Supplementary Information

Unveiling the "Template-dependent" Inhibition on the Viral Transcription of SARS-CoV-2

Xueying Luo[1,2], Xiaowei Wang[3], Yuan Yao[3], Xin Gao[4,5], Lu Zhang[1,2,6]*

[1] State Key Laboratory of Structural Chemistry, Fujian Institute of Research on the Structure of Matter, Chinese Academy of Sciences, 350002, Fuzhou, Fujian, China

[2] University of Chinese Academy of Sciences, 100049, Beijing, China

[3] Department of Chemical and Biological Engineering, Department of Mathematics, The Hong Kong University of Science and Technology, 999077, Kowloon, Hong Kong

[4] Computer Science Program, Computer, Electrical and Mathematical Sciences and Engineering (CEMSE) Division, King Abdullah University of Science and Technology (KAUST), Thuwal, 23955-6900, Kingdom of Saudi Arabia

[5] KAUST Computational Bioscience Research Center (CBRC), King Abdullah University of Science and Technology, Thuwal, 23955-6900, Kingdom of Saudi Arabia

[6] Fujian Provincial Key Laboratory of Theoretical and Computational Chemistry, 361005, Fujian, China

* To whom correspondence should be addressed. Email: luzhang@fjirsm.ac.cn

Supplementary Figures

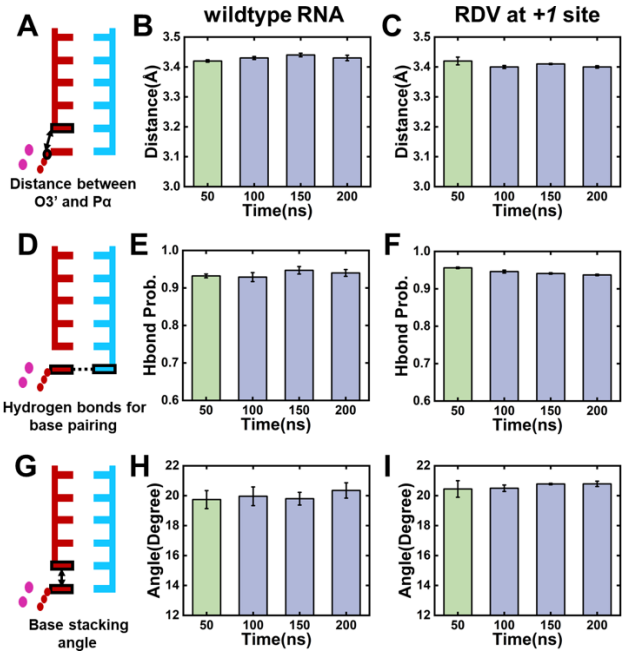

Figure S1. Validation of the convergence of the simulations. (A) A cartoon showing the distance between P_α atom of NTP and the O3' atom of the 3'-terminal nucleotide in the nascent strand. (B) The P_α-O3' distance calculated using 5×50ns (in green), 5×100ns, 5×150ns and 5×200ns (in light purple) simulations for RdRp with wildtype RNA system ("wildtype RNA"). (C) Similar to (B) but for the RdRp with Remdesivir (RDV) embedded at *+1* site of the template strand ("RDV at *+1* site"). (D) A cartoon showing the hydrogen bonds for base pair at the active site. (E) Hydrogen bonding probability for the UTP:A pair at the active site in the "wildtype RNA" system calculated using 5×50ns (in green), 5×100ns, 5×150ns and 5×200ns (in light purple) simulations. (F) Similar to (E) but for the UTP:RDV pair at the active site in the "RDV at *+1* site" system. (G) A cartoon showing the base stacking between UTP and the 3'-terminal nucleotide in the nascent strand. (H) The base stacking angle computed using 5×50ns (in green), 5×100ns, 5×150ns and 5×200ns (in light purple) simulations for the "wildtype RNA" system. (I) Similar to (H) but for the "RDV at *+1* site" system.

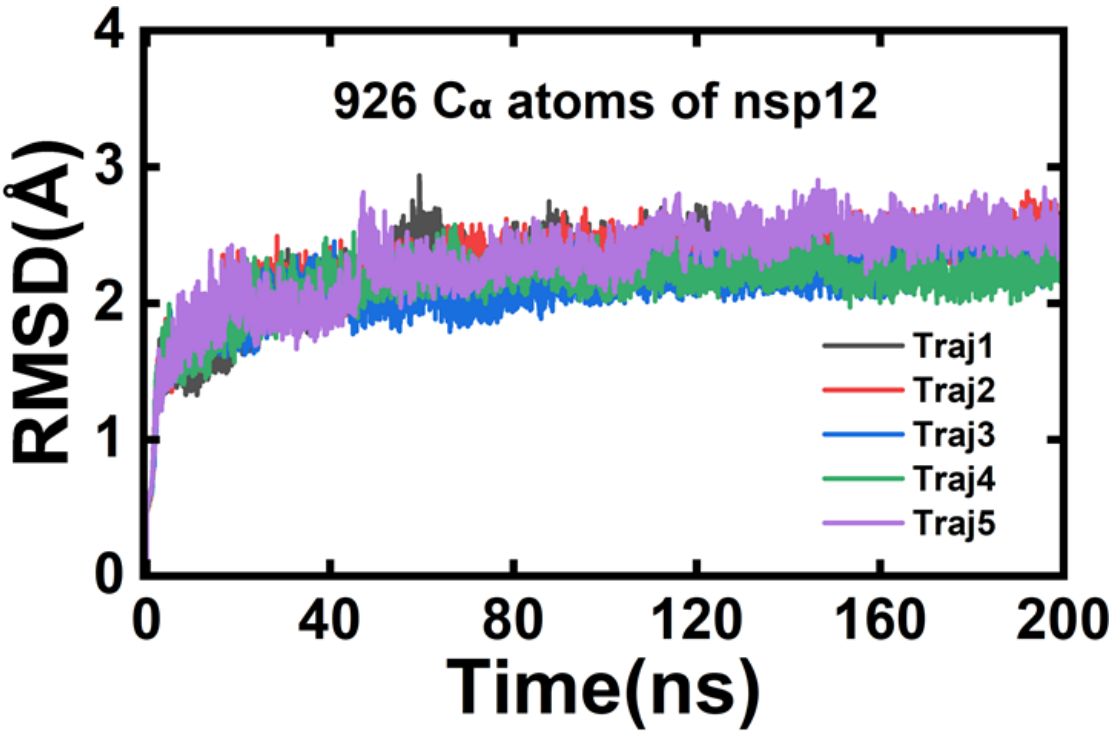

Figure S2. The RMSD of C_α atoms of nsp12 versus time in the five 200ns MD simulations. All the 926 C_α atoms in nsp12 were used for the structural alignment and RMSD calculations.

