Supplementary Information

Role of 1’-Ribose Cyano Substitution for Remdesivir to Effectively Inhibit both Nucleotide Addition and Proofreading in SARS-CoV-2 Viral RNA Replication

Lu Zhang¹,²*, Dong Zhang¹,², Congmin Yuan³,⁴, Xiaowei Wang³,⁴, Yongfang Li¹,², Xilin Jia¹,², Xin Gao⁵, Hui-Ling Yen⁶*, Peter Pak-Hang Cheung³,⁴*, Xuhui Huang³,⁴*

¹State Key Laboratory of Structural Chemistry, Fujian Institute of Research on the Structure of Matter, Chinese Academy of Sciences, Fuzhou, Fujian, China
²University of Chinese Academy of Sciences, Beijing, China
³Department of Chemistry, Centre of Systems Biology and Human Health, State Key Laboratory of Molecular Neuroscience, The Hong Kong University of Science and Technology, Kowloon, Hong Kong
⁴The Hong Kong University of Science and Technology-Shenzhen Research Institute, Hi-Tech Park, Nanshan, Shenzhen 518057, China
⁵Computational Bioscience Research Center, Computer, Electrical and Mathematical Sciences Engineering Division, King Abdullah University of Science and Technology, Saudi Arabia
⁶School of Public Health, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong

*correspondence could be addressed to: luzhang@fjirsm.ac.cn (L. Z.) or hyen@hku.hk (H.-L. Y.) or ppcheung@ust.hk (P. P.-H. C.) or xuhuihuang@ust.hk (X. H.)
Supplementary Figures:

Fig. S1 Validation of our model by comparing protein-nucleotide interactions with a holo cryo-EM structure (PDB ID: 7BV2). The detailed information about the atom pairs used for the calculations is tabulated in the bottom panel.

