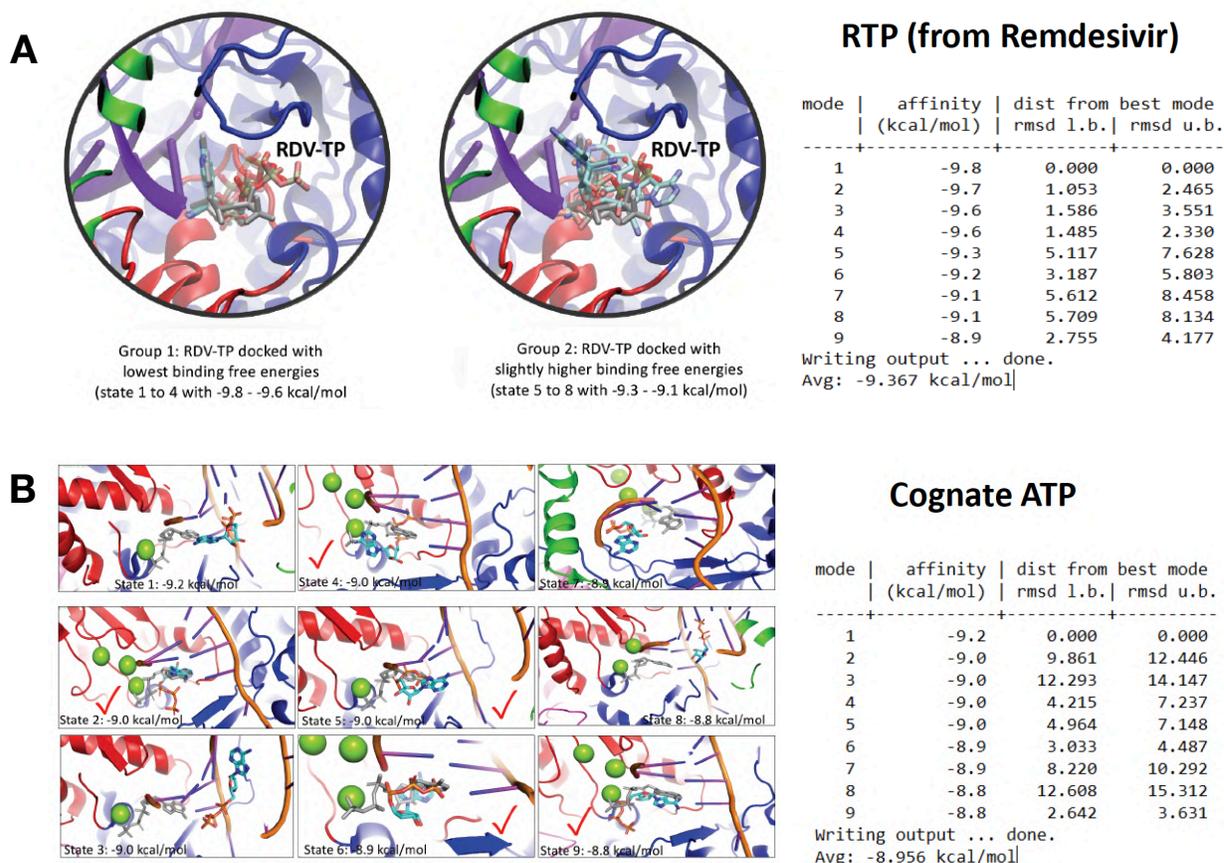


Figure S3: Docking of RDV-TP (or RTP) and ATP onto a modeled initial binding (active-site open) structure of SARS-CoV-2 RdRp (PDB: 7BTf). **A.** RDV-TP docking show comparatively stabilized docking structures (grouped into two). The palm, fingers, thumb subdomains are shown in red, blue, and green, and RNA in violet. The modeled RDV-TP (positioning from PDB: 7BV2) is shown in gray, and the docking structures of RDV-TP are shown with colored atoms. **B.** ATP docking shows diverse configurations and less stabilized configurations. The obtained docking energetics are listed on the right side for both systems (using AutoDock) [6]

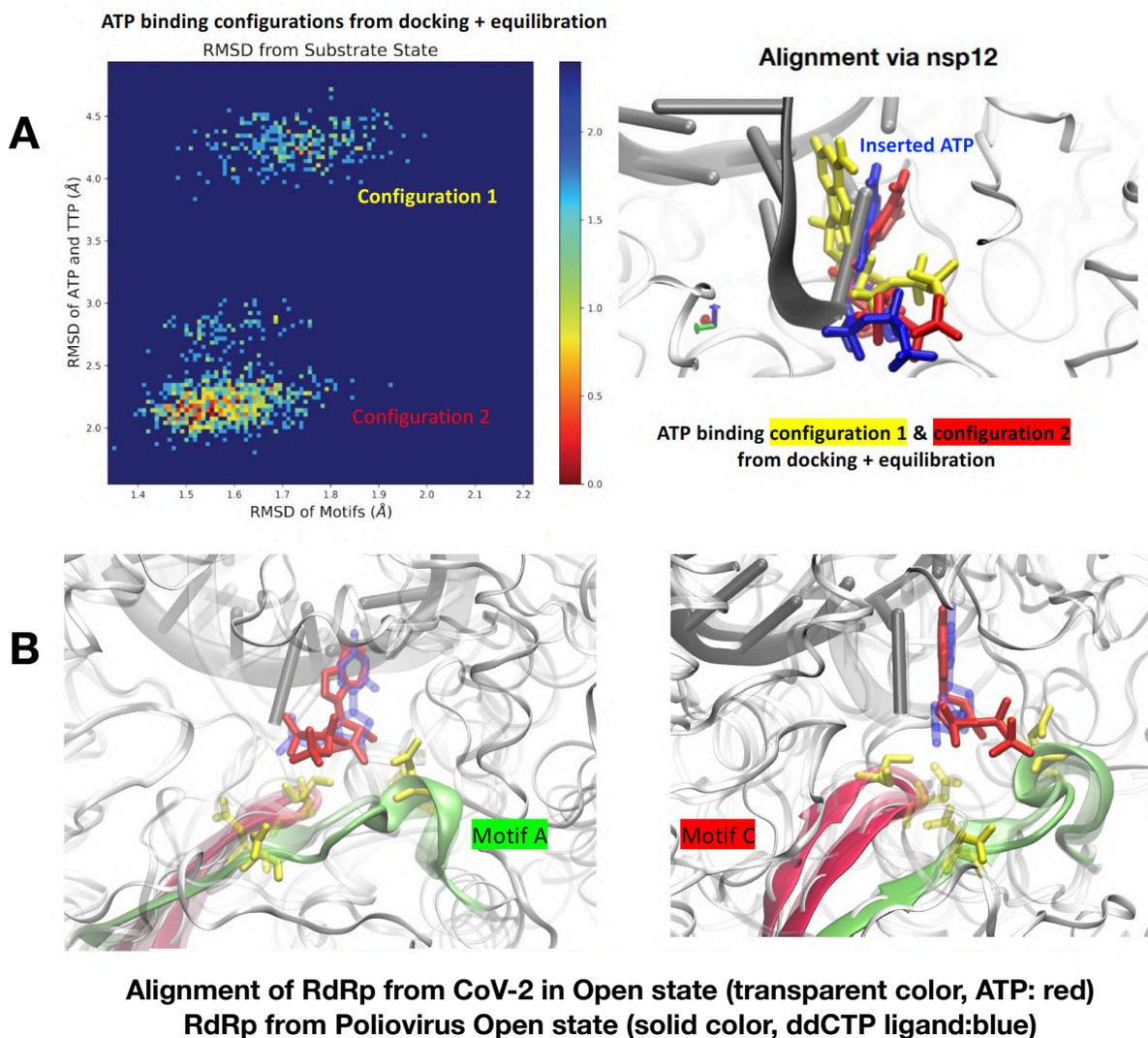


Figure S4: Examining ATP binding configurations to the open form active site CoV-2 RdRp. **A.** The MD equilibration of the optimal docking complex of ATP to the CoV-2 RdRp structure. The equilibrated configurations were measured by RMSDs for both structural motifs (A-G) and ATP+template nt (uracil) with respect to the substrate insertion complex. Two dominant configurations of ATP have been identified, both of which are quite close to the insertion configuration (as our modeled open or initial binding complex of ATP to the RdRp, see Methods 2.1). **B.** The alignment of our modeled ATP bound open equilibrated form CoV-2 RdRp with that of the poliovirus (PV) RdRp, shown in two views for better visualization.