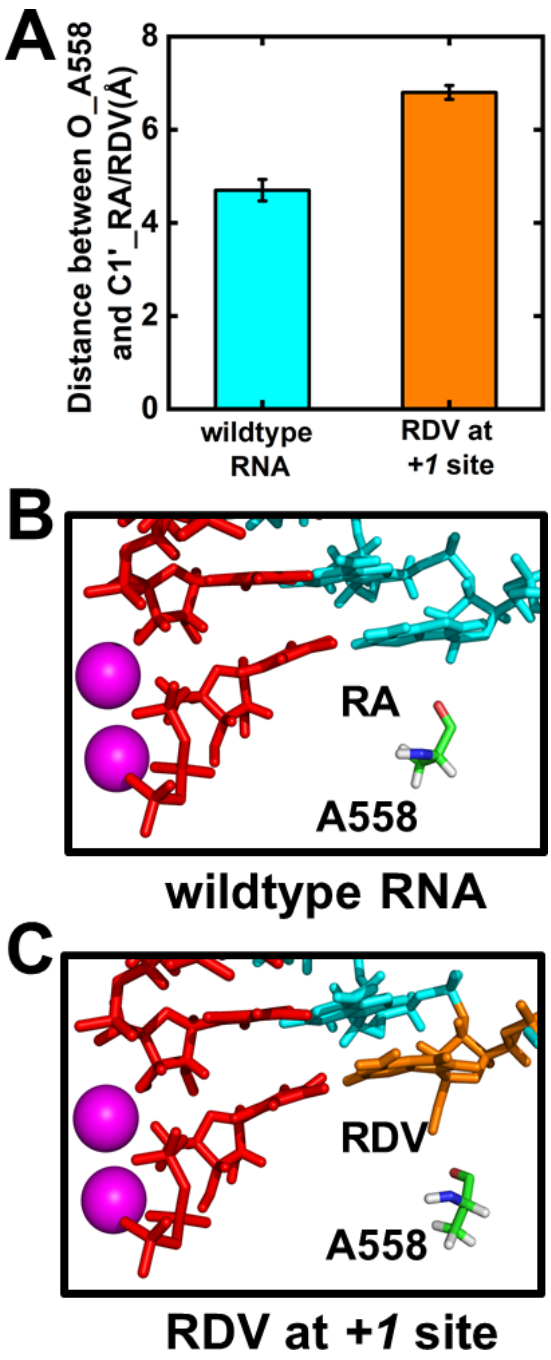

Figure S3. (A) The distance between the oxygen atom in the backbone of A558 and C1' atom of RDV/Adenosine (RA) for RdRp systems with wildtype-RNA (cyan) and RDV at *+1* site (orange). (B) Typical conformation exhibiting A558 with RA in the wildtype RNA system. (C) Typical conformation showing A558 moving away from RDV at *+1* site of the template strand.

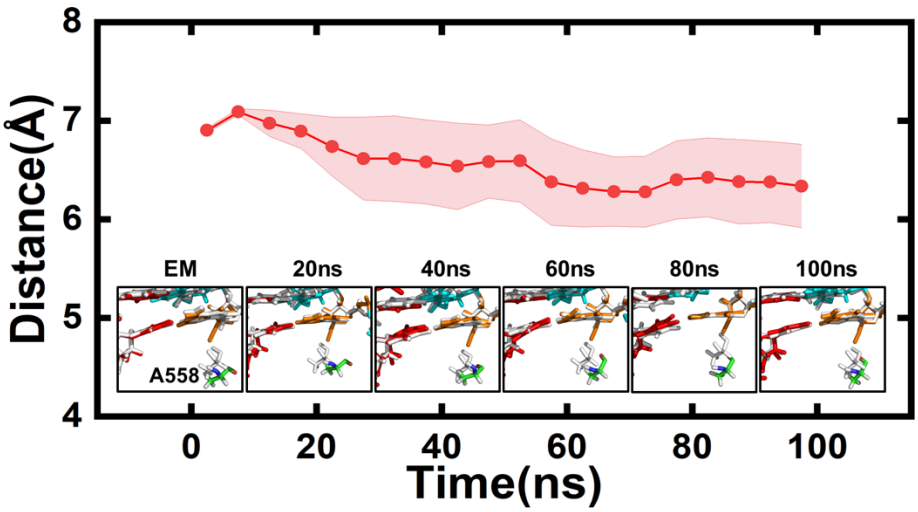

Figure S4. The conformations of A558 along the MD simulations. The red line with circles denotes the distance between A558 and the C1' atom of Remdesivir at *+1* site of template strand along the MD trajectories. Each symbol labels the mean distance averaged by five trajectories using 5ns as a window. The transparent red background represents the standard deviations. The energy minimized conformation as well as the MD snapshots at 20ns, 40ns, 60ns, 80ns and 100ns from one random trajectory of the "RDV at *+1* site" system are shown as inset, where the energy minimized conformation of the "wildtype RNA" system is shown in white for comparison.

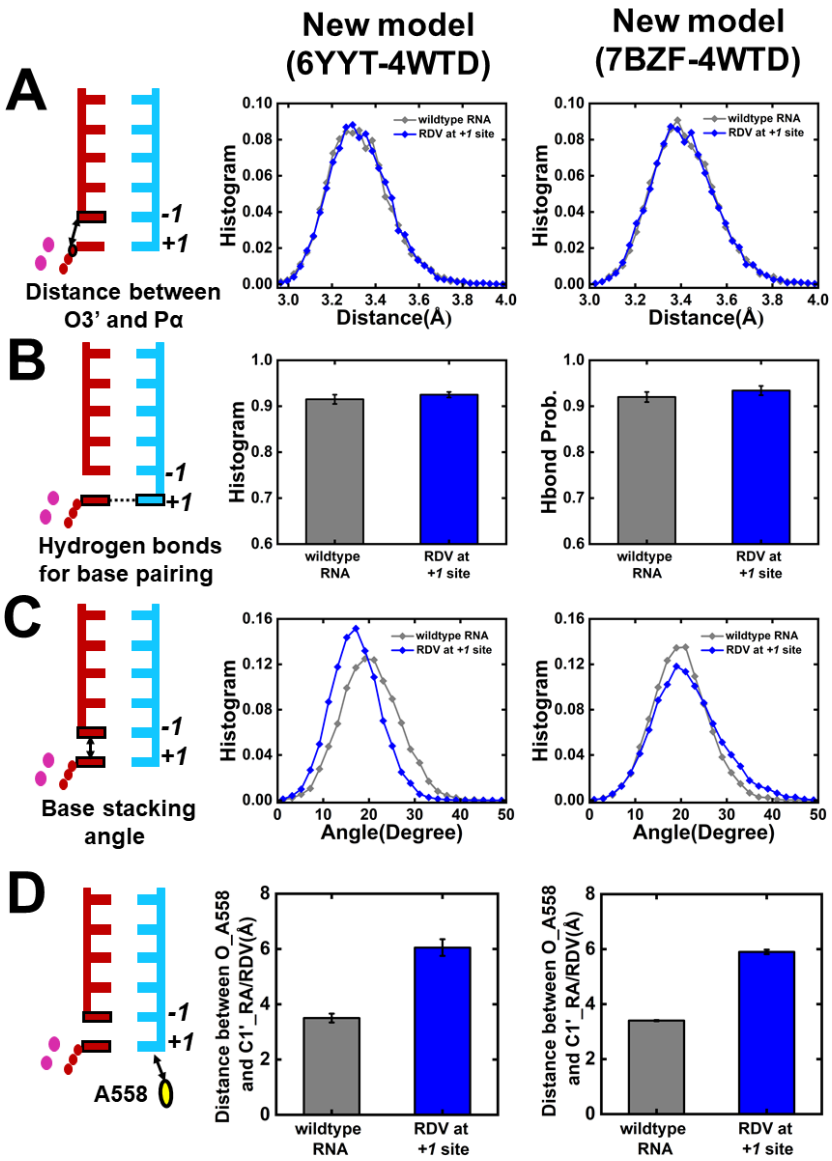

Figure S5. Examination over the stability of active site using MD simulations initiated from different models. (A) The distance between P_α atom of NTP and the O3' atom of the 3'-terminal nucleotide of the nascent strand. (B) The hydrogen bonding probability for the base pair at the active site. (C) The base stacking angle between NTP and the 3'-terminal nucleotide. (D) The distance between A558 and adenosine(RA) /Remdesivir(RDV) at the *+1* site of template strand. In (A)-(D), the first column is the cartoon showing the structural feature under investigation, and the second and third columns demonstrate the results for the "6YYT-4WTD" model and the "7BZF-4WTD" model, respectively.