<table>
<thead>
<tr>
<th>Index</th>
<th>Chain</th>
<th>Resid</th>
<th>Residue</th>
<th>Interacting atom</th>
<th>Chain</th>
<th>Site</th>
<th>Interacting atom</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>577</td>
<td>Lys</td>
<td>NZ</td>
<td>T</td>
<td>i+3</td>
<td>phosphate backbone</td>
</tr>
<tr>
<td>2</td>
<td>A</td>
<td>595</td>
<td>Tyr</td>
<td>OH</td>
<td>T</td>
<td>i+6</td>
<td>phosphate backbone</td>
</tr>
<tr>
<td>3</td>
<td>A</td>
<td>814</td>
<td>Ser</td>
<td>OG</td>
<td>P</td>
<td>i+1</td>
<td>phosphate backbone</td>
</tr>
<tr>
<td>4</td>
<td>A</td>
<td>836</td>
<td>Arg</td>
<td>NH1/NH2</td>
<td>P</td>
<td>i+2</td>
<td>phosphate backbone</td>
</tr>
<tr>
<td>5</td>
<td>A</td>
<td>858</td>
<td>Arg</td>
<td>NH1/NH2</td>
<td>P</td>
<td>i+4</td>
<td>phosphate backbone</td>
</tr>
<tr>
<td>6</td>
<td>A</td>
<td>684</td>
<td>Asp</td>
<td>O</td>
<td>T</td>
<td>i+1</td>
<td>O2'</td>
</tr>
<tr>
<td>7</td>
<td>A</td>
<td>689</td>
<td>Tyr</td>
<td>OH</td>
<td>T</td>
<td>i+2</td>
<td>O2'</td>
</tr>
<tr>
<td>8</td>
<td>A</td>
<td>592</td>
<td>Ser</td>
<td>OG</td>
<td>T</td>
<td>i+4</td>
<td>O2'</td>
</tr>
<tr>
<td>9</td>
<td>A</td>
<td>592</td>
<td>Ser</td>
<td>OG</td>
<td>T</td>
<td>i+4</td>
<td>O3'</td>
</tr>
<tr>
<td>10</td>
<td>A</td>
<td>759</td>
<td>Ser</td>
<td>OG</td>
<td>P</td>
<td>i+1</td>
<td>O2'</td>
</tr>
<tr>
<td>11</td>
<td>A</td>
<td>865</td>
<td>Asp</td>
<td>OD1/OD2</td>
<td>P</td>
<td>i+4</td>
<td>O2'</td>
</tr>
</tbody>
</table>
Fig. S2 Investigation of nucleotide addition in RdRp when RDV is at $i+3$ or $i+4$ site. The two cartoons in the top horizontal panel denote the site where RDV is positioned (in orange). The three cartoons in the left vertical panel describe the types of distances used for calculation. For clarification, only the molecules involved in the calculations are colored. Each of the six histograms is calculated for the corresponding distance (cartoon in the left vertical panel) using the model with RDV located a specific site (cartoon in top horizontal panel). (A)-(B) Histogram of distance between the P$\alpha$ of ATP and the O3’ atom of the 3’-terminal nucleotide when RDV is at $i+3$ (A) or $i+4$ (B) site. (C)-(D) Histogram of distance between the base of ATP and the base of template nucleotide at $i+1$ site when RDV is at $i+3$ (C) or $i+4$ (D) site. (E)-(F) Histogram of distance between the base of ATP and the base of 3’-terminal nucleotide when RDV is at $i+3$ (E) or $i+4$ (F) site. The histogram of the distances for wildtype-RNA as a reference is shown in light grey. See SI Section 4.1 for details about the base-to-base calculations in (C)-(F).
Fig. S3 Hydrogen bond probability for RDV:U pair at the post-T state with RDV at $i+1$, $i+2$, $i+3$, or $i+4$ site. The top panel is the cartoon model of post-T state with RDV at a specific site (in orange). See SI Section 4.3 for details about the calculation of hydrogen bond probability.
Fig. S4 Interaction between RDV at $i$ site and the surrounding residues in comparison with that of cryo-EM structure (PDBID: 7BV2). (A) Schematic representation of RDV at $i$ site (orange), surrounded by its base-paired template nucleotide (the base of which is shown in cyan) and protein residues. (B) Distances between RDV and its surrounding protein residues/nucleobase calculated using MD conformations (in black circles), in comparison with those calculated from the cryo-EM structure (in red triangles). (C) Details of the atoms used for the distance calculations.
**Fig. S5 Arg858 does not have steric clash with the 1’-cyano group of RDV at i+4 site.** (A) MD snapshot showing the configuration of Arg858 next to RDV at i+4 site. The minimum distance between amino groups of Arg858 and 1’-cyano group of RDV is ~11.4 Å based on an ensemble of 20×30 ns MD conformations with RDV at i+4 site. The distances were calculated by bootstrap algorithm. In each bootstrap sample, 20 trajectories were randomly selected with replacement. The means and standard deviations were estimated by 20 bootstrap samples. (B) Cryo-EM structure (PDBID: 7BV2) with RDV modelled at i-4 site. The minimum distance between amino groups of Arg858 and 1’-cyano group of RDV is 9.5 Å.
Fig. S6 Sequence alignment shows the salt bridge D865-K593+ is conserved among different coronaviruses, excepting that Lys (K) is replaced with Arg (R) in two human coronaviruses. See SI Section 5 for details about the sequence alignment.
Fig. S7 The O3'-MgA distance over time for 20 replicas of 30 ns MD simulations of nsp14-nsp10 complex containing wildtype-RNA.
Fig. S8 The O3'-MgA distance over time for 20 replicas of 30 ns MD simulations of nsp14-nsp10 complex containing single-stranded RNA with RDV at 3'-terminal.
Fig. S9 Illustration of the basepair stepping twist angle.
Fig. S10  δ-δ' repulsion between the bases of fludarabine (FDR)-TP and the template nucleotide destabilizes FDR-TP in the active site (i site) of RdRp. The top panel: A schematic representation of the interactions between the bases of FDR-TP:U pair. The hydrogen bonding interactions are shown in dashed lines. The arrow denotes the δ-δ' repulsion. The bottom panel: Histogram of distance between the C2 atom of FDR-TP (labelled in the top panel) and the δ' charged oxygen atom of template nucleotide when FDR-TP is at i site (shown in light purple), in comparison with the histogram of the corresponding distance when ATP is at i site (shown in dark grey).
Fig. S11 Stability of base pairing and base stacking in RdRp with NTP analogue at either 
i or i+1 site. The left vertical panel describes the molecules used for the distance calculations. 
(A) The distance between the base of ATP/NTP analogue and the base of the template 
nucleotide at the active site when ATP/NTP analogue is at i site. (B) The distance between the 
base of ATP and the base of the template nucleotide at the active site when adenine/NTP 
analogue is at i+1 site. (C) The distance between the base of ATP/NTP analogue and the base 
of 3’-terminal nucleotide when ATP/NTP analogue is at i site. (D) The distance between the 
base of ATP and the base of 3’-terminal nucleotide when NTP analogue is at i+1 site. See SI 
Section 4.1 for details about the base-to-base distance calculations. For the distances in (A)- 
(D), bootstrap algorithm was used to estimate the mean values and standard deviations by using 
20 samples. Each sample contains 20 trajectories randomly selected with replacement from the 
ensemble.
Fig. S12 Stability of the cleavage site when NTP analogue is at 3’-terminal of the nascent strand in ExoN. (A) The distance between two Mg$^{2+}$ ions when adenine/NTP analogue is at the 3’-terminal in ExoN. (B) The twist angle between the adenine/NTP analogue and its neighboring nucleotide when adenine/NTP analogue is at 3’-terminal in ExoN. For the distances in (A) and the angles in (B), bootstrap algorithm was used to estimate the mean values and standard deviations by using 20 samples. Each sample contains 20 trajectories randomly selected with replacement from the ensemble.
Fig. S13 Representative conformation of the cleavage site of the ExoN domain with NTP analogue at 3’-terminal. Asn104 was shown in sphere. See SI Section 4.8 for details.
Supplementary Text:

1. Structural modeling

1.1 nsp12-nsp7-nsp8 complex The cryo-EM structure of nsp12-nsp7-nsp8 complex (PDBID: 6NUR)\(^1\) was used as the basis to construct the RdRp of SARS-CoV-2. The nearly identical amino acid sequences of nsp12-nsp7-nsp8 complex between SARS-CoV and SARS-CoV-2 (~97.1% sequence identity) render our modelling highly feasible. First, after filling in the missing residues (IDs from 897 to 906) in nsp12, we modelled the nsp12, nsp7 and nsp8 of SARS-CoV-2 based on the corresponding protein subunits of SARS-CoV by modeller9.21.\(^2\)

For the homology modelling of nsp12, nsp7 or nsp8, we generated 20 modelled structures and selected the one with the optimal Discrete Optimized Protein Energy (DOPE) assessment score\(^3\) as our final model. Second, the double stranded RNA (dsRNA), ATP and Mg\(^{2+}\) ions in the active site were modelled by structural alignment to the norovirus RdRp (PDBID: 3H5Y\(^4\)) using Pymol\(^5\). To facilitate the modelling of remdesivir and other adenine analogues at \(i\) or \(i+1\) site, we used Coot10.13\(^6\) to mutate the nucleotides in the corresponding sites to ATP:U and A:U base pairs, respectively. Third, after the structural alignment between SARS-CoV-2 and norovirus\(^4\) RdRps, we found Ser682 of SARS-CoV-2 RdRp is homologous to the Ser300 of norovirus RdRp, but their side chains are at different orientations. To maintain interactions between the side chain of Ser682 and the O2’ atom of ATP, we replaced the coordinates of Ser682 with those of Ser300 from the aligned norovirus RdRp\(^4\). Fourth, the protonation states of histidine residues were predicted using propka3.0 module\(^7\) in the pdbpqr2.2.1\(^8\) package, followed by manual inspection to ensure that the coordination between the Zn\(^{2+}\) ion and the corresponding histidines (residue IDs 695 and 242 in nsp12) was maintained. Accordingly, histidines with residue IDs 295, 309, 642, 872, and 892 in nsp12, as well as histidine with residue ID 36 in nsp7, have N\(_e\) atom protonated; the remaining histidine residues have N\(_\delta\) atom protonated. The whole complex was placed in a dodecahedron box with the box edges at least 12 Å away from the complex surface. The box was filled with TIP3P water molecules\(^9\), and sufficient counter ions were added to neutralize the whole system. This nsp12-nsp7-nsp8 model containing wildtype RNA with ATP in the active site (\(i\) site) serves as the structural basis to model NTP analogues in RdRp.