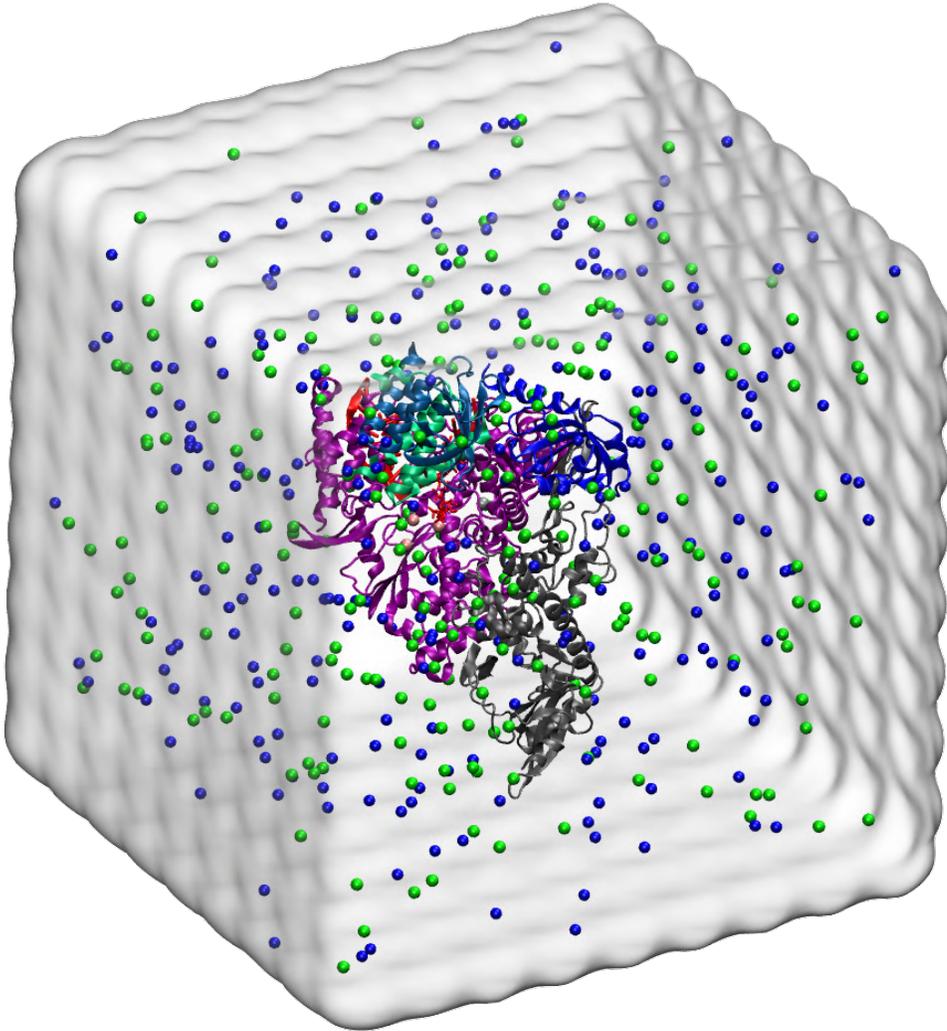


Figure S5: All-atom molecular dynamics simulation box. The size of the box on average: 15.7nm x 15.7 nm x 15.7 nm, containing an average of 382,000 atoms.