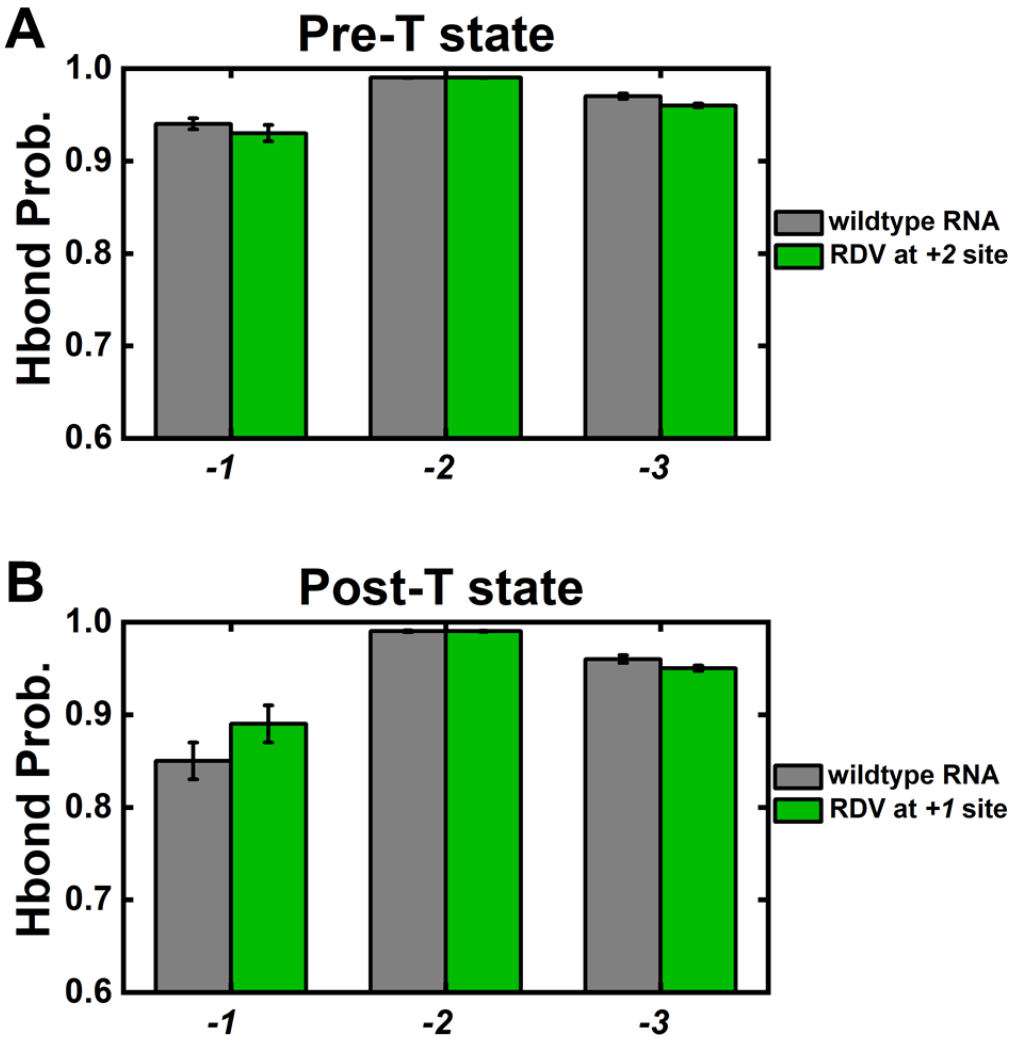

Figure S6. Evaluation of the hydrogen bonding probability between the template and nascent strand at both the pre- (pre-T) and post-translocation (post-T) states. (A) The hydrogen bonding probability when RdRp is in the pre-T state with wildtype RNA (grey) and RDV at *+2* site of the template strand (green). (B) The hydrogen bonding probability when RdRp is in the post-T state with the wildtype RNA (grey) and RDV at *+1* site of the template strand (green).

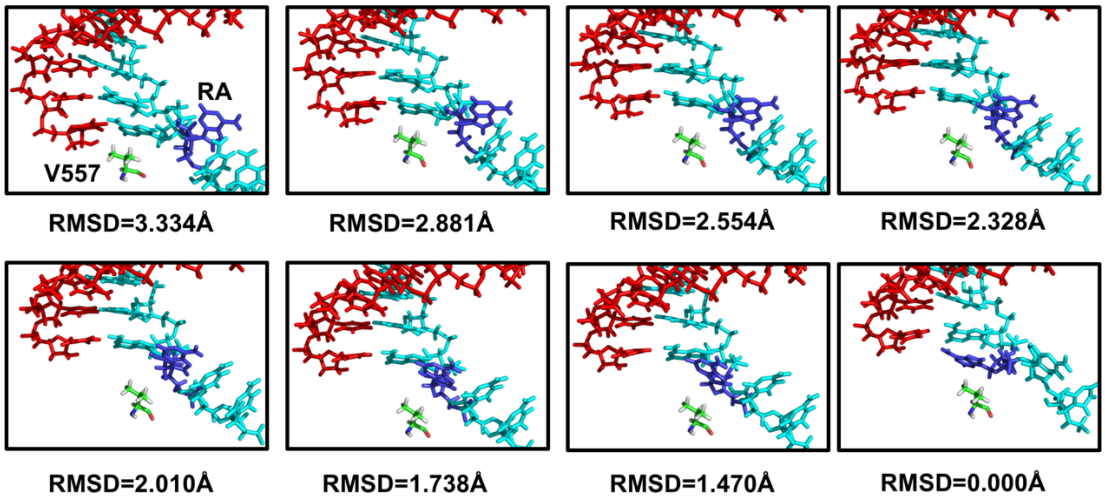

Figure S7. The conformations along the preliminary translocation pathway generated by Climber algorithm. The adenosine (RA) translocating from *+2* to *+1* site is shown in blue. For each conformation, the RMSD relative to the conformation of the post-translocation state is shown below, and all the heavy atoms of nucleotides were included for the RMSD calculations.