1.2 nsp14-nsp10 complex The crystal structure of nsp14-nsp10 complex of SARS-CoV (PDBID: 5C8S\(^10\)) serves as our modelling template to build the nsp14-nsp10 complex of SARS-CoV-2. First, due to the ~95.7% sequence identity of nsp14-nsp10 complex between
SARS-CoV and SARS-CoV-2, we constructed the model of nsp14-nsp10 complex of SARS-CoV-2 directly by using Pymol\textsuperscript{3} to generate the mutations based on the crystal structure of SARS-CoV\textsuperscript{10} and ensuring that the side chains of mutated residues could maintain the original orientations. Second, we used modeller\textsuperscript{9.21} to fill in the missing amino acids (residue IDs 454-464) in nsp14. Third, the ExoN domain of SARS-CoV shares a similar architecture as the proofreading ExoN domain of DNA polymerase I Klenow fragment (PDBID: 1KLN\textsuperscript{11}) and the ε-subunit of DNA polymerase III (PDBID: 1J53\textsuperscript{12}). Hence, we modelled the single-stranded RNA and Mg\textsuperscript{2+} ions in the catalytic cleavage site by aligning the protein residues in the proofreading domains of these proteins. Specifically, D90, E92, E191, D273, and H268 in nsp14 of SARS-CoV-2, D12, E14, D103, D167 and H162 in the ε-subunit of DNA polymerase III, and E357, D424, D501 and Y497 of ExoN domain of DNA polymerase I were used for the structural alignment. The modelled single-stranded RNA contains three nucleotides, because three base pairs are required to be melted to allow the 3’-terminal of nascent strand to access the cleavage site\textsuperscript{11}. In particular, the 3’-terminal nucleotide and two Mg\textsuperscript{2+} ions were modelled based on the structural alignment to the ε-subunit of DNA polymerase III\textsuperscript{12}, while the remaining two nucleotides were modelled by aligning to the ExoN domain of DNA polymerase I Klenow fragment\textsuperscript{11}. Fourth, the alignment between ExoN domains of SARS-CoV-2 and ε-subunit of DNA polymerase III\textsuperscript{12} indicate the protein residues (D90, E92, D273 and H268) in SARS-CoV-2 are homologous to the D12, E14, D167 and H162 in DNA polymerase III. To maintain the coordination between these residues and the Mg\textsuperscript{2+} ions in the cleavage site, we replaced the coordinates of D90, E92, D273 and H268 with those of their homologous residues in the aligned ε-subunit of DNA polymerase III\textsuperscript{12}. We also extracted water molecules coordinated with the Mg\textsuperscript{2+} ions from the aligned ε-subunit of DNA polymerase III\textsuperscript{12} to maintain their coordination with the Mg\textsuperscript{2+} ions. Fifth, to facilitate the modelling adenine analogue at the 3’-terminal, we used Coot\textsuperscript{10.13} to mutate the 3’-terminal nucleotide to adenine. Sixth, protonation states of histidine residues were predicted using the procedure illustrated in Section 1.1 for nsp12-nsp8-nsp7 complex. Specifically, residues 26, 264, 268, 330, 373, 424, 486 and 487 in nsp14 were protonated at N\textsubscript{6}, while other histidine residues were protonated at N\textsubscript{ε}. We did not include the methyl donor and acceptor in our current model, because their binding to the MTase domain of nsp14 does not show obvious structural variations\textsuperscript{10}. The whole complex was solvated with TIP3P water molecules\textsuperscript{9} in a dodecahedron box. The box edges are at least 12 Å away from the complex surface. Enough counter ions were added to neutralize the whole system. This nsp14-nsp10 model has wildtype RNA with
adenine at the 3’-terminal in ExoN domain and serves the structural basis to model the NTP analogues in ExoN.

2. Molecular Dynamics Simulation Set-up

2.1 Force field parameters We used amber99sb-ildn force field\textsuperscript{13} to simulate protein and RNA. To generate force field parameters of NTP analogues including remdesivir, favipiravir, vidarabine and fludarabine, we followed a similar scheme as used in amber99sb-ildn force field\textsuperscript{13}. In particular, the initial structures for generating force field parameters for remdesivir and favipiravir were built by modifying the natural adenine from 3’-terminal of nascent strand in nsp12 to the corresponding NTP analogues. For generating force field parameters of vidarabine and fludarabine, the initial structures were acquired from DrugBank\textsuperscript{14}. When deriving the partial charges for NTP analogues, both the 3’- and 5’-terminals of the NTP analogues were truncated by hydroxyl groups. Geometry optimization was performed using HF/6-31G*, followed by single-point calculation with the same method and basis set in Gaussian16\textsuperscript{15}. The obtained electrostatic potential was used to generate the partial charges using the Restricted Electrostatic Potential approach\textsuperscript{16-17}. To make the partial charges of NTP analogues compatible with the existing parameters of RNA molecules, the partial charges of 5’-terminal and 3’-terminal hydroxyl groups were assigned with the empirical values (H5T=0.4295, O5’=–0.6223, H3T=0.4376, O3’=–0.6541) as used in amber99sb-ildn force field\textsuperscript{13}, and kept constant in the charge fitting. Bonded and Lennard-Jones parameters for NTP analogues were derived from those used for adenine in the amber99sb-ildn force field\textsuperscript{13}, excepting those with obvious discrepancy from adenine were instead from General Amber Force Field\textsuperscript{18-19}. They include the 1’-cyano group as well as the five-membered ring for remdesivir, the base for favipiravir and the fluorine-involved interactions for fludarabine. For the NTP analogues or adenine in the triphosphate form, parameters for the triphosphate tail were taken from those developed by Meagher \textit{et al.}\textsuperscript{20}.

2.2 MD simulations for nsp12-nsp7-nsp8 complex

2.2.1 MD simulations for nsp12-nsp7-nsp8 complex containing wildtype RNA with ATP in the active site (i) site

We performed multiple steps of energy minimization and position restraint simulations to gradually relax and fully equilibrate the simulation complex containing wildtype RNA with ATP in the active site as follows:
(a) One 10,000-steps energy minimization on the whole system by position restraining the heavy atoms of nucleotides and Mg$^{2+}$ ions with a force constant of 10 kJ×mol$^{-1}$×Å$^{-2}$;
(b) Another 10,000-steps energy minimization without restrain;
(c) One 200 ps position restraint simulation under NVT ensemble (T=300K) with a force constant of 10 kJ×mol$^{-1}$×Å$^{-2}$ on all the heavy atoms of the complex;
(d) Another 500 ps position restraint simulation under NPT ensemble (T = 300 K, P=1 bar) with a force constant of 10 kJ×mol$^{-1}$×Å$^{-2}$ on all the heavy atoms of the complex;
(e) One 10 ns simulation under NPT ensemble (T = 298 K, P = 1 bar) by releasing the position restrain on protein while retaining the position restrain on nucleotides and Mg$^{2+}$ ions;
(f) Another 10 ns simulation under NPT ensemble (T = 298 K, P = 1 bar) was performed without position restraint;
(g) One 100 ns simulation under NPT ensemble (T = 298 K, P = 1 bar) to fully equilibrate the whole system at T = 298 K and P = 1 bar.