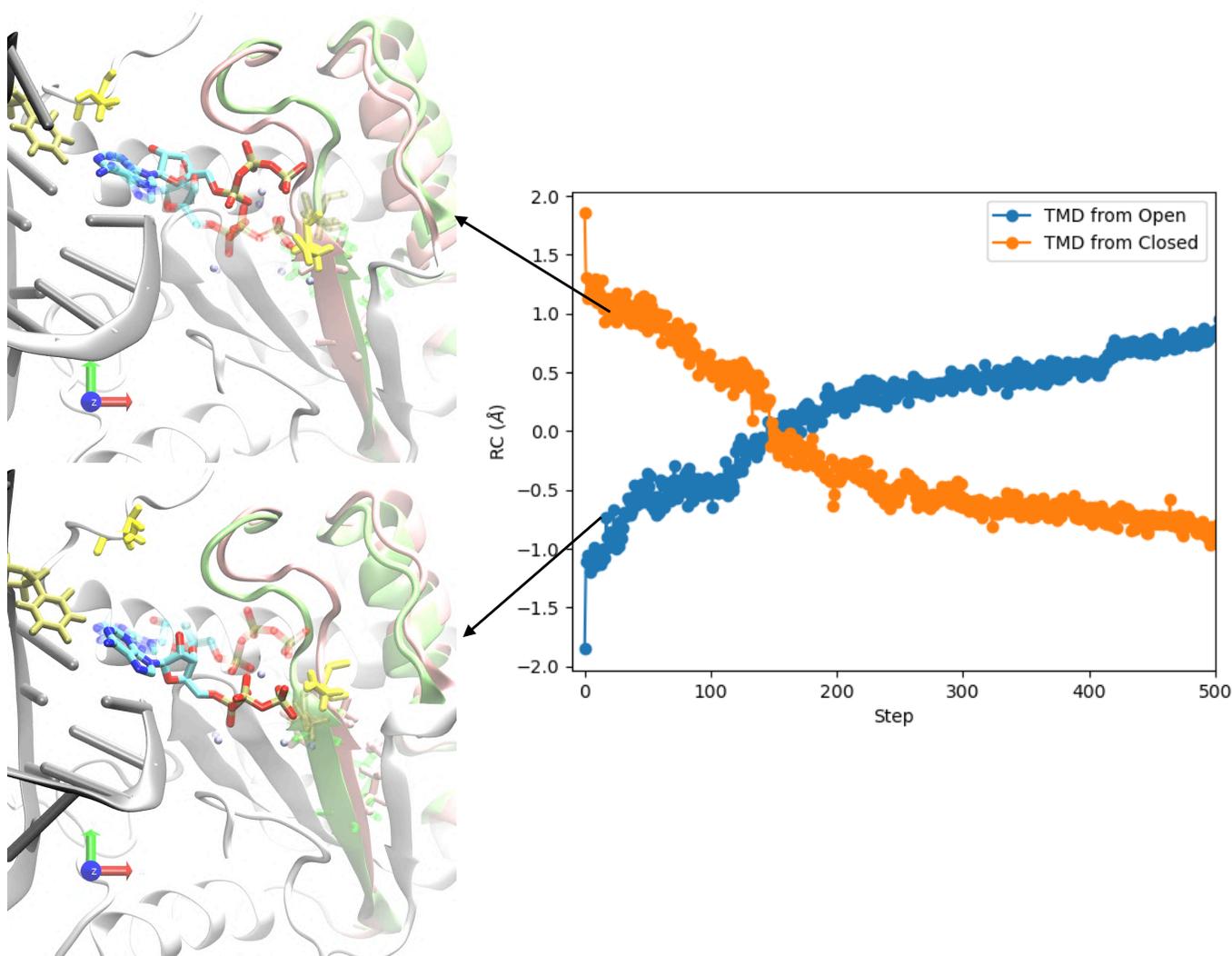


Figure S6: Implementation of targeted molecular dynamics (TMD) simulations for constructing NTP insertion path to be utilized in the umbrella sampling simulations. (Left) The initial and final structures of each respective paths: backward (bottom) and forward (upper). With motifs A/D colored pink for the starting structure and green for the target structure. Representations are colored to compare with PV RdRp in [9]. (Right) The implementation of the TMD simulations forward and backward. Where TMD from open refers to starting from the initial binding complex (forward path), and TMD from closed is starting from the insertion complex (backward path). Structures are selected every 0.1 Å from the two paths until they meet in the middle along the reaction coordinate (RC). Such a constructed path is then used for the umbrella sampling simulation.

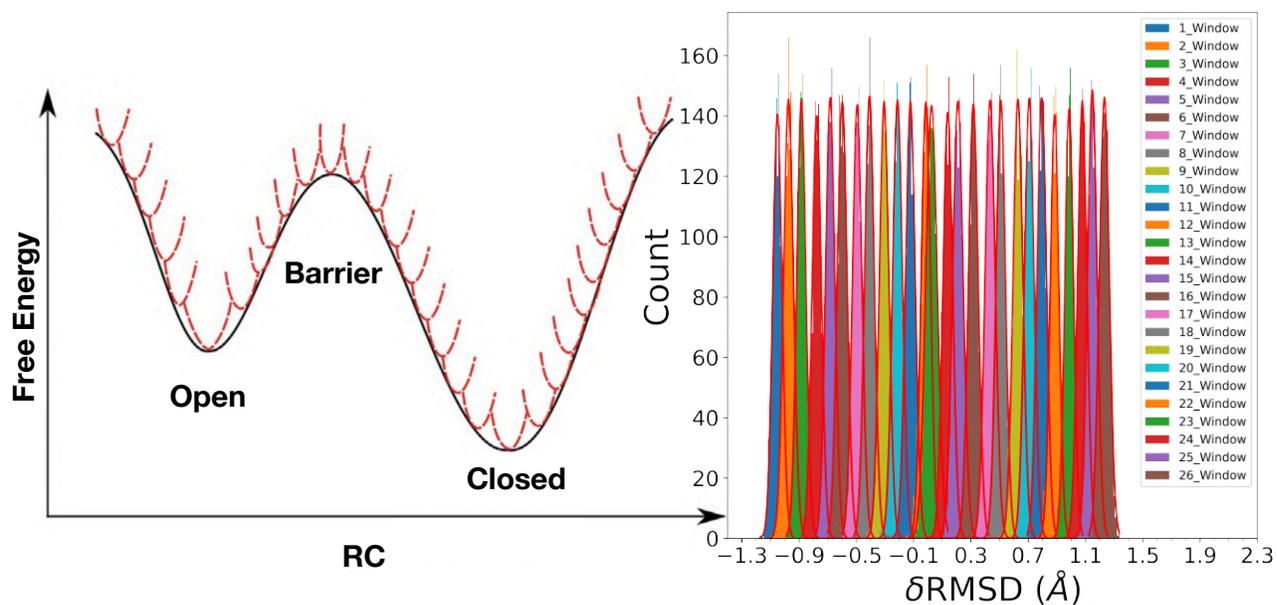


Figure S7: Conducting the umbrella sampling simulations for the NTP insertion. (Left) The schematics of the umbrella sampling simulation strategies (figure adapted from [10]). (Right) The overlap of simulated windows, where the RC is centered every  $0.1\text{\AA}$ , with the initial simulation structure taken from the forward/backward TMD paths (see **Fig S6**).

Table S2: Summary of target MD and Umbrella Sampling parameters. The force constant used from the TMD simulations were carried over and used for the respective umbrella sampling simulations. Large force constants were used for the ATP simulations and smaller ones for RDV-TP simulations. Where the () in RDV-TP or RTP systems indicate the initial binding structure (open for the active site open state).

RC	Force Constant $\left(\frac{kcal}{mol\text{\AA}^2}\right)$	RC Range( $\text{\AA}$ )	Number of Windows
Motifs + ATP	501	-1.2 to 1.3	27
Motifs + ATP + Template	501	-1.1 to 1.3	26
Motifs + RTP(Open Stacking)	125	-1.0 to 1.0	21
Motifs + RTP(Open Stacking) + Template	125	-1.6 to 1.6	34
Motifs + RTP(Open Base-pairing) + Template	250	-1.5 to 1.5	32