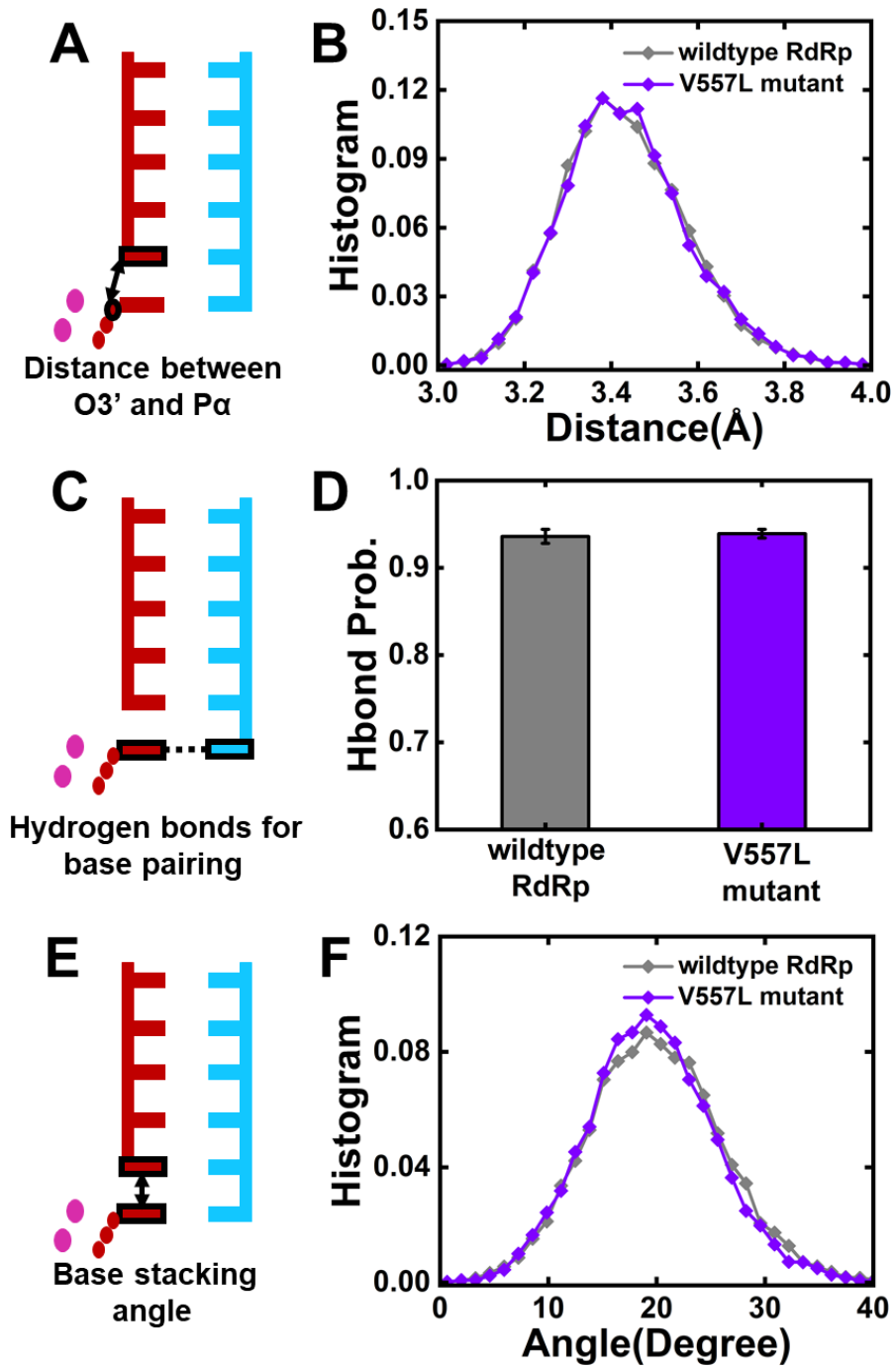

Figure S8. Estimation the capability of maintaining the pre-catalytic conformation for wildtype-RdRp and RdRp with V557L single-point mutation. (A) Cartoon model showing the distance between P_α atom of the NTP and O3' atom of the 3'-terminal nucleotide of product strand. (B) The histogram of P_α-O3' distance. (C) Cartoon model showing the hydrogen bonds for base pairing at the active site. (D) The mean values and standard deviations of hydrogen bonding probability. (E) Cartoon model showing the dihedral angle between the base of NTP and the base of 3'-terminal nucleotide of product strand. (F) The histogram of the base stacking dihedral angles. In (B), (D) and (F), the results of wildtype-RdRp (grey) are in comparison with that of V557L mutant RdRp (violet).

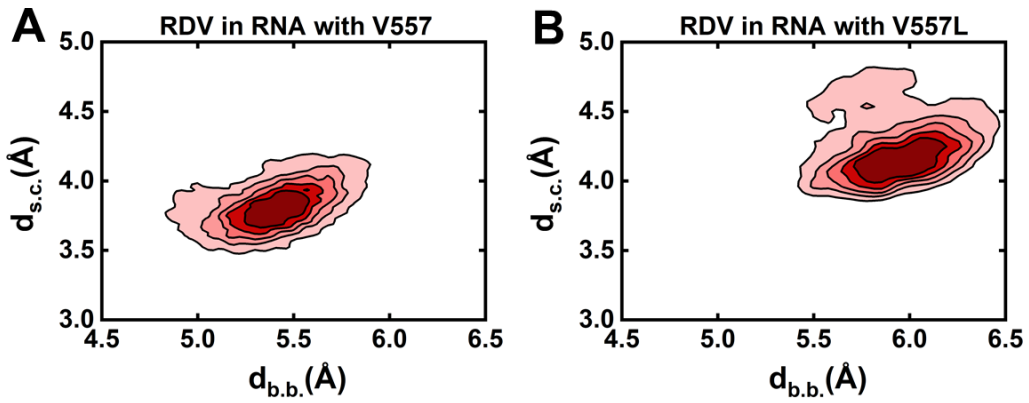

Figure S9. Investigation into the orientation of V557/V557L in the system containing RDV in the template strand. (A) Density plot of $d_{b.b.}$ versus $d_{s.c.}$ for wildtype RdRp with RDV at *+2* site. (B) Similar to (A) but for RdRp upon V557L mutation. The simulations of RdRp with RDV at *+2* site of template strand in the pre-translocation state were adopted for the computation.

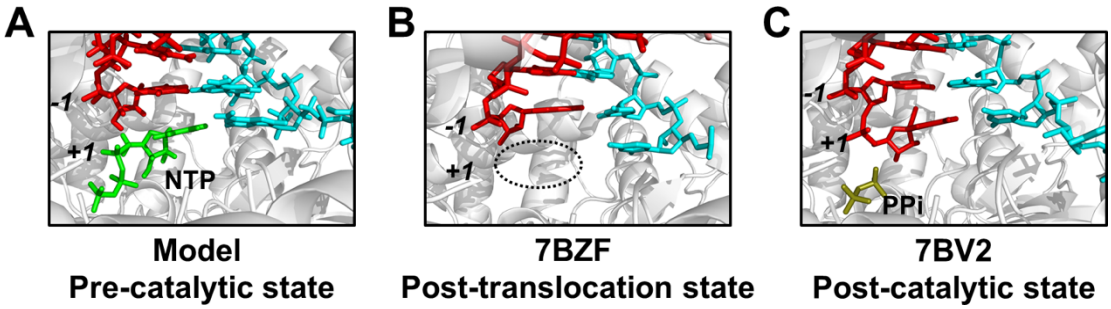

Figure S10. The dsRNA conformation at the active site in our model and two cryo-EM structures of SARS-CoV-2 RdRp. (A) Our model with NTP (in green) bound at the active site (*+1* site) and not covalently bonded with the 3'-terminal of nascent strand at *-1* site (in red). (B) The 7BZF structure with an unoccupied active site and the 3'-terminal nucleotide of the nascent strand is at *-1* site. (C) The 7BV2 structure with the 3'-terminal of nascent strand (in red) occupying the active site (*+1* site) and the pyrophosphate ion (PPi, in dark yellow) is also located at the active site. In (A)-(C), the nascent strand and the template strand are shown in red and cyan, respectively.

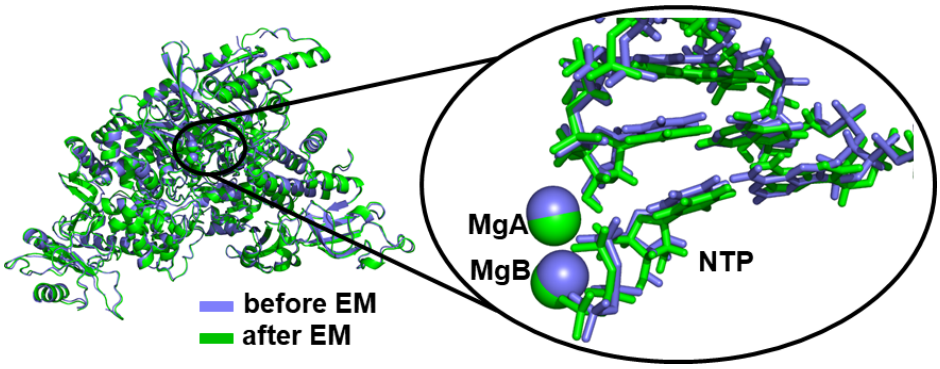

Figure S11. Comparison of the conformation before (in steelblue) and after (in green) energy minimization.

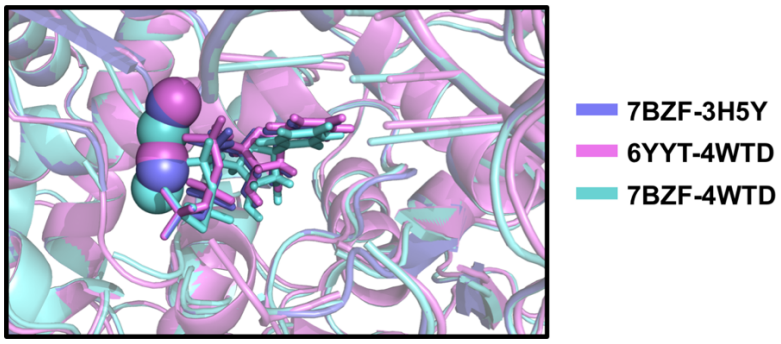

Figure S12. Comparison of the initial positions of NTP (shown in sticks) and Mg^{2+} ions (shown in spheres) in our model with that in the two new models.