It’s worthy to note that the structural alignment between SARS-CoV-2 RdRp and norovirus RdRp enables us to pinpoint the residues that may be important for stabilizing the active site. Therefore, we maintained these interactions by adding harmonic constraints in steps (a)-(f). These interactions include: (i) MgA with its coordinated Asp618, Asp760, Asp761 and one P$_\alpha$ oxygen of ATP; (ii) MgB with Asp618, Tyr619, Asp760 and three oxygen atoms from P$_\alpha$, P$_\beta$ and P$_\gamma$ of ATP; (iii) hydrogen bond between the 2'-hydroxyl group of ATP and Asn691$^{1,4}$; (iv) hydrogen bond between the 2'-hydroxyl group of ATP and Thr680$^1$; (v) hydrogen bond between the 3'-hydroxyl group of ATP and Asp623$^{1,4}$. The harmonic constraints were removed in step (g) for full equilibration of the whole system.

The last configuration of the 100 ns simulation was used to seed 20 replicas of 30 ns MD simulations under NVT ensemble, with simulated annealing from 50 K to 300 K in the first 2 ns. In the 20×30 ns simulations, we added the harmonic constraint to maintain the coordination between the Zn$^{2+}$ ions with their coordinated protein residues (Zn$^{2+}$ ions are located distantly (> 25 Å) from the ATP in the active site): one Zn$^{2+}$ ion is coordinated by His295, Cys301, Cys306 and Cys310; the other one is coordinated by Cys487, His642, Cys645 and Cys646. This is to avoid the Zn$^{2+}$ ions, especially the ones on the surface, from diffusing into the solvent. We applied V-rescale thermostat with the coupling time constant of 0.1 ps. The long-range electrostatic interactions beyond the cut-off at 12 Å were treated with the Particle-Mesh Ewald method. Lennard-Jones interactions were smoothly switched off from 10 Å to 12 Å. The
neighbour list was updated every 10 steps. An integration time step of 2.0 fs was used and the LINCS algorithm\textsuperscript{23} was applied to constrain all bonds. We saved the snapshots every 20 ps, and conformations after 5 ns were collected for subsequent structural analyses. All simulations were performed with Gromacs 5.0\textsuperscript{24}.

### 2.2.2 MD simulations for nsp12-nsp7-nsp8 complex containing NTP analogues

We first modelled each NTP analogue in the corresponding site of the nascent strand based on the last configuration of the 100 ns equilibrated simulation containing wildtype RNA with ATP in the active site (step (g) in Section 2.2.1):

- Modeling NTP analogues at \(i\) site: we replaced ATP with each NTP analogue in the triphosphate form at the active site (\(i\) site);
- Modeling NTP analogues at \(i+1\) site: we substituted the 3’-adenine with each NTP analogue;
- Modeling the post-translocation (post-T) state with remdesivir at \(i+2, i+3\) or \(i+4\) site: we replaced the nucleotide in the corresponding site with remdesivir;
- Modeling the pre-translocation (pre-T) state with remdesivir at \(i, i+1, i+2\) or \(i+3\) site: we added covalent bond between the O3’ atom of 3’-adenine and the P\(\alpha\) atom of ATP, and replaced the triphosphate tail with hydrogen atom. We removed MgB from the model. The nucleotide at corresponding upstream site of the nascent strand was substituted with remdesivir.

For each model with NTP analogue, the whole system was relaxed gradually for the equilibration:

(a) One 10,000-steps energy minimization with position restraint (10 kJ\(\times\)mol\(^{-1}\times\text{Å}^{-2}\)) on the heavy atoms of nucleotides and Mg\(^{2+}\) ions;
(b) Another 10,000-steps energy minimization without position restraint;
(c) The energy minimized system was further relaxed by 200 ps simulation under NVT ensemble (\(T=300\) K) with position restrain (10 kJ\(\times\)mol\(^{-1}\times\text{Å}^{-2}\)) on all the heavy atoms of the system;
(d) Another 500 ps simulation under NPT ensemble (\(T=300\) K, \(P=1\) bar) with position restrain (10 kJ\(\times\)mol\(^{-1}\times\text{Å}^{-2}\)) on all the heavy atoms of the system.