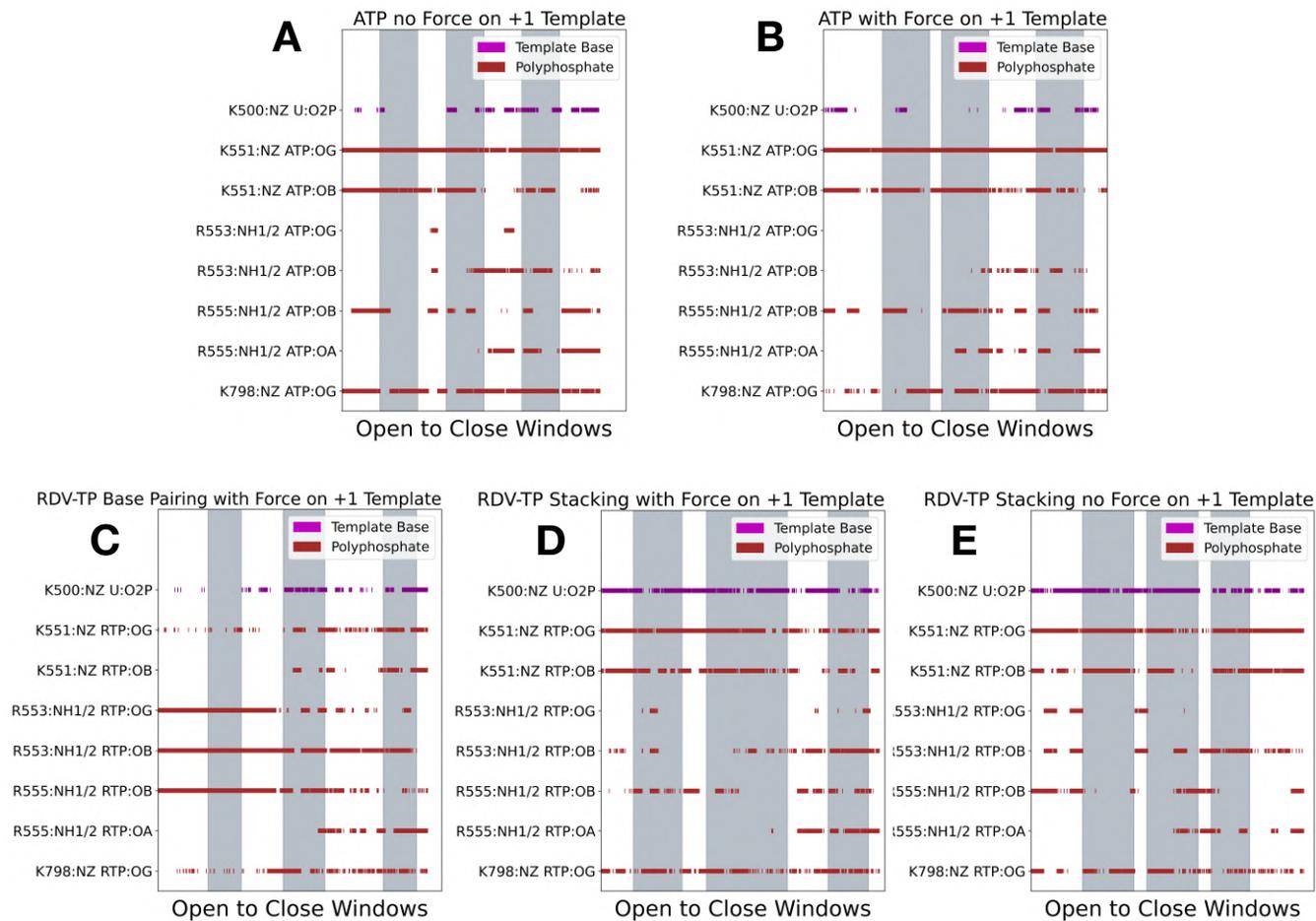


Figure S8: Salt-bridge electrostatic interactions with ATP/RDV-TP triphosphates. Here we can identify the positively charged residues (Lys and Arg) which can form salt bridges with the negatively charged oxygen's along the polyphosphate. Distances are measured from the positive charge center (NZ nitrogen in Lys and CZ carbon in Arg) and the negative charge (O1G,O2G,O3G,O1B,O2B,O1A, and O2A in the NTP or O1P and O2P in the template backbone), if the charges are less than 5Å [11] a salt bridge is identified.

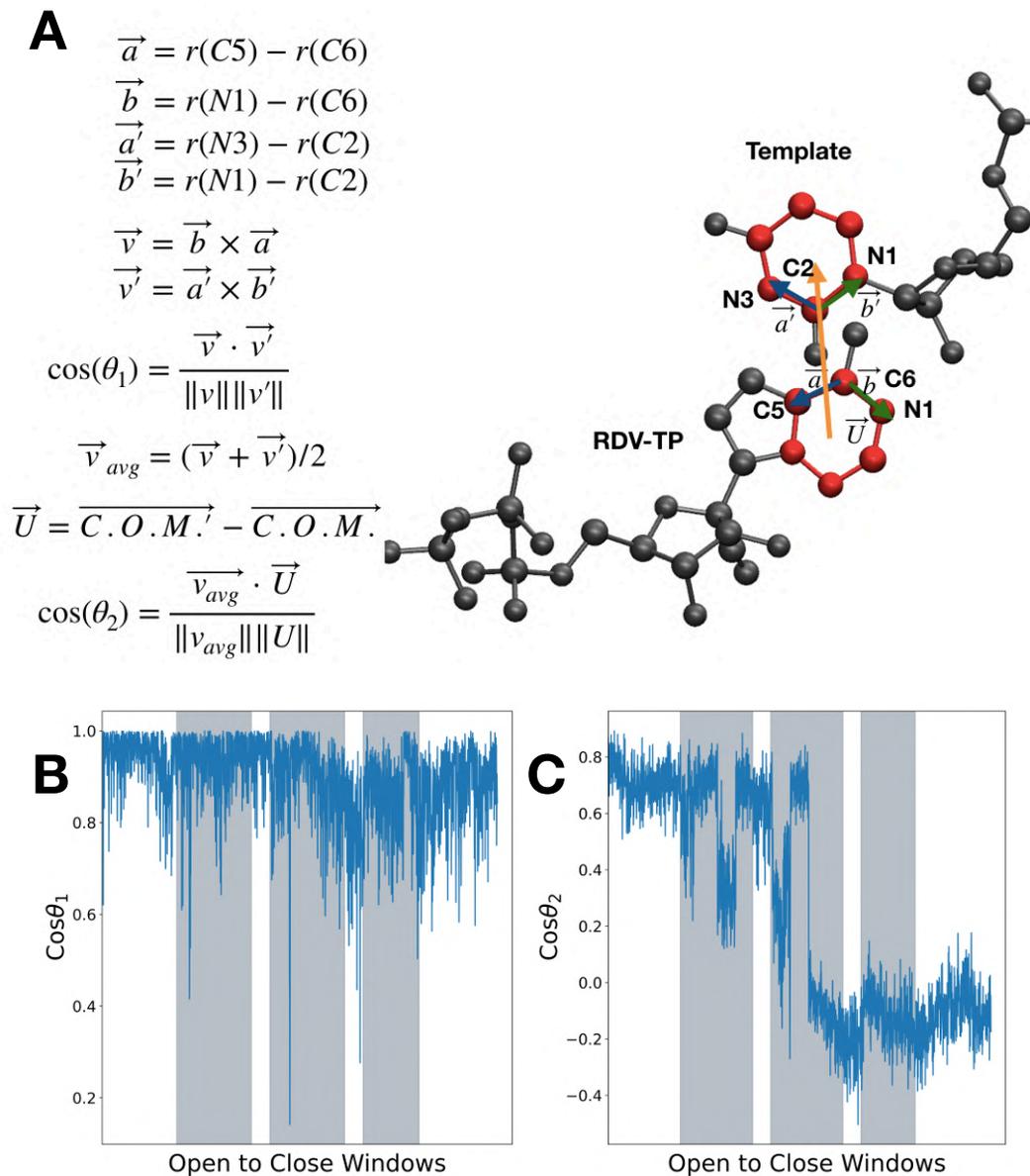


Figure S9: Schematics for base-stacking measurements. **A** The stacking is determined by measuring if the two base rings are parallel with reasonable overlapping. **B** The first  $\cos \theta_1$  is calculated from atoms within the six numbered rings in the bases. **C** The second  $\cos \theta_2$  is calculated via considering the center of mass (C.O.M.) of the two six numbered rings. Reasonable base stacking is formed when  $\cos \theta_1 > 0.8$  ( $\theta_1$  close to zero or the two rings being parallel) and  $\cos \theta_2 > 0.6$  ( $\theta_2$  not far from zero or the two rings overlap). Measurements shown are from the RDV-TP with base stacking (without force on template nt) insertion.

