Chem Catalysis, Volume *2*

Supplemental information

Mechanism of reaction of RNA-dependent RNA polymerase from SARS-CoV-2

Juan Aranda, Milosz Wieczór, Montserrat Terrazas, Isabelle Brun-Heath, and Modesto Orozco

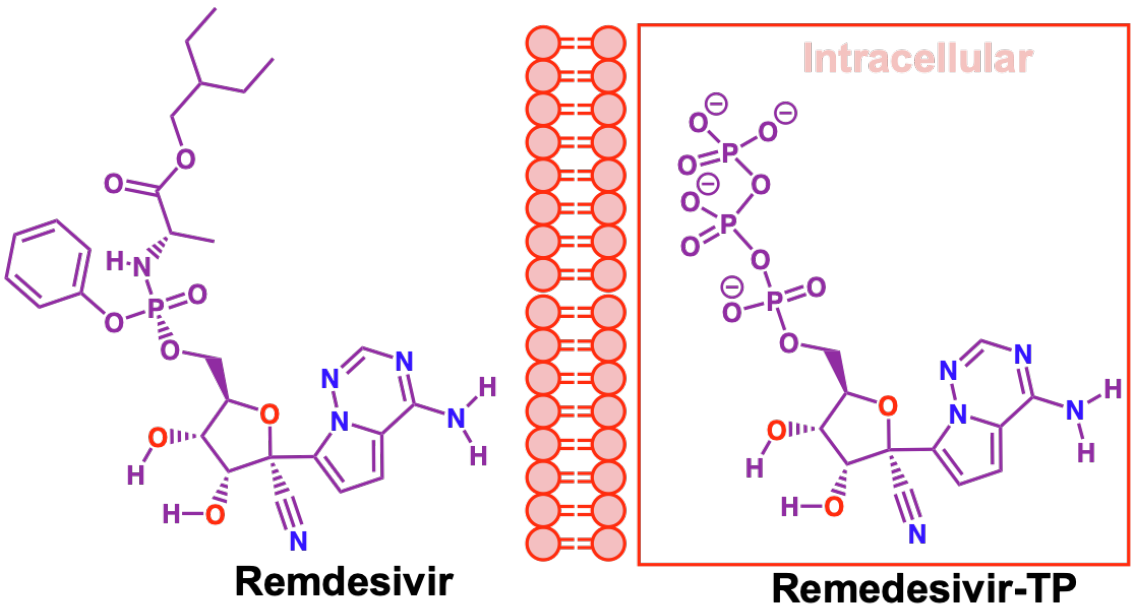

Figure S1. Chemical structures of Remdesirvir in its prodrug form and as a triphosphate nucleoside inside the cell.