The last configuration of the position restraint simulation was utilized to run 20 replicas of 30 ns NVT simulations (\(T=300\) K) with different random seeds. Temperature was gradually
increased from 50 K to 300 K in the first 2 ns using simulated annealing. In the 20×30 ns simulations, harmonic constraint was added between the Zn$^{2+}$ ions and their coordinated protein residues to maintain their coordination (one Zn$^{2+}$ ion with His295, Cys301, Cys306 and Cys310; the other Zn$^{2+}$ ion with Cys487, His642, Cys645 and Cys646). Same parameters as used in Section 2.2.1 were utilized for MD simulations. MD snapshot was saved every 20 ps and conformations after 5 ns were collected for subsequent structural analyses.

2.3 MD simulations for nsp14-nsp10 complex

2.3.1 MD simulations for nsp14-nsp10 complex containing wildtype RNA with adenine at the 3’-terminal of nascent strand

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

(a) One 10,000-steps energy minimization with position restraint (10 kJ×mol$^{-1}$×Å$^{-2}$) on the heavy atoms of nucleotides and Mg$^{2+}$ ions;
(b) Another 10,000-steps energy minimization on the whole system without position restraint;
(c) One position restraint simulation of 200 ps under NVT ensemble (T=300 K) by using a restraint force constant of 10 kJ×mol$^{-1}$×Å$^{-2}$ on all the heavy atoms of the nsp14-nsp10 complex;
(d) Another 500 ps NPT ensemble (T=300 K, P=1 bar) with position restraint (10 kJ×mol$^{-1}$×Å$^{-2}$) on all the heavy atoms of the nsp14-nsp10 complex;
(e) One 10 ns NPT simulation (T=300 K, P=1 bar) with position restraint on the nucleotides (10 kJ×mol$^{-1}$×Å$^{-2}$) and Mg$^{2+}$ ions (100 kJ×mol$^{-1}$×Å$^{-2}$). In this step, we added a stronger restraining force constant (100 kJ×mol$^{-1}$×Å$^{-2}$) on Mg$^{2+}$ ions, in order to reduce the perturbation caused by the full relaxation of the surrounding protein residues;
(f) The position restraint on nucleotides was released and the system was simulated for another 30 ns under NVT ensemble (T=300 K) for full equilibration. Simulated annealing was performed from 50 K to 300 K in the first 2 ns.

We have added harmonic constraint between Mg$^{2+}$ ions and their coordinated residues in steps (a)-(d). The coordination network of two Mg$^{2+}$ ions includes: MgA with carboxyl group of Asp90, oxygen atom in the phosphate backbone of 3’-adenine and the O3’ atom of the guanine next to the 3’-adenine; MgB with carboxyl group of Asp90, carboxyl group of Asp273,
carboxyl group of Glu92 and oxygen atom in the phosphate backbone of 3’-adenine. The constraint has been removed after step (d).

The last configuration of the 30 ns simulation was used to seed 20 replicas of 30 ns NVT simulations (T=300 K), including the simulated annealing from 50 K to 300 K in the first 2 ns. It’s worthy to note that when NTP analogue is incorporated, the nascent RNA strand will backtrack from nsp12 and protrude into the cleavage site of nsp14 for excision. However, there is no available structural information for the binding interfaces between nsp12 and nsp14. Therefore, in order to recapitulate the condition in which the backtracked RNA enters the active site of nsp14 through the binding interface between nsp12 and nsp14, we placed a position restraint (100 kJ×mol⁻¹×Å⁻²) on the 5’-nucleotide and the phosphate atom of the neighboring nucleotide to mimic the conformational constrain imposed by the upstream RNA strand. In addition, we also added a weak position restraint (10 kJ×mol⁻¹×Å⁻²) on the Mg²⁺ ions to maintain their coordination. Furthermore, harmonic constraint was added between the Zn²⁺ ions and their coordinated protein residues to maintain their coordination (Zn²⁺ ions are located distantly (> 14 Å) from nucleotides in the cleavage site). The coordination network of the five Zn²⁺ ions includes: (i) Zn²⁺ ion (residue ID 201 in nsp10) with Cys74, Cys77, His83 and Cys90; (ii) Zn²⁺ ion (residue ID 202 in nsp10) with Cys117, Cys120, Cys128 and Cys130; (iii) Zn²⁺ ion (residue ID 601 in nsp14) with Cys207, Cys210, Cys226 and His229; (iv) Zn²⁺ ion (residue ID 602 in nsp14) with His257, Cys261, His264 and Cys279; (v) Zn²⁺ ion (residue ID 603 in nsp14) with Cys452, Cys477, Cys484 and His487. The same parameters as used above for the RdRp production simulations (Section 2.2.1) were applied. The MD snapshot was saved every 20 ps and conformations after 5 ns were collected for subsequent structural analyses.