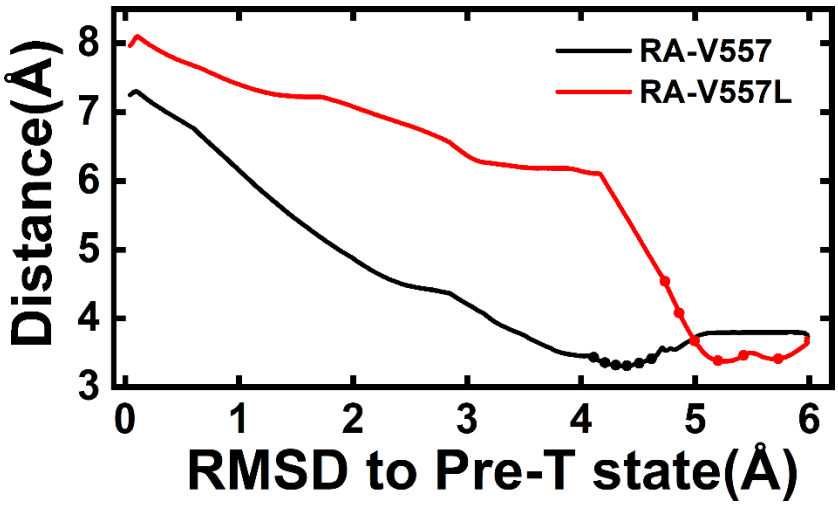

Figure S13. Identifying the translocating intermediates based on the distance between RA and V557/V557L versus the RMSD of dsRNA to the pre-T state during translocation. The distance was measured as the minimum distance between the heavy atoms of the sidechain of V557(black)/V557L(red) and all heavy atoms of RA. The RMSD of the heavy atoms of dsRNA was computed against the pre-T state.

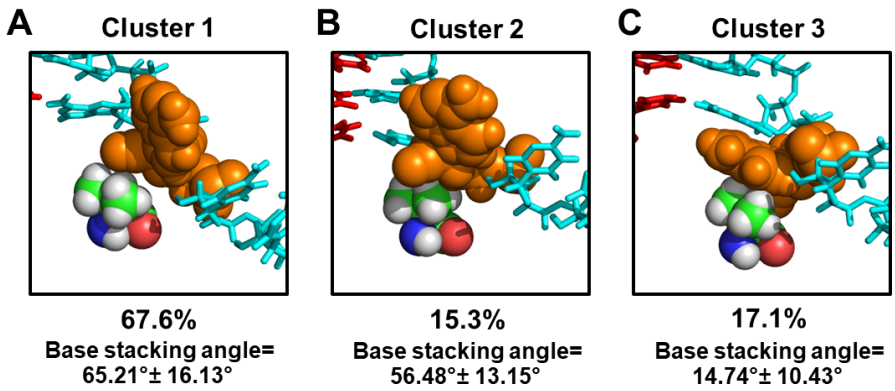

Figure S14. K-center clustering based on the conformational similarity of the MD conformations initiated from the translocation pathway of Remdesivir from *+2* to *+1* site in wildtype RdRp. The representative conformation from each cluster is shown. The representative conformation is defined as the conformation with the base stacking angle close to the mean value of the cluster. Below each representative conformation, the population of the respective cluster is labelled. The mean values and standard deviations of the base stacking angles for the conformations within each cluster were also computed and displayed.

Supplementary methods

1. MD simulations of SARS-CoV-2 RdRp with UTP occupying at the active site

1.1. SARS-CoV-2 RdRp with wildtype double-stranded(ds) RNA

1.1.1. Construction of the structural model

In order to investigate if the NTP incorporation is directly perturbed at the active site when Remdesivir is present at *+1* site of the template strand, we need to construct a model of pre-catalytic state with NTP binding at the closed active site (Fig. S10A). The cryo-EM structure of SARS-CoV-2 nsp12-nsp7-nsp8 RdRp complex in the post-translocation state (PDBID:7BZF[1]) was used as the structural basis for constructing our atomic model. We chose 7BZF cryo-EM structure as the structural basis because the RdRp is in the post-translocation state with an unoccupied active site (*+1* site), i.e. the 3'-terminal of the nascent strand is at *-1* site (Fig. S10B). This structure provides similar dsRNA conformations (shown in red and cyan in Fig. S10) as in the pre-catalytic state (Fig. S10A), which allows us to directly obtain the NTP and Mg^{2+} ions by structural alignment to the norovirus RdRp (PDBID: 3H5Y[2]) in the pre-catalytic state. Missing atoms and residues were modelled using Modeller 9.21[3]. To achieve a structural model of SARS-CoV-2 RdRp with a closed active site, the side chains of surrounding residues D760, D761, D618, Y619 and D623 were modified by aligning to the norovirus RdRp by C_α atoms in proteins[4,5]. Harmonic constraints were added between Zn^{2+} ions and their coordinated protein residues including C301, C306, C310, H295, C487, C645, H642 and C646. Two extra nucleotides in the downstream region of the template strand (*+2* site and *+3* site) were modelled based on the cryo-EM structure of SARS-CoV-2 RdRp in the pre-translocation state (PDBID:7C2K[1]). The Remdesivir incorporated at *-1* site of nascent strand in the cryo-EM structure (PDBID:7BZF[1]) was modified using Coot0.8.7[6] to ensure the nascent strand contains natural nucleotides. The base pair at the active site (*+1* site) was modified to UTP:A using Coot0.8.7[6] to allow for the subsequent substitution of A with Remdesivir to investigate the inhibitory effect. The protonation states of amino acids were predicted using propka3.0 module[7] in the pdbpqr2.2.19 package[8]. The whole complex was solvated with 70,041 TIP3P water molecules[9] in a dodecahedral box with the box edge at least 12 Å away from the surface of the complex. Sufficient counter ions (31 sodium ions) were inserted to neutralize the system.

1.1.2. MD simulations

We used ff14SB amber force field [10-13] and amber chiOL3 force field to simulate the protein and RNA, respectively. The parameters of Remdesivir obtained in the previous study[5] were used and modified to be compatible with amber OL15 force field[10-13]. For UTP or Remdesivir in the triphosphate form, parameters for the triphosphate moiety were taken from those developed by Meagher et al[14].

The whole system was energy minimized by two steps. First, a 10,000-steps energy minimization was performed by position restraining the heavy atoms with a force constant of 10 $kJ\times mol^{-1}\times Å^{-2}$. Afterward, another 10,000-steps energy

minimization was performed without restrains. No obvious conformational change has been observed with a RMSD as low as ~ 0.6Å (Fig. S11) when all the C_α atoms, the heavy atoms of nucleotides, NTP and the two Mg^{2+} ions were included for the RMSD calculation. After energy minimization, one 200 ps position restraint simulation was performed under NVT ensemble (T=300K) by restraining the heavy atom with a force constant of 10 kJ×mol^{-1}×Å^{-2}, followed by another 500 ps position restraint simulation under NPT ensemble (T=300 K, P=1 bar). The last conformation of the position restraint simulation was used to conduct one 50 ns simulations under NPT ensemble (T=300 K, P=1 bar). The last conformation of the 50 ns equilibrated simulation was adopted to seed five 50 ns MD simulations under NVT ensemble (T=300 K) with different initial velocities. Temperature annealing was performed from 50 K to 300 K in the first 2 ns. The V-rescale thermostat[15] was applied with a coupling constant of 0.1 ps. The Lennard-Jones interactions were smoothly switched off from 10 Å to 12 Å. The short-range electrostatic interactions were cut off at 12 Å and the long-range electrostatic interactions were treated with the Particle-Mesh Ewald (PME) summation method[16]. The neighbor list was updated every 10 steps. The MD snapshot was saved with an interval of 20 ps. The first 10 ns was removed before performing the analysis.

To evaluate the convergence of sampling, we have further extended the simulations to 200 ns to achieve the conformational ensemble of 5×200ns for examining the stability of NTP and its capability to maintain the catalytically active conformation at the active site (Fig. S1).

All the simulations in the current work were performed by Gromacs 5.0[17].

1.2. Other SARS-CoV-2 RdRp models with UTP occupying at the active site

1.2.1. Model construction

(1) RdRp with Remdesivir embedded at template strand

The last conformation of the 50 ns equilibrated simulation of RdRp with wildtype dsRNA (Section 1.1.2) was used as the structural basis, and we manually modified the adenosine at *+1* site of template strand to RDV, in order to achieve the structural model of RdRp with RDV embedded at *+1* site of the template strand.