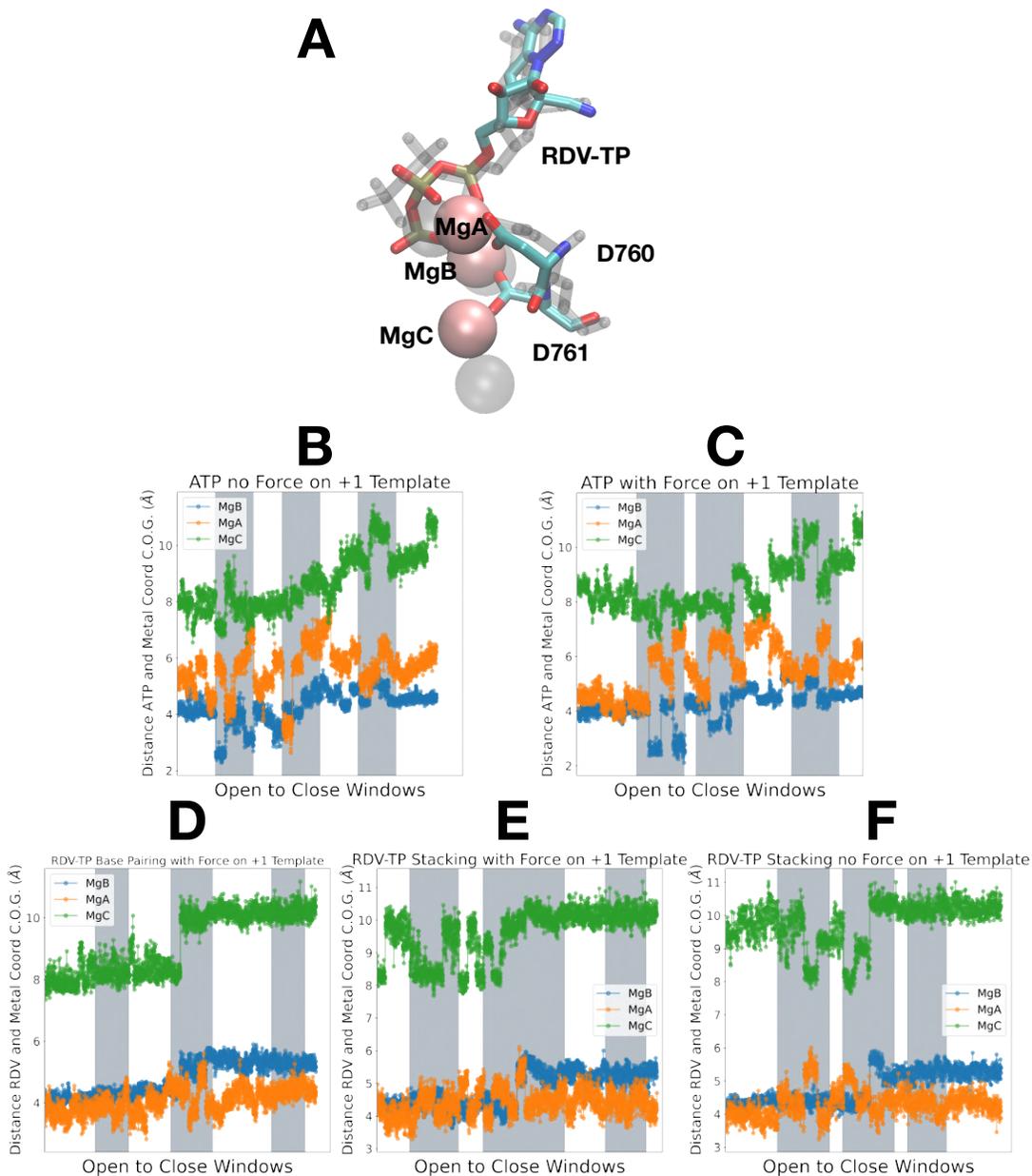


Figure S10: **A** The modeled insertion state structure of RDV-TP (grey transparent) aligned with that of the equilibrated insertion state one (in color). MgA is in coordination with residues D760 and D761. MgB is coordinated by the  $\beta$  and  $\gamma$  phosphate oxygens. **B-F** Distances measured between the center of geometry (C.O.G.) of the NTP and  $Mg^{2+}$  ions in the umbrella sampling simulations (from open to barrier and to the closed state). MgA is in coordination with the catalytic D760 and D761 residues as well as the 3' end primer backbone phosphate. MgB is in coordination with the phosphate group ( $\beta$  &  $\gamma$  phosphate oxygens) [12]. MgC is unlikely to be involved in catalysis or product release as it stays comparatively far from the NTP.

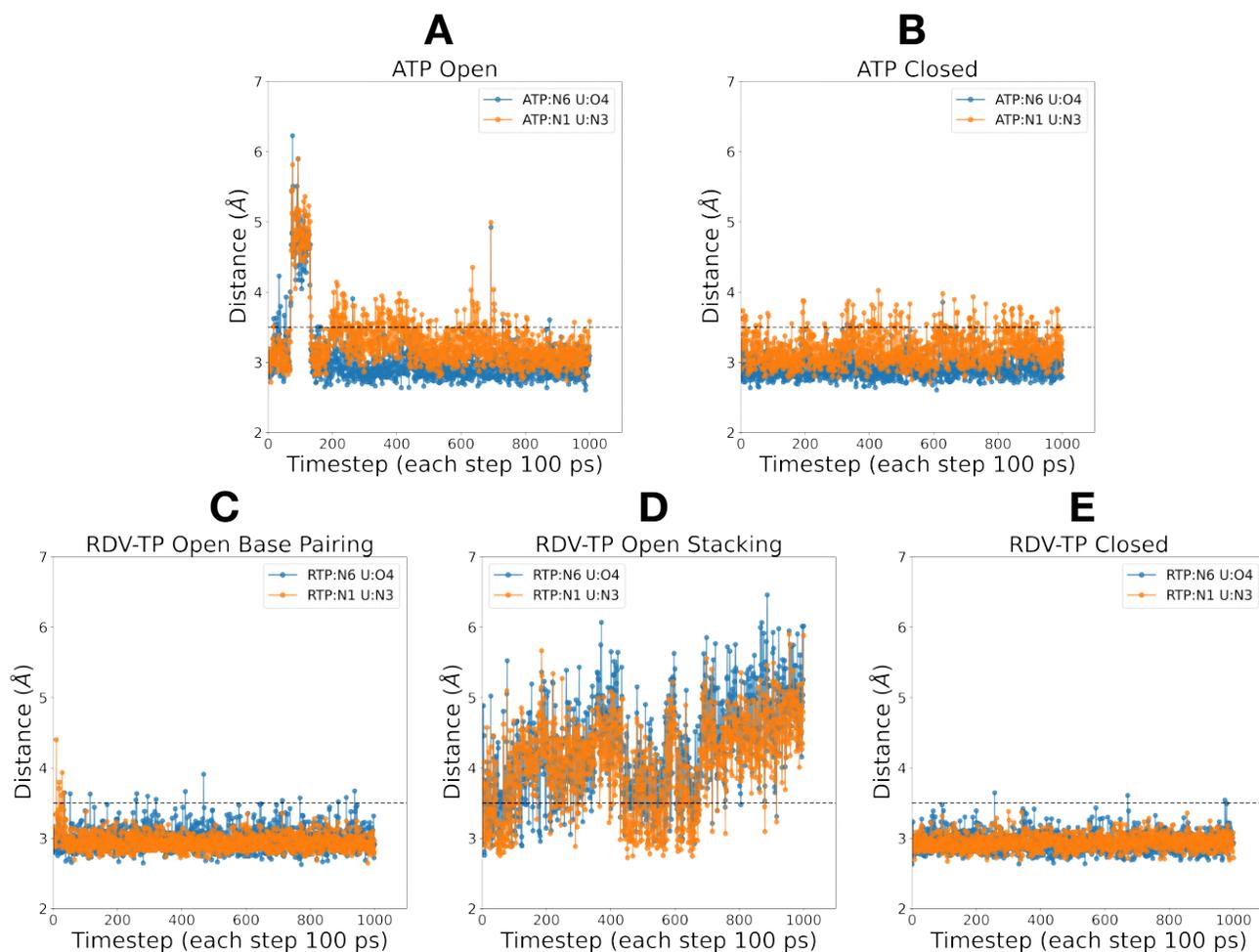


Figure S11: Expected hydrogen bond (HB) distance between ATP/RDV-TP and the +1 template Uracil (**A-E**). The distance pairs measured are the heavy atoms from the nucleotide triphosphate N6 with U:O4 and N1 and U:N3. The dashed black line indicate the cutoff (3.5 Å) for a HB. The NTPs in the insertion complexes (**B&E**) form significantly more stable HB than in in the initial binding forms. An RDV-TP in initial-binding forms a base stacking configuration with the template nt (**D**), in which hydrogen bonds are rarely formed.