2.3.2 MD simulations for nsp14-nsp10 complex containing each NTP analogue at the 3’-terminal of nascent strand

We constructed the system containing NTP analogue at the 3’-terminal of nascent strand based on the last configuration of the 30 ns equilibrated simulation containing 3’-adenine. Specifically, the 3’-adenine was replaced with each of the NTP analogues.

For each system, we performed multiple steps of energy minimization and position restraint simulations for full equilibration as follows:

(a) One 10,000-steps energy minimization with position restraint (10 kJ×mol⁻¹×Å⁻²) on the heavy atoms of nucleotides and Mg²⁺ ions;
(b) Another 10,000-steps energy minimization on the whole system without position restraint;

(c) One 200 ps NVT (T=300 K) position restraint simulation with a force constant of 10 kJ mol\(^{-1}\) Å\(^{-2}\) on all the heavy atoms of the complex;

(d) Another 500 ps NPT (T=300 K, P=1 bar) position restraint simulation with a force constant of 10 kJ mol\(^{-1}\) Å\(^{-2}\) on all the heavy atoms of the complex.

The last configuration of the position restraint simulations was used to randomly seed 20 replicas of 30 ns MD simulations under NVT ensemble (T=300 K), with simulated annealing from 50 K to 300 K in the first 2 ns. Harmonic constraint was added to maintain the coordination between Zn\(^{2+}\) ions and their coordinated protein residues, as described in the Section 2.3.1. Position restraint was applied on the 5’-terminal nucleotides and Mg\(^{2+}\) ions in the same way as used in the production simulation in Section 2.3.1. The MD snapshot was saved every 20 ps and conformations after 5 ns were collected for subsequent structural analyses. Same parameters as used above for the RdRp simulations (Section 2.2.1) were utilized.

3. Generation of translocation pathway

To investigate if the translocation of remdesivir from \(i+3\) to \(i+4\) site is inhibited, we used the Climber algorithm\(^{25-26}\) to generate a translocation pathway.

3.1 Modelling the pre-T and post-T states of SARS-CoV2-RdRp. First, we constructed the pre-T conformation based on the configuration of pre-T state with remdesivir at \(i+3\) site after position restraint simulations (Section 2.2.2) and replaced remdesivir with adenine. Second, we constructed post-T conformation based on the configuration of post-T state with remdesivir at \(i+4\) site after position restraint simulations (Section 2.2.2) and replaced remdesivir with adenine. Third, in the translocation, all the nucleotides are translocating forward by one base pair position. Our structural model of SARS-CoV-2 RdRp contains only the coordinates of nucleotides from \(i-1\) to \(i+7\) site, which enables us to simulate the translocation of nucleotides from \(i-1\) to \(i+6\) site in the pre-T conformation to \(i\) to \(i+7\) site in the post-T conformation. Therefore, we removed the most upstream base pair of nucleotides at \(i+7\) site in the pre-T conformation, as well as the ATP and the most downstream nucleotide at \(i-1\) site in the post-T conformation. Nucleotides were mutated by Coot\(^{10.13}\) to make the nucleotides’ sequence consistent between pre-T and post-T states. Fourth, energy minimization was performed for the pre-T and post-T conformations using the Energy Calculation and Dynamics (ENCAD)
simulation program. The energy minimized pre-T and post-T conformations were used as the initial and final conformation for generating the preliminary pathway, respectively.

3.2 Generation of translocation pathway. We used Climber algorithm to generate a low-energy pathway for the translocation of double-stranded RNA (dsRNA) containing an adenine moving from $i+3$ to $i+4$ site in SARS-CoV-2 RdRp. An external energy was applied on dsDNA to drive the system from the initial pre-T state to the final post-T state. In particular, this external energy is consisted of a series of harmonic potentials applying on distances between atom pairs (one atom belongs to dsRNA, while the other atom belongs to protein and ions). To generate the translocation pathway, we performed a 350-step Climber simulation, which gradually drives the dsRNA from the pre-T state to the post-T state. In this simulation, the system has succeeded in translocation with a RMSD of 1.8 Å to the post-T state. Next, we chose conformations along this translocation pathway every 10 steps, and replaced the adenine (translocating from $i+3$ to $i+4$ site) with remdesivir. Subsequent energy minimization of each of these conformations was then performed by Gromacs 5.0.

4. Structural analyses

4.1 Base-to-base distance for the base pairing/base stacking at the active site of RdRp
Adenine and NTP analogues have two conjugated rings in the base, excepting that favipiravir has only one six-membered ring in the base. In this regard, to make the base-to-base distance calculations consistent for adenine and all NTP analogues, we used the center of mass (c.o.m.) of the ring directly connected to the ribose to determine the base position of the nucleotides involved in the distance calculations. Only heavy atoms constituted the ring were considered in the c.o.m. calculation.