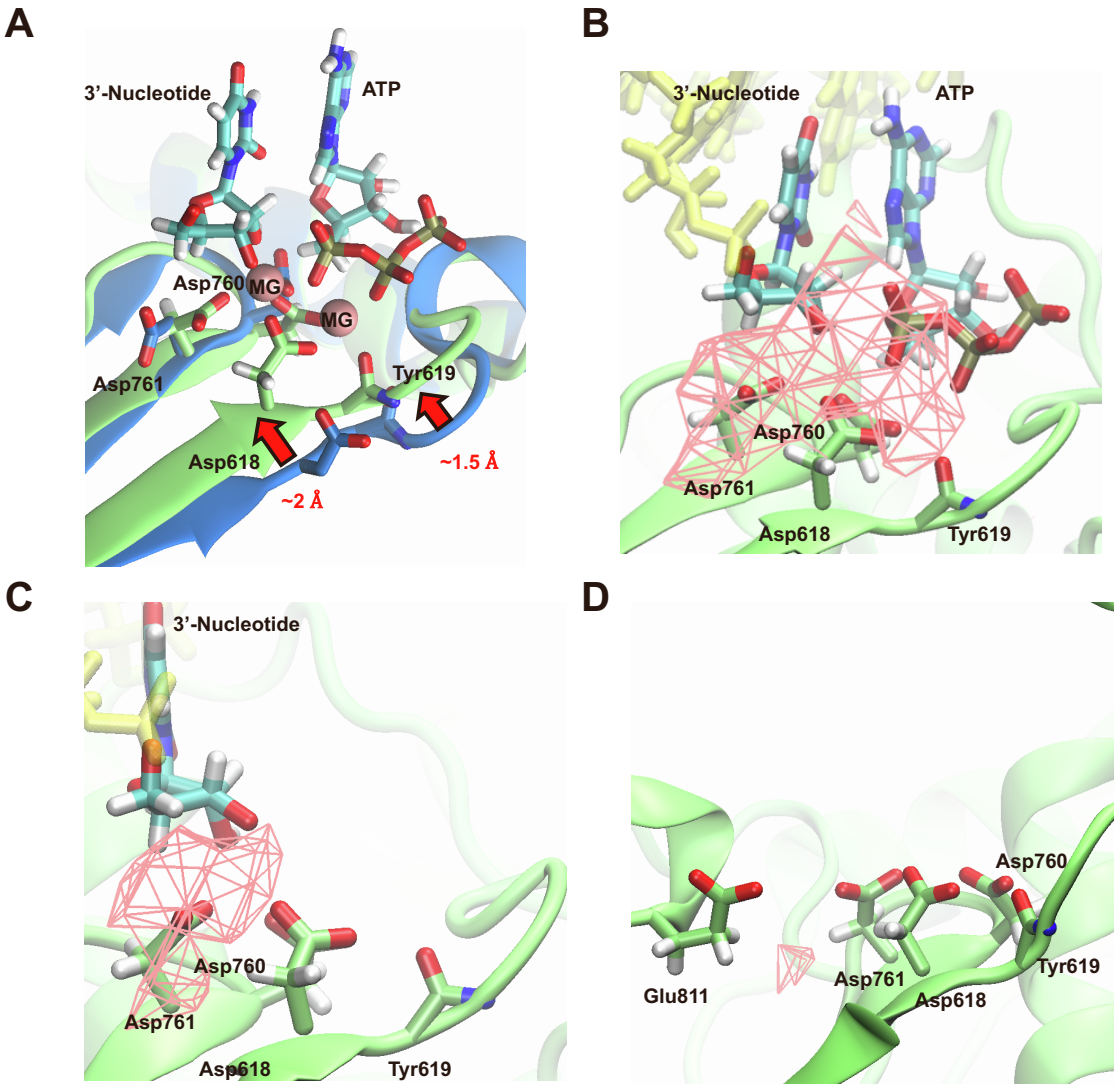

Figure S2. A Structural alignment of SARS-CoV-2 RdRp (PDB ID: 6M71, in blue) with the obtained catalytically competent structure of this study (in green). Values next to arrows indicate the difference (in Å) between structures, measured as the distance between Cα's corresponding to residues Asp_{618} and Tyr_{619}, respectively. **B** Classical molecular interaction potential (CMIP) calculations performed in the system containing RdRp with an RNA duplex and an ATP molecule displays a high interaction energy region (displayed as a pink wireframe) which corresponds to the binding of two Mg^{2+} cations. **C** CMIP calculations performed in the system containing RdRp with an RNA duplex displays a high energy region which corresponds to the binding of one Mg^{2+} cation. **D** CMIP calculations performed in RdRp alone show a very low interaction energy region, which is displaced from the active site residues towards Glu_{811}.