(2) V557L mutant RdRp with wildtype dsRNA

Based on the last conformation of the 50 ns equilibrated simulation of wildtype RdRp with wildtype dsRNA (Section 1.1.2), we made the Val-to-Leu mutation by pymol[18] to achieve the model of V557L mutant RdRp with wildtype dsRNA.

1.2.2. MD simulations

For each of the above models, we performed the MD simulations as below:

(a) 10,000-steps energy minimization by position restraining the heavy atoms with a force constant of 10 kJ×mol^{-1}×Å^{-2};
(b) 10,000-steps energy minimization without restrains;
(c) 200 ps position restraint simulation under NVT ensemble (T=300 K) by restraining the heavy atom with a force constant of 10 kJ×mol^{-1}×Å^{-2};
(d) 500 ps position restraint simulation under NPT ensemble (T=300 K, P=1 bar);

(e) the last conformation of the position restraint NPT simulation was used to randomly seed five 50 ns MD simulations under NVT (T=300 K) ensemble with temperature annealing from 50 K to 300 K in the first 2 ns.

In the simulations, the same parameters as used in Section 1.1.2 were utilized. For the model constructed in Section 1.2.1(1), we have extended the simulations to 200 ns for convergence validation (Fig. S1).

2. MD simulations of SARS-CoV-2 RdRp in the pre-translocation state

2.1. SARS-CoV-2 RdRp with wildtype dsRNA

2.1.1. Model construction

The cryo-EM structure of SARS-CoV-2 RdRp (PDBID:7C2K[1]) was used to construct the model of RdRp in the pre-translocation (pre-T) state. The missing side chains of protein residues were modelled by aligning the structure to the cryo-EM structure of SARS-CoV-2 RdRp (PDBID:6YYT[19]) based on C_α using Modeller 9.21[3]. Based pair at *+1* site was mutated to U:A and the nucleotide at *+2* site of the template strand was mutated to A using Coot0.8.7[6], to be compatible with the sequence of the post-translocation state. The whole complex was solvated with TIP3P water molecules[9] in a dodecahedral box (148.1Å×148.1Å×148.1Å) and neutralized by counter ions.

2.1.2. MD simulations

We performed multiple steps of energy minimization and position restraint simulations to gradually relax and fully equilibrate the simulation complex as follows:

(a) 10,000-steps energy minimization by position restraining the heavy atoms with a force constant of 10 kJ×mol^{-1}×Å^{-2};
(b) 10,000-steps energy minimization without restrains;
(c) 200 ps position restraint simulation under NVT ensemble (T=300K) by restraining the heavy atom with a force constant of 10 kJ×mol^{-1}×Å^{-2};
(d) 500 ps position restraint simulation under NPT ensemble (T=300 K, P=1 bar);
(e) the last conformation of position restraint simulation was used to conduct one 50 ns simulations under NPT ensemble (T=300 K, P=1 bar);
(f) the last conformation of the 50ns equilibrated simulation was adopted to seed five 50 ns MD simulations under NVT (T=300 K) ensemble with different initial velocities and the temperature was gradually increased from 50 K to 300 K in the first 2 ns.

In the simulations, the same parameters as used in Section 1.1.2 were utilized.

2.2. Other SARS-CoV-2 RdRp models in the pre-translocation state

2.2.1. Model construction

(1) Wildtype RdRp with Remdesivir embedded at template strand

The last conformation of the 50 ns equilibrated simulation of RdRp with wildtype dsRNA (Section 2.1.2) was used as the structural basis. RDV was modelled to the *+2* site in the template strand to achieve the structural model of RdRp in the pre-T state with RDV embedded at *+2* site of the template strand.

(2) V557L mutant RdRp with wildtype dsRNA

We used the last conformation of the 50 ns equilibrated simulation of RdRp in the pre-T state (Section 2.1.2) and made the Val-to-Leu mutation by pymol[18] to achieve the model of V557L mutant RdRp in the pre-T state with wildtype dsRNA.

(3) V557L mutant RdRp with Remdesivir embedded at template strand

Base on the structure of RdRp in the pre-T state with V557L mutant constructed in Section 2.2.1(2), we modified the adenosine at *+2* site of template strand to achieve the structural model of V557L mutant RdRp with Remdesivir at *+2* site.

2.2.2. MD simulations

For each of the above models, we performed the multiple steps as described in Section 1.2.2 for the MD simulations.

3. Generating the preliminary pathways connecting the pre- and post-translocation states

3.1. Constructing the structural models of pre- and post-translocation states

The structural model of pre-translocation state constructed in Section 2.1.1 was used as the structural basis and one base pair located at the upstream terminal of the template-nascent duplex was removed to achieve a pre-T model containing nucleotides from *+4* to *-9* site. To investigate the translocation by one base pair, we would require a post-translocation (post-T) model with nucleotides from *+3* to *-10* site. To achieve such a RdRp model, the cryo-EM structure (PDBID:7BZF[1]) was used as the structural basis. Missing atoms and residues were modelled using Modeller 9.21[3]. As the cryo-EM structure contains the nucleotides from *+1* to *-9* site, we added one extra pair in the upstream region of the template-nascent duplex and two extra nucleotides in the downstream region of the template strand by aligning to the cryo-EM structure of SARS-CoV-2 RdRp (PDBID:7C2K[1]). The sequence of the nucleotide in the post-T state was also mutated by Coot0.8.7[6] to ensure the sequence is shifted by one base-pair position compared to the pre-T state.

To generate preliminary pathway for RdRp V557L mutation, we collected the MD conformations based on the model constructed in Section 2.2.1(2) and performed K-center clustering[20] based on the root mean squared deviation of the heavy atoms of V557L. We generated 5 states in total with the most populated state containing 80.2% MD conformations. The central conformation of this dominant state of V557L mutant RdRp was then chosen as the basis for constructing the pre-T and post-T states with V557L mutation. In particular, for the V557L mutant RdRp in the pre-T state, the base pair at the most upstream region of the RNA duplex was removed. Such a pre-T model also lays the foundation for constructing the V557L mutant RdRp in the post-T state, where the 3'-terminal nucleotide at the product strand and the downstream terminal nucleotide of template strand were removed from the pre-T model. The sequence of the nucleotide in the post-T state was also mutated by Coot0.8.7[6] to ensure the sequence is consistent with the pre-T state.

3.2. Generating the preliminary pathways

We first performed energy minimization for both the pre- and post-T models using Energy Calculation and Dynamics (ENCAD)[21] algorithm. Afterwards, we applied Climber algorithm[22,23] to generate the preliminary pathways to connect the energy minimized pre- and post-T conformations of the SARS-CoV-2 RdRp with dsRNA (Fig. S7). A 500-steps Climber simulation was performed to gradually drive the dsRNA from the pre-T state to the post-T state with an external energy applied on dsRNA while the protein is frozen. These procedures were used to generate each of the following pathways:

(a) from *+2* site to *+1* site for the SARS-CoV-2 RdRp.

(b) from *+2* site to *+1* site for the SARS-CoV-2 RdRp with V557L mutation.

For each pathway, the translocating conformations showing short distances between V557/V557L and translocating RA in the template strand were chosen as the translocating intermediates (Fig. S13). To investigate the translocation of Remdesivir, the translocating adenosine was substituted by Remdesivir in each of the intermediate conformations.

3.3. Simulation of the translocating intermediates

For each of the translocating intermediates, the RdRp complex was solvated with TIP3P water molecules[9] in a dodecahedral box with box edge at least 12 Å away from the complex surface and sufficient counter ions were inserted to ensure the whole system neutral. Multiple steps of energy minimization and position restraint simulations were performed for full equilibration as follows:

(a) 10,000-steps energy minimization by position restraining the heavy atoms with a force constant of 10 $kJ\times mol^{-1}\times Å^{-2}$;

(b) 10,000-steps energy minimization was performed without restrains;

(c) 200 ps position restraint simulation was performed under NVT ensemble (T=300 K) by restraining the heavy atom with a force constant of 10 $kJ\times mol^{-1}\times Å^{-2}$;

(d) 500 ps position restraint simulation under NPT ensemble (T=300 K, P=1 bar);

(e) the last conformation of the NPT equilibration was used to conduct one 10 ns simulation under NPT ensemble (T=300 K, P=1 bar) with position restrains were applied on two nucleotides at the downstream terminal of the template strand. The V-rescale thermostat[15] was applied with a coupling constant of 0.1 ps. The Parrinello-Rahman[24,25] barostat was applied with a coupling constant of 2 ps. Other simulation parameters were the same as those described in section 1.1.2.