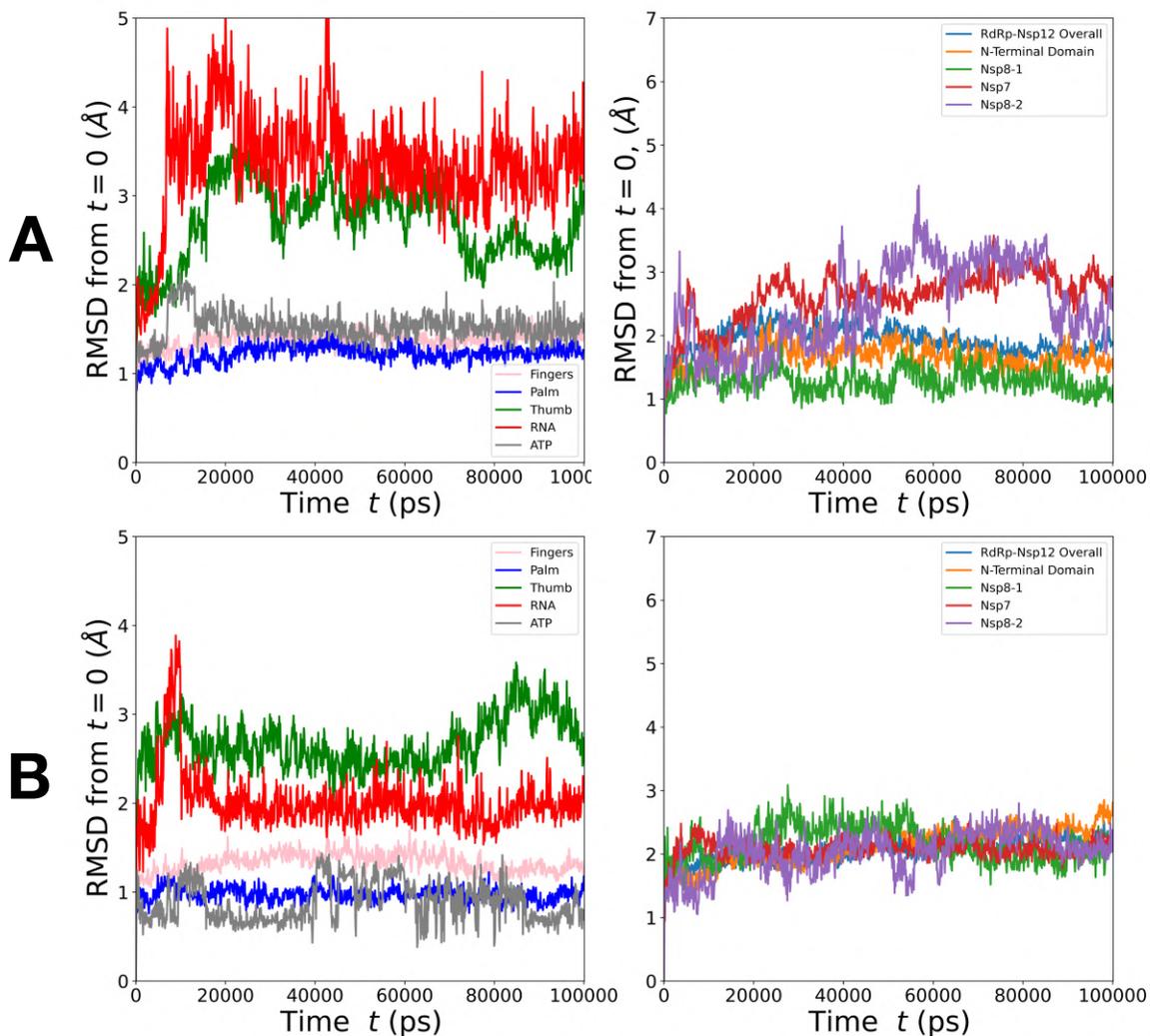


Figure S12: Measured RMSD from equilibration simulations. (Left) Subdomains, RNA, and ATP. (Right) Cofactors (ns7 and nsp8), Nsp12 and N-Terminus domain. **A** ATP initial binding complex. **B** ATP insertion complex. The insertion complex appears to be more stable than the initial binding complex.

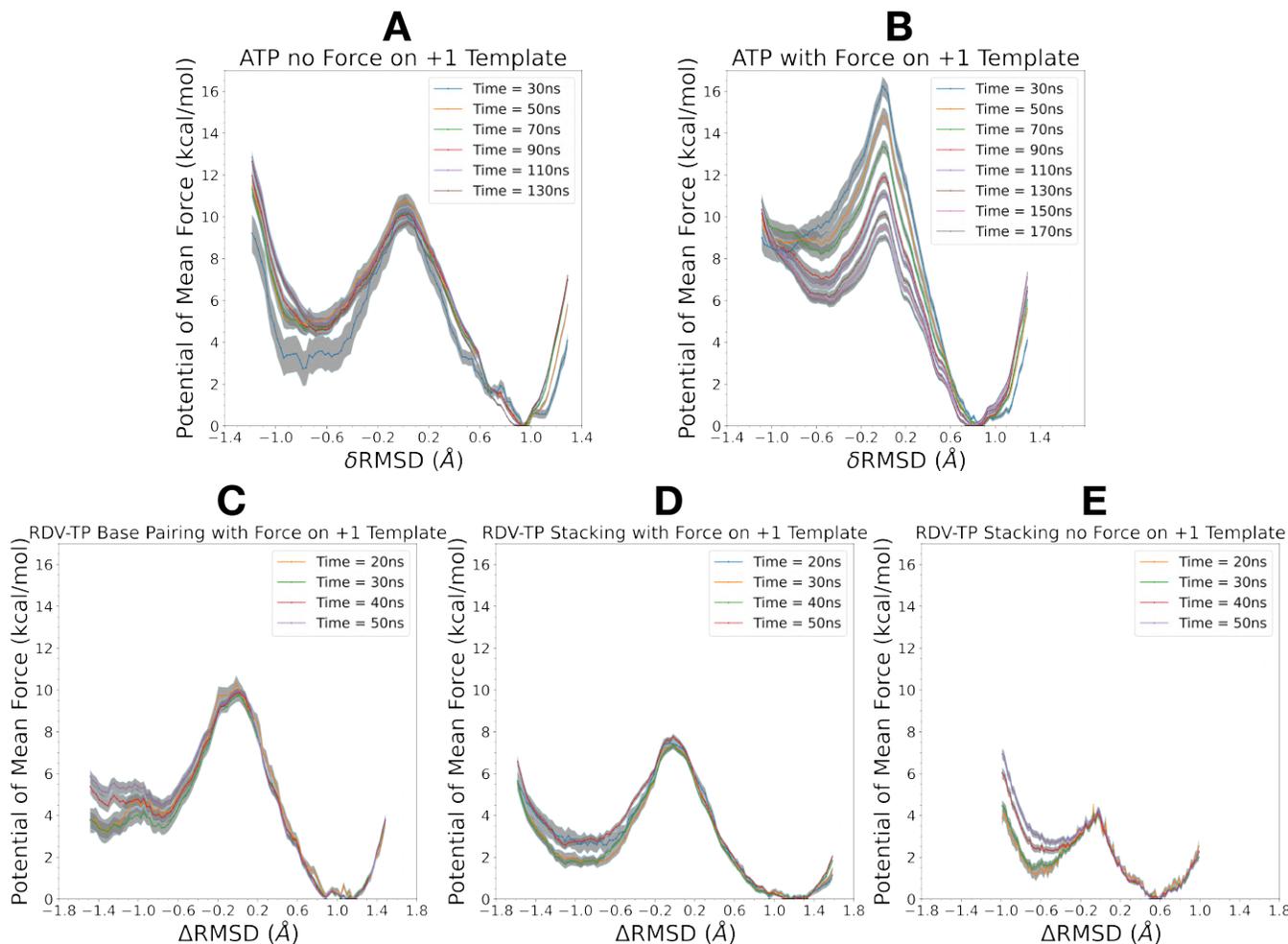


Figure S13: Convergence plots of all PMFs constructed with bootstrapping error analysis for each set of data [13,14]. As more data are accumulated with the extended simulations, the PMF further converges. Early data collected is removed as for pre-equilibration, 20ns for ATP and 10ns for RDV-TP systems. **A&B** ATP PMF's with no force on template and with force on template, respectively. **C** RDV-TP base pairing with force on +1 template nt. **D** RDV-TP stacking with force on + 1 template nt. **E** RDV-TP stacking with no force on + 1 template. **D-E** Only 50ns of data from each window was needed to reach convergence.