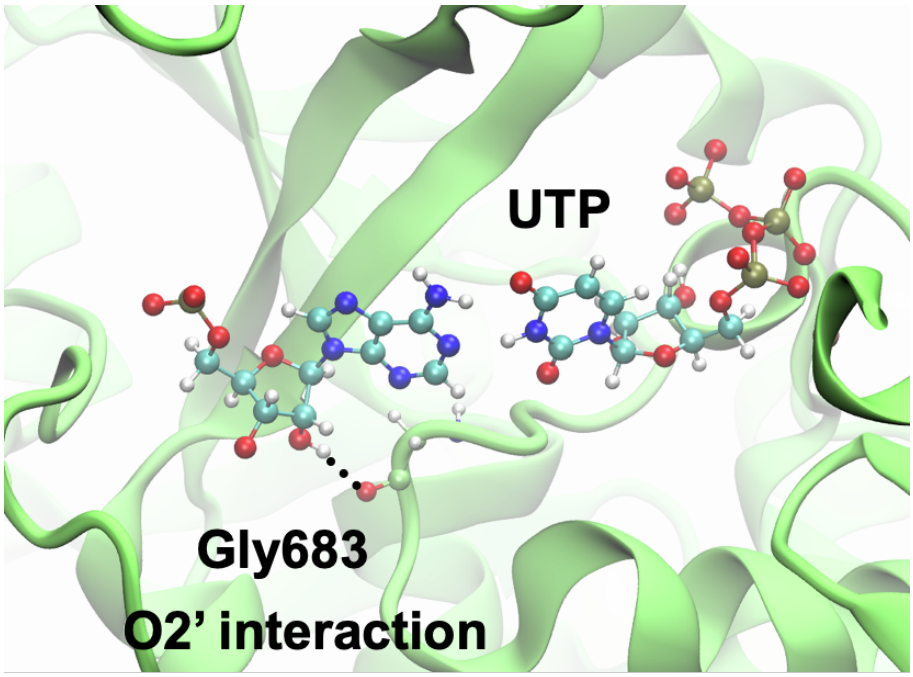

Figure S3. Gly_{683} forms a hydrogen bond interaction with the O2' hydroxyl group of the template nucleotide in "i" position, recognizing the entry of a ribonucleotide template.

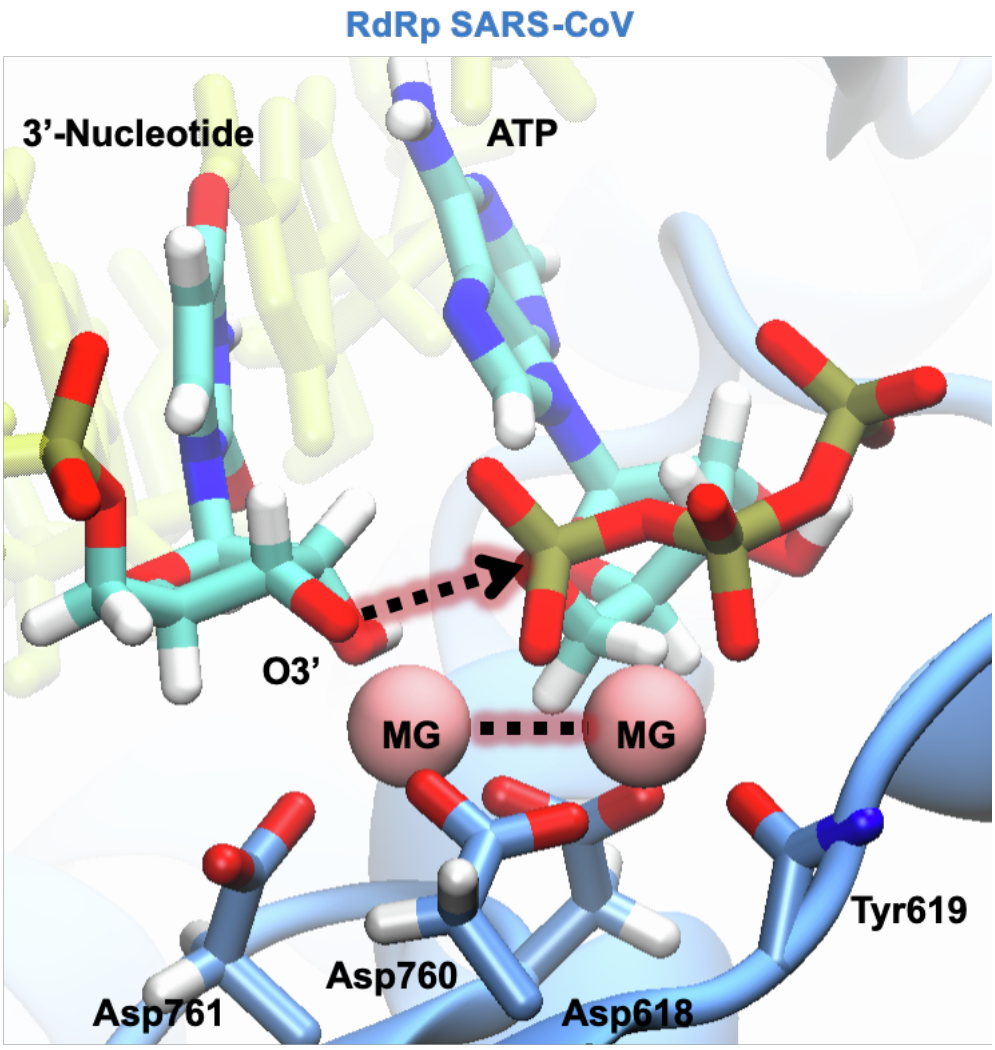

Figure S4. SARS-CoV RdRp's active site structure displays a very similar arrangement as compared to SARS-CoV-2 RdRp's during our 0.5 μs MD simulations.