4.2 Root Mean Square Deviation (RMSD) calculations of MD conformations based on our structural model against the cryo-EM structures
We compared RMSDs of protein residues that are present in both of our structural model and cryo-EM structures (cryo-EM structures contain longer N-terminal tail). In particular, we used the Cα atoms of residues 117 to 919 (excluding residues 896 to 905) for comparing the structural similarity of protein between our structural model and the apo cryo-EM structure (PDBID: 6M71). We included the Cα atoms of residues 118 to 919 (excluding residues 896 to 905) for comparing the protein in our structural model with that in the apo (PDBID: 7BV1) and holo (PDBID: 7BV2) cryo-EM structures. When comparing the structural similarity of nucleic acids, we only included
nucleotides present in both our structural model and the holo cryo-EM structure. Specifically, nucleotides in the template strand from $i-1$ to $i+7$ sites and in the nascent strand from $i+1$ to $i+7$ sites were included, and their phosphate backbones and heavy atoms in the ribose ring were used for the RMSD calculations. RMSDs were calculated based on the MD conformational ensemble using bootstrap algorithm. In each bootstrap sample, 20 trajectories were randomly selected with replacement. The mean values and standard deviations were estimated by 20 bootstrap samples.

4.3 Hydrogen bond probability of base pairs Hydrogen bond probability was calculated by the apparent number of hydrogen bonds divided by the optimal number of hydrogen bonds. Bootstrap algorithm was used by generating 20 bootstrap samples to calculate hydrogen bond probability. For each sample, 20 trajectories were randomly selected with replacement.

4.4 Interaction energy between 1’-cyano group and the side chains of K593 and D865 We collected the MD conformations of pre-T state with remdesivir at $i+3$ site to calculate the interaction energy between 1’-cyano group of remdesivir and the side chains of K593 and D865. Only the conformations with remdesivir:U in their canonical base paired configuration (two hydrogen bonds formed between remdesivir and uracil) were considered in the calculations. To set up the calculations, both electrostatic interactions and Lennard-Jones interactions were cut off at 30 Å. Water and counter ions were removed to simply the calculations. The side chain of K593, the side chain of D865 and the 1’-cyano group of remdesivir were treated as three energy groups, and the non-bonded interaction energies among the three energy groups were calculated. Bootstrap method was applied to estimate the means and standard deviations. Specifically, after removing the MD conformations in the first 5 ns, canonical base paired configurations (two hydrogen bonds for the remdesivir:U pair) were observed in 17 simulations. Thus, we generated 17 bootstrap samples, each of which contains 17 random trajectories selected with replacement. Afterward, the mean values and standard deviations were derived from the 17 samples.

4.5 Distance between 1’-cyano group and K593/D865/S861 in the translocation of remdesivir from $i+3$ to $i+4$ site The nitrogen atom in the 1’-cyano group of remdesivir was used to calculate its distance to the nitrogen in the quaternary amine group of K593 and the backbone oxygen atom of S861. Because the two oxygen atoms in the carbonyl group of D865 are chemically equivalent, we calculated the minimum distance between the nitrogen atom in
the 1’-cyano group of remdesivir and the two oxygen atoms in the carbonyl group of D865 as the “cyano-D865 distance” in Fig. 4C.

4.6 RMSD to the pre-T state in the translocation of remdesivir from i+3 to i+4 site We aligned the energy minimized conformations along the translocation pathway to the energy minimized pre-T conformation by the Cα atoms of protein and then calculated RMSD of the heavy atoms of nucleotides and remdesivir.

4.7 Distance between Asn104 and adenine or NTP analogue at the 3’-terminal site in ExoN To make the calculations consistent for adenine and all NTP analogues (as explained in 3.1), we used the c.o.m. of the ring directly connected to the ribose to determine the base position of 3’-terminal adenine/NTP analogue. For Asn104, the nitrogen atom in the amide group was used for the distance calculations.

4.8 Selection of representative conformations in ExoN We performed K-center clustering based on the RMSD of heavy atoms in nucleotides and Asn104. The MD conformations were divided into 10 clusters. Only the center conformation in the cluster with population > 10% is present in Figs. 5E, 5F and S13.

5. Sequence alignment

The sequence of nsp12 SARS-CoV-2 was compared with the GenBank sequences of several CoVs (Fig. S6): Human coronavirus NL63 (HCoV-NL63, accession code YP_003766.2), swine acute diarrhea syndrome coronavirus (SADS-CoV, accession code QID98967.1), Human coronavirus NL63 (HCoV-HKU1, accession code AGW27852.1), Middle East respiratory syndrome-related coronavirus (MERS-CoV, accession code YP_009047223.1), SARS-CoV (accession code AEA10937.1). The sequence alignment was performed by Clustal Omega1.2.4.
6. Reference