3.4. K-center clustering on the MD conformations

To further validate if the conformation in Fig. 3G is a representative rather than a high-energy and low populated conformation, we have collected the MD conformations with Remdesivir from *+2* to *+1* site in wildtype RdRp (as shown in Section 3.3) and performed K-center clustering based on the conformational similarity. In specific, the heavy atoms of V557, Remdesivir and its upstream nucleotide were included in the

measurement of conformational similarity. The results of clustering have demonstrated one cluster with dominant population (67.6%) while each of the remaining two clusters possesses ~15% of MD conformations (Fig. S14). Interestingly, the highest populated cluster shows the largest base stacking angle (Fig. S14A), i.e., the base of Remdesivir is significantly tilted relative to the base of the upstream nucleotide with the mean angle of 65.2°±16.1°. The other two clusters are showing the mean angles of 56.5°±13.1° and 14.7°±10.4°, respectively (Fig. S14B and S14C). This analysis has suggested that the conformation as shown in Fig. 3G with an obviously tilted base relative to its upstream nucleotide is a highly populated and representative conformation rather than a high-energy conformation along the translocation pathway for wildtype RdRp with V557.

4. MD simulations of other models with UTP occupying at the active site

We have noticed that there exist other cryo-EM structures of SARS-CoV-2 RdRp[4,19,26-29], which could also serve as the basis for the construction of structural model. For example, the 7BV2 structure demonstrates RdRp in the post-catalytic state with the byproduct pyrophosphate ion bound at the active site (Fig. S10C). In this structure, the 3'-terminal of the nascent strand occupies the active site (*+1* site). However, the target model of the pre-catalytic state requires that the 3'-terminal of nascent strand locates at *-1* site and the NTP is at the *+1* site while not covalently bound with the nascent strand (Fig. S10A). Therefore, the dsRNA in 7BV2 structure adopts the conformation different from that in the pre-catalytic state (shown in red and cyan in Fig. S10A and S10C). Although 7BV2 structure could also serve as an alternative basis for model construction upon multiple-step modifications, the different dsRNA conformational states in 7BV2 structure (Fig. S10C) and the target model (Fig. S10A) render it more challenging to use 7BV2 as the basis to construct our initial model for MD simulations. We have noted that there is an alternative cryo-EM structure of SARS-CoV-2 RdRp in the post-translocation state (PDBID: 6YYT[19]), the same state as that (PDBID: 7BZF[1]) used in our modelling. Moreover, this structure has been used in a previous work[30,31] for modelling the pre-catalytic state. In this section, we have followed the previous work to construct two extra models and performed MD simulations to validate if our conclusions are robust when different experimental structures were used for the initial modelling.

4.1. Constructing the structural models based on other experimental structures

We have constructed two new models based on different experimental structures and re-examined the stability of the active site by MD simulations. In our current model, the initial positions of NTP and Mg^{2+} ions in the active site were obtained by aligning the 7BZF[1] structure of SARS-CoV-2 RdRp to the norovirus RdRp in the pre-catalytic state (PDBID: 3H5Y[2]) based on the C_α atoms using Pymol programme[18], as the active site configuration is conserved across the viral RdRps. This model is named as "7BZF-3H5Y". To achieve a new model, we have followed the previous work[31] to construct the model (named as "6YYT-4WTD") based on another cryo-EM structure of SARS-CoV-2 RdRp in the post-translocation state (PDBID: 6YYT[19]), and used the HCV RdRp (PDBID: 4WTD[32]) in the pre-catalytic state as the reference for structural alignment to

obtain the initial positions of NTP and Mg^{2+} ions. Besides, we have also built an alternative model (named as "7BZF-4WTD") by using the same SARS-CoV-2 RdRp (PDBID: 7BZF[1]) as used in our study but obtaining the NTP and Mg^{2+} ions from HCV RdRp (PDBID: 4WTD[32]). In the two new models, the positions of NTP and Mg ions are a bit different from that in our current model (Fig. S12), which can thus well serve the purpose to examine if our approaches and results are sensitive to the initial model.

4.1.1. "7BZF-4WTD" model

We have followed a similar procedure as performed in Section 1.1.1 to construct this model, excepting that the positions of NTP and Mg^{2+} were obtained by aligning to the HCV RdRp (PDBID: 4WTD[32]) based on all the C_α atoms using Pymol[18]. In specific, the adenosine diphosphate (ADP) in HCV RdRp was extracted and placed at the active site of the cryo-EM structure of SARS-CoV-2 RdRp (PDBID: 7BZF[1]). The "PG" phosphate atom as well as the three oxygen atoms connected to "PG" were taken from our "7BZF-3H5Y" model and chemical bond is manually added between "PG" and the bridge oxygen atom in ADP to create an ATP molecule at the active site. Mutation was further made by Coot0.8.7[6] to modify ATP to UTP, in order to generate UTP:A pair at the active site. The two Mn^{2+} ions were taken from HCV RdRp and changed to Mg^{2+} ions. The protein residues (D618, D760, D761, S282, G283, R555 and K545) that define the closed state of active site in SARS-CoV-2 RdRp were taken from the counterpart residues (D220, D318, D319, S682, G683, R158 and K141) in HCV RdRp. Other modifications were performed as that shown in Section 1.1.1.

4.1.2. "6YYT-4WTD" model

In this model, another cryo-EM structure of SARS-CoV-2 RdRp in the post-translocation state (PDBID: 6YYT[19]) was used as the structural basis and the positions of NTP and Mg^{2+} were obtained by aligning to the HCV RdRp (PDBID: 4WTD[32]). In specific, we have aligned the 6YYT structure with our model based on all the C_α atoms using Pymol[18], and the missing residues in 6YYT structure (residues with IDs 1-76 and 98-117 in nsp12) were modelled from the corresponding residues in our model. Residues with IDs 6-79 in nsp8-1 and IDs 6-81 in nsp8-2, the four unpaired nucleotides in the upstream region of product strand, as well as the nucleotides with chain IDs U and Q flanked by nsp8 extensions were deleted to ensure the new model possesses the same lengths of nsp8 and RNA strands as our model. To obtain the positions of NTP and Mg^{2+} ions, we have followed the previous work[31] and aligned cryo-EM structure of SARS-CoV-2 RdRp (PDBID: 6YYT[19]) with HCV RdRp (PDBID: 4WTD[32]) based on several protein residues (residues S501, G590, G683 and S814 in SARS-CoV-2 RdRp, A97, F193, G283 and S367 in HCV RdRp). After the structural alignment, the UTP:A base pair was modelled in a similar way as that illustrated in Section 4.1.1. The conformations of the protein residues that define the closed state of active site were also taken from the corresponding residues in HCV RdRp as that shown in Section 4.1.1.

4.2. MD simulations

For each of the above models, we performed the MD simulations as below:

(a) 10,000-steps energy minimization by position restraining the heavy atoms

with a force constant of 10 kJ×mol^{-1}×Å^{-2};

(b) 10,000-steps energy minimization without restrains;

(c) 200 ps position restraint simulation under NVT ensemble (T=300 K) by restraining the heavy atom with a force constant of 10 kJ×mol^{-1}×Å^{-2};

(d) 500 ps position restraint simulation under NPT ensemble (T=300 K, P=1 bar);

(e) the last conformation of the position restraint NPT simulation was used to randomly seed five 50 ns MD simulations under NVT (T=300 K) ensemble with temperature annealing from 50 K to 300 K in the first 2 ns. In the simulations, the same parameters as used in Section 1.1.2 were utilized.