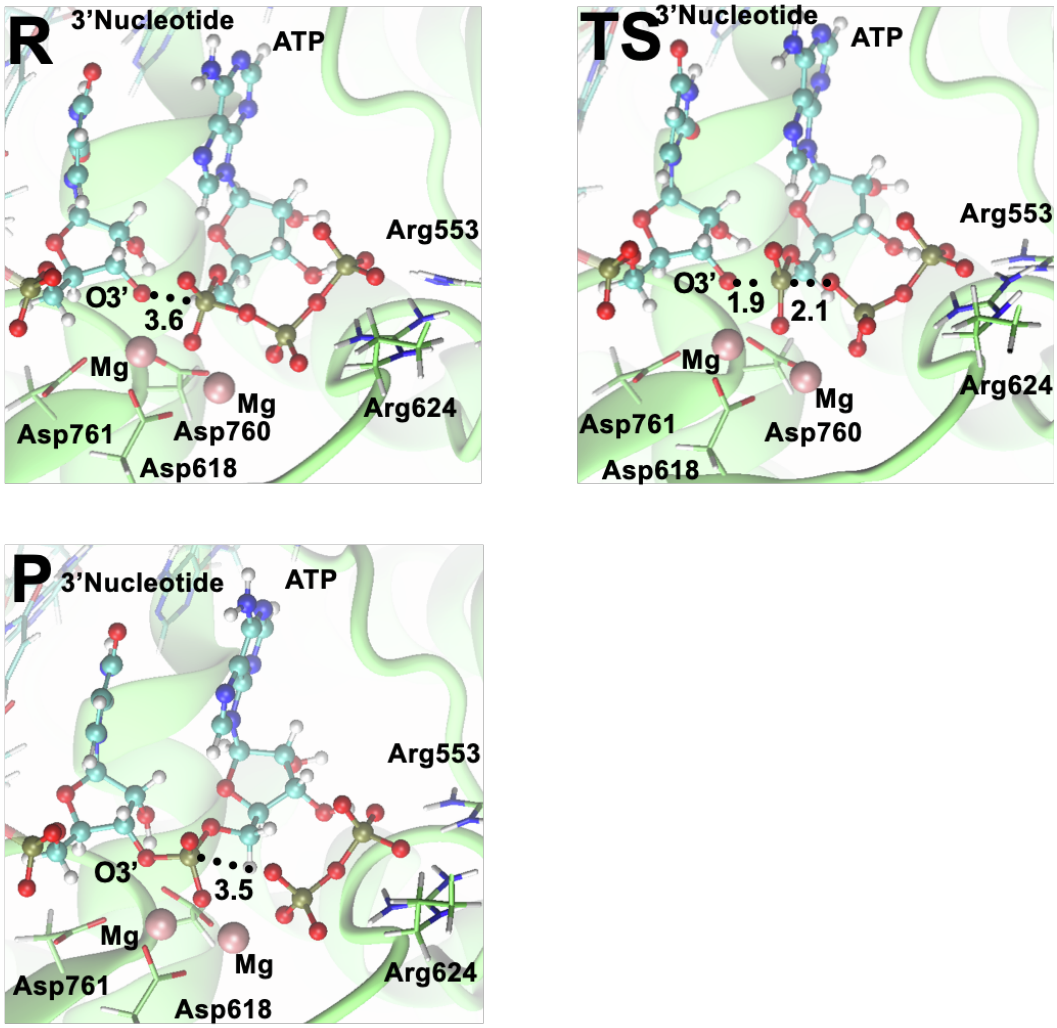

Figure S5. Active site insights of RdRp when a A is being incorporated to a nascent viral RNA strand. Reactant, Transition and Product states representative structures are depicted. Average distances involved in the phosphoryl transfer are shown in Å and depicted as dotted lines.