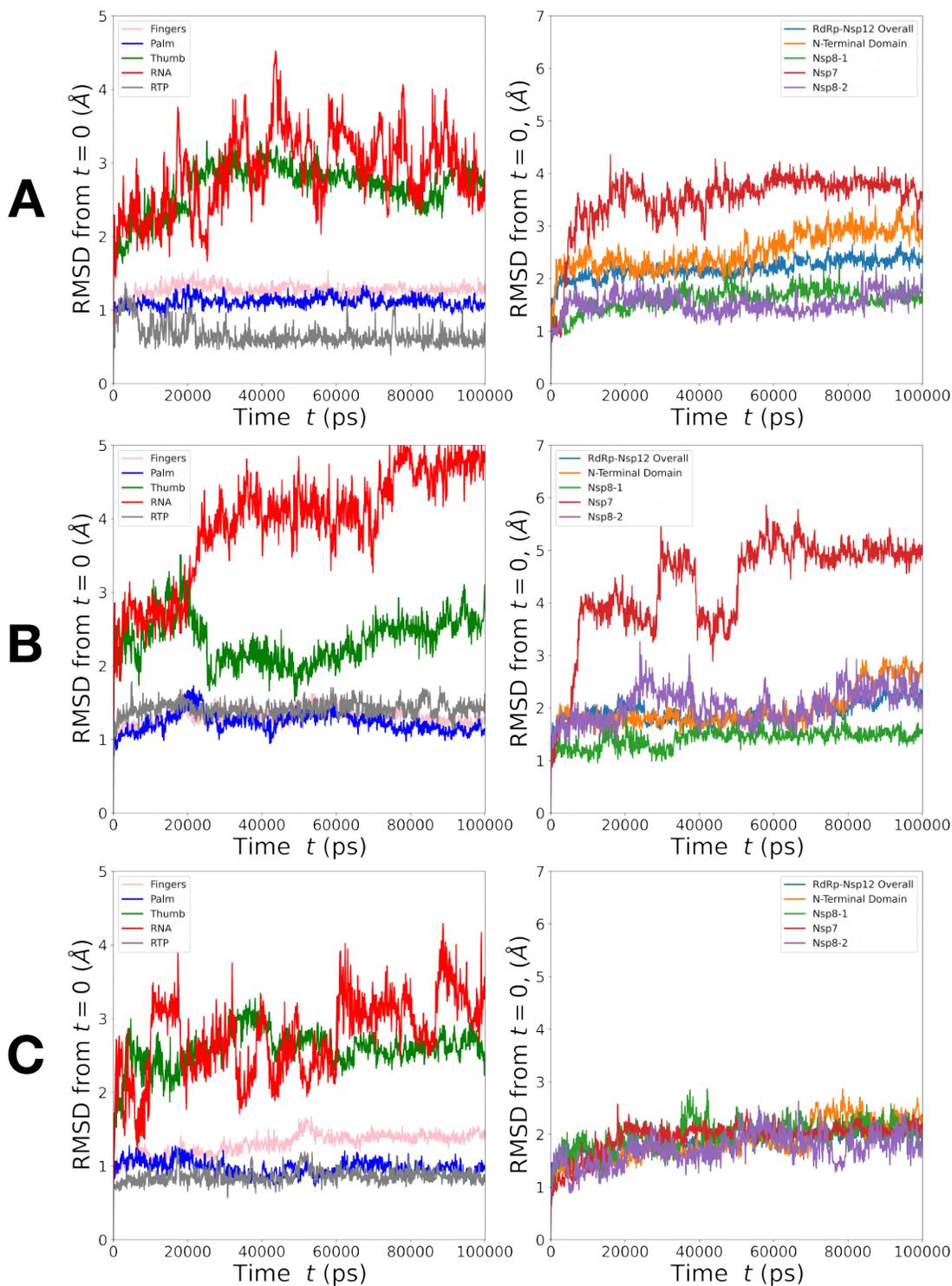


Figure S14: Measured RMSD from RTP (RDV-TP) equilibration simulations. (Left) Subdomains, RNA, and NTP. (Right) Cofactors (nsp and nsp8), Nsp12 and N-Terminus domain. **A** RDV-TP initial binding base pairing configuration. **B** RDV-TP initial binding stacking configuration. **C** RDV-TP insertion complex. The RDV-TP insertion complex appears to be more stable than the initial binding configurations.