Reference

(1) Wang, Q.; Wu, J.; Wang, H.; Gao, Y.; Liu, Q.; Mu, A.; Ji, W.; Yan, L.; Zhu, Y.; Zhu, C.; et al. Structural Basis for RNA Replication by the SARS-CoV-2 Polymerase. *Cell* **2020,** *182*, 417-428.

(2) Zamyatkin, D. F.; Parra, F.; Machin, A.; Grochulski, P.;Ng, K. K. S. Binding of 2'-Amino-2'-Deoxycytidine-5'-Triphosphate to Norovirus Polymerase Induces Rearrangement of the Active Site. *J. Mol. Biol.* **2009,** *390*, 10-16.

(3) Webb, B.;Sali, A. Comparative Protein Structure Modeling Using Modeller. *Curr. Protoc. Bioinf.* **2016,** *54*, 561-5637.

(4) Kirchdoerfer, R. N.;Ward, A. B. Structure of the SARS-CoV Nsp12 Polymerase Bound to Nsp7 and Nsp8 Co-factors. *Nat. Commun.* **2019,** *10*, 2342.

(5) Zhang, L.; Zhang, D.; Wang, X.; Yuan, C.; Li, Y.; Jia, X.; Gao, X.; Yen, H. L.; Cheung, P. P.-H.;Huang, X. 1'-Ribose Cyano Substitution Allows Remdesivir to Effectively Inhibit Nucleotide Addition and Proofreading During SARS-CoV-2 Viral RNA Replication. *Phys. Chem. Chem. Phys.* **2021,** *23*, 5852-5863.

(6) Emsley, P.; Lohkamp, B.; Scott, W. G.;Cowtan, K. Features and Development of Coot. *Acta Crystallogr. D.* **2010,** *66*, 486-501.

(7) Olsson, M. H. M.; Sondergaard, C. R.; Rostkowski, M.; Jensen, J. H. Propka3: Consistent Treatment of Internal and Surface Residues in Empirical pK(a) Predictions. *J. Chem. Theory Comput.* **2011,** *7*, 525-537.

(8) Dolinsky, T. J.; Nielsen, J. E.; McCammon, J. A.; Baker, N. A. PDB2PQR: An Automated Pipeline for the Setup of Poisson-Boltzmann Electrostatics Calculations. *Nucleic Acids Res.* **2004,** *32*, W665-667.

(9) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983,** *79*, 926-935.

(10) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015,** *11*, 3696-3713.

(11) Krepl, M.; Zgarbova, M.; Stadlbauer, P.; Otyepka, M.; Banas, P.; Koca, J.; Cheatham, T. E.; Jurecka, P.;Sponer, J. Reference Simulations of Noncanonical Nucleic Acids with Different Chi Variants of the AMBER Force Field: Quadruplex DNA, Quadruplex RNA and Z-DNA. *J. Chem. Theory Comput.* **2012,** *8*, 2506-2520.

(12) Zgarbova, M.; Luque, F. J.; Sponer, J.; Cheatham, T. E.; Otyepka, M.; Jurecka, P. Toward Improved Description of DNA Backbone: Revisiting Epsilon and Zeta Torsion Force Field Parameters. *J. Chem. Theory Comput.* **2013,** *9*, 2339-2354.

(13) Zgarbova, M.; Otyepka, M.; Sponer, J.; Mladek, A.; Banas, P.; Cheatham, T.; Jurecka, P. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *J. Chem. Theory Comput.* **2011,** *7*, 2886-2902.

(14) Meagher, K.; Redman, L.; Carlson, H. Development of Polyphosphate Parameters for Use with the AMBER Force Field. *J. Comput. Chem.* **2003,** *24*, 1016-1025.

(15) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling Through Velocity Rescaling. *J. Chem. Phys.* **2007,** *126*, 014101.

(16) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995,** *103*, 8577-8593.

(17) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations Through Multi-level Parallelism from Laptops to Supercomputers. *SoftwareX.* **2015,** *1*, 19-25.

(18) Yuan, S.; Chan, H. C. S.; Hu, Z. Using PyMOL as A Platform for Computational Drug Design. *WIREs Comput. Mol. Sci.* **2017,** *7*, e1298.

(19) Hillen, H. S.; Kokic, G.; Farnung, L.; Dienemann, C.; Tegunov, D.; Cramer, P. Structure of Replicating SARS-CoV-2 Polymerase. *Nature* **2020,** *584*, 154-156.

(20) Bowman, G. R.; Beauchamp, K. A.; Boxer, G.; Pande, V. S. Progress and Challenges in the Automated Construction of Markov State Models for Full Protein Systems. *J. Chem. Phys.* **2009,** *131*, 124101.

(21) Levitt, M. Molecular-Dynamics of Native Protein.1.Computer-Simulation of Trajectories. *J. Mol. Biol.* **1983,** *168*, 595-617.

(22) Weiss, D. R.; Levitt, M. Can Morphing Methods Predict Intermediate Structures? *J. Mol. Biol.* **2009,** *385*, 665-674.

(23) Silva, D. A.; Weiss, D. R.; Pardo, A. F.; Da, L. T.; Levitt, M.; Wang, D.; Huang, X. Millisecond Dynamics of RNA Polymerase II Translocation at Atomic Resolution. *Proc. Natl. Acad. Sci. U. S. A.* **2014,** *111*, 7665-7670.

(24) Nosé, S.; Klein, M. L. Constant Pressure Molecular Dynamics for Molecular Systems. *Mol. Phys.* **2006,** *50*, 1055-1076.

(25) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **2010,** *132*, 154104.

(26) Yin, W.; Mao, C.; Luan, X.; Shen, D.-D.; Shen, Q.; Su, H.; Wang, X.; Zhou, F.; Zhao, W.; Gao, M.; et al. Structural Basis for Inhibition of the RNA-Dependent RNA Polymerase from SARS-CoV-2 by Remdesivir. *Science*. **2020**, *368*, 1499-1504.

(27) Peng, Q.; Peng, R.; Yuan, B.; Wang, M.; Zhao, J.; Fu, L.; Qi, J.; Shi, Y. Structural Basis of SARS-CoV-2 Polymerase Inhibition by Favipiravir. *Innovation*. **2021**, *2*, 100080.

(28) Kokic, G.; Hillen, H. S.; Tegunov, D.; Dienemann, C.; Seitz, F.; Schmitzova, J.; Farnung, L.; Siewert, A.; Hobartner, C.; Cramer, P. Mechanism of SARS-CoV-2 Polymerase Stalling by Remdesivir. *Nat. Commun*. **2021**, *12*, 279.

(29) Naydenova, K.; Muir, K. W.; Wu, L. F.; Zhang, Z.; Coscia, F.; Peet, M. J.; Castro-Hartmann, P.; Qian, P.; Sader, K.; Dent, K.; et al. Structure of the SARS-CoV-2 RNA-Dependent RNA Polymerase in the Presence of Favipiravir-RTP. *Proc. Natl. Acad. Sci. U. S. A.* **2021**, *118*, e2021946118.

(30) Gordon, C. J.; Tchesnokov, E. P.; Woolner, E.; Perry, J. K.; Feng, J. Y.; Porter, D. P.; Gotte, M. Remdesivir is A Direct-acting Antiviral that Inhibits RNA-Dependent RNA Polymerase from Severe Acute Respiratory Syndrome Coronavirus 2 with High Potency. *J. Biol. Chem.* **2020**, *295*, 6785-6797.

(31) Tchesnokov, E. P.; Gordon, C. J.; Woolner, E.; Kocinkova, D.; Perry, J. K.; Feng, J. Y.; Porter, D. P.; Gotte, M. Template-dependent Inhibition of Coronavirus RNA-

Dependent RNA Polymerase by Remdesivir Reveals a Second Mechanism of Action. *J. Biol. Chem.* **2020**, *295*, 16156-16165.

(32) Appleby, T.; Perry, J.; Murakami, E.; Barauskas, O.; Feng, J.; Cho, A.; Iii, D.; Wetmore, D.; McGrath, M.; Ray, A.; et al. Structural Basis for RNA Replication by the Hepatitis C Virus Polymerase. *Science.* **2015**, *347*, 771-775.