Supplementary Information

1’-cyano substitution of Remdesivir “template-dependent”

inhibiting the transcription of SARS-CoV-2

Xueying Luo1,2, Xiaowei Wang3, Xin Gao4,5, Lu Zhang1,2,6*

1 State Key Laboratory of Structural Chemistry, Fujian Institute of Research on the Structure of Matter, Chinese Academy of Sciences, Fuzhou, Fujian, China
2 University of Chinese Academy of Sciences, Beijing, China
3 Department of Chemical and Biological Engineering, Centre of Systems Biology and Human Health, State Key Laboratory of Molecular Neuroscience, The Hong Kong University of Science and Technology, 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
6 Fujian Provincial Key Laboratory of Theoretical and Computational Chemistry, Fujian, China

* To whom correspondence should be addressed. Email: luzhang@fjirsm.ac.cn
Figure S1. (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 S2. 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 S3. 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 Pα atom of the NTP and O3’ atom of the 3’-terminal nucleotides. (B) The histogram of Pα-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. (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 S4. V557L shows similar configuration in both the RdRp with wildtype RNA and the RdRp with RDV at the template strand. Typical conformations of V557L with RA (A) and RDV (B) in SARS-CoV-2 RdRp.
Figure S5. Selection of the 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.
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

The cryo-EM structure of SARS-CoV-2 nsp12-nsp7-nsp8 RdRp complex in the post-translocation state (PDBID:7BZF) was used as the structural basis for constructing our atomic model. Missing atoms and residues were modelled using Modeller 9.21. The NTP and Mg$^{2+}$ ions in the active site (+1 site) were modelled by aligning to the RdRp of Norovirus (PDBID:3H5Y) based on Cα atoms using Pymol. 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 Cα atoms in proteins. Harmo nic 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). The base pairs at both +1 and -1 site were mutated to UTP:A and U:A using Coot0.8.7. The protonation states of amino acids were predicted using propka3.0 module in the pdbpqr2.2.19 package. The whole complex was solvated with 70,041 TIP3P water molecules 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 and amber chiOL3 force field to simulate the protein and RNA, respectively. The parameters of Remdesivir obtained in the previous study were used and modified to be compatible with amber OL15 force field. For UTP or Remdesivir in the triphosphate form, parameters for the triphosphate tail were taken from those developed by Meagher et al. 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×mol$^{-1}$×Å$^{-2}$. Afterward, another 10,000-steps energy minimization was performed without restraints. 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 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\textsuperscript{17}. 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.

All the simulations in the current work were performed by Gromacs 5.0\textsuperscript{18}.

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 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\textsuperscript{4} 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\textsuperscript{1}×Å\textsuperscript{2};
(b) 10,000-steps energy minimization without restraints;
(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\textsuperscript{1}×Å\textsuperscript{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.

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\textsuperscript{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\textsuperscript{19}) based on Ca using Modeller 9.21\textsuperscript{2}. 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\textsuperscript{7}. The whole complex was solvated with TIP3P water molecules\textsuperscript{10} 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 restraints;
(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 gradually increases 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 pymol⁴ 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 structures, we performed the multiple steps as described in Section 1.2.2 for 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 contains 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) was used as the structural basis. Missing atoms and residues were modelled using Modeller 9.21. 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). The sequence of the nucleotide in the post-T state was also mutated by Coot0.8.7 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 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 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) algorithm. Afterwards, we applied Climber algorithm to generate the preliminary pathways to connect the energy minimized pre- and post-T conformations of the SARS-CoV-2 RdRp with dsRNA. 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 was 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. S5). 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 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×mol⁻¹×Å⁻²;
(b) 10,000-steps energy minimization was performed without restraints;
(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×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 NPT equilibration was used to conduct one 10 ns simulation under NPT ensemble (T=300 K, P=1 bar) with position restraints were applied on two nucleotides at the downstream terminal of the template strand. The V-rescale thermostat was applied with a coupling constant of 0.1 ps. The Parrinello-Rahman barostat was applied with a coupling constant of 2 ps. Other simulation parameters were the same as those described in section 1.1.2.