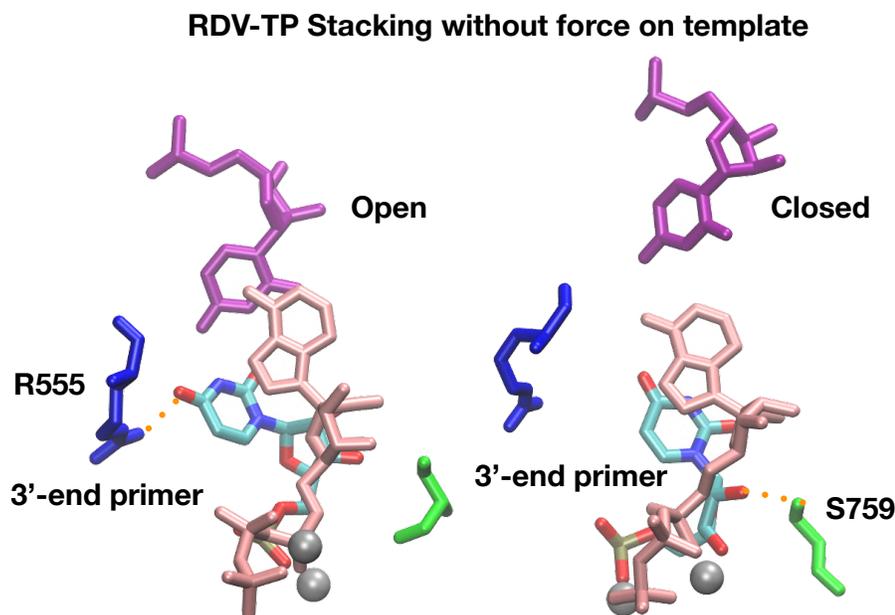
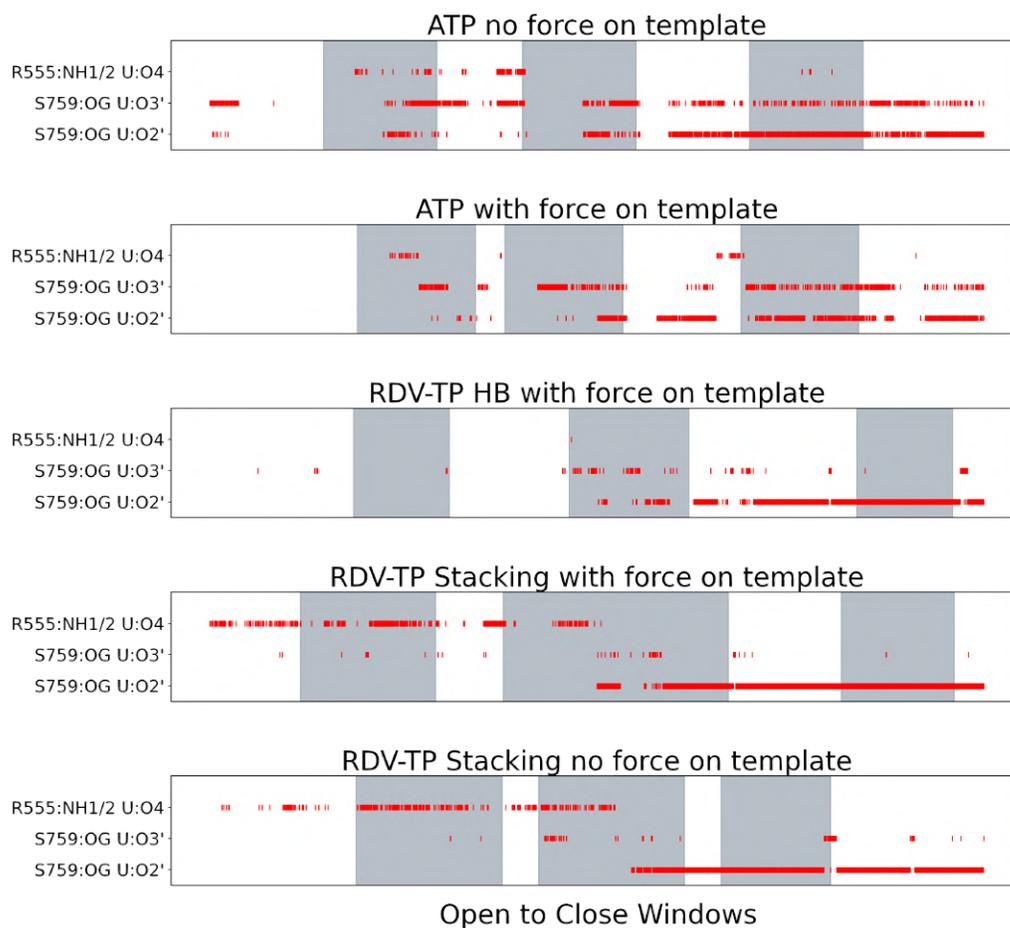


Figure S15: Hydrogen bond analyses around the 3'-end RNA primer nt (uracil base). (Upper) S759 forms frequent HBs with the 3'-end primer into both RDV and ATP insertion states, but only does that for ATP initial binding but not RDV-TP initial binding systems. R555 forms stable HB with 3'-end primer O4 oxygen in the RDV stacking initial binding state but no stable HB for RDV base pairing. (Lower) Open and closed states for RDV-TP from stacking insertion (no force on the template) showing the R555 and S759 HBs on the 3' end primer, respectively.

## Movie S1

Forward path (from open to closed) from the TMD for ATP insertion (with force on the template nt). The representation follows the same described in **Fig S6**.

## Movie S2

Forward path (from open to closed) from the TMD for RDV-TP (initial binding stacking and no force on template nt). The representation follows the same described in **Fig S6**.

## References

- [1] Andrej Šali and Tom L. Blundell. Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology*, 234(3):779–815, 12 1993.
- [2] Ivan Ivani, Pablo D. Dans, Agnes Noy, Alberto Pérez, Ignacio Faustino, Adam Hospital, Jürgen Walther, Pau Andrio, Ramon Goñi, Alexandra Balaceanu, Guillem Portella, Federica Battistini, Josep Lluís Gelpí, Carlos González, Michele Vendruscolo, Charles A. Laughton, Sarah A. Harris, David A. Case, and Modesto Orozco. Parmbsc1: A refined force field for DNA simulations. *Nature Methods*, 13(1):55–58, 12 2015.
- [3] Junmei Wang, Romain M. Wolf, James W. Caldwell, Peter A. Kollman, and David A. Case. Development and testing of a general Amber force field. *Journal of Computational Chemistry*, 25(9):1157–1174, 7 2004.
- [4] Junmei Wang, Piotr Cieplak, and Peter A Kollman. How Well Does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules? Keywords: additive force field; nonadditive force field; restrained electrostatic potential (RESP); torsion. *Journal of Computational Chemistry*, 21(12):1049–1074, 2000.
- [5] Christopher I. Bayly, Kenneth M Merz, David M Ferguson, Wendy D Cornell, Thomas Fox, James W Caldwell, Peter A Kollman, Piotr Cieplak, Ian R Gould, and David C Spellmeyer. A Second Generation Force Field for the Simulation of Proteins, Nucleic

- Acids, and Organic Molecules. *Journal of the American Chemical Society*, 117(19):5179–5197, 1995.
- [6] Garrett M. Morris, Huey Ruth, William Lindstrom, Michel F. Sanner, Richard K. Belew, David S. Goodsell, and Arthur J. Olson. Software news and updates AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry*, 30(16):2785–2791, 12 2009.
- [7] Dmitry Temiakov, Vsevolod Patlan, Michael Anikin, William T. McAllister, Shigeyuki Yokoyama, and Dmitry G. Vassylyev. Structural basis for substrate selection by T7 RNA polymerase. *Cell*, 116(3):381–391, 2 2004.
- [8] Elizabeth Jurrus, Dave Engel, Keith Star, Kyle Monson, Juan Brandi, Lisa E. Felberg, David H. Brookes, Leighton Wilson, Jiahui Chen, Karina Liles, Minju Chun, Peter Li, David W. Gohara, Todd Dolinsky, Robert Konecny, David R. Koes, Jens Erik Nielsen, Teresa Head-Gordon, Weihua Geng, Robert Krasny, Guo Wei Wei, Michael J. Holst, J. Andrew McCammon, and Nathan A. Baker. Improvements to the APBS biomolecular solvation software suite. *Protein Science*, 27(1):112–128, 1 2018.
- [9] Grace Campagnola, Seth McDonald, Stéphanie Beaucourt, Marco Vignuzzi, and Olve B Peersen. Structure-Function Relationships Underlying the Replication Fidelity of Viral RNA-Dependent RNA Polymerases. *Journal of Virology*, 89(1):275–286, 2015.
- [10] Qinghua Liao. Enhanced sampling and free energy calculations for protein simulations. In *Progress in Molecular Biology and Translational Science*, volume 170, pages 177–213. Elsevier B.V., 1 2020.
- [11] Andrew S. Thomas and and Adrian H. Elcock\*. Molecular Simulations Suggest Protein Salt Bridges Are Uniquely Suited to Life at High Temperatures. *Journal of the American Chemical Society*, 126(7):2208–2214, 2 2004.
- [12] T A Steitz and J A Steitz. A general two-metal-ion mechanism for catalytic RNA. *Proceedings of the National Academy of Sciences*, 90(14):6498–6502, 7 1993.

- [13] Bradley Efron and R.J. Tibshirani. *An Introduction to the Bootstrap*. Chapman and Hall/CRC, 5 1994.
- [14] Chunhong Long, E. Chao, Lin Tai Da, and Jin Yu. Determining selection free energetics from nucleotide pre-insertion to insertion in viral T7 RNA polymerase transcription fidelity control. *Nucleic Acids Research*, 47(9):4721–4735, 5 2019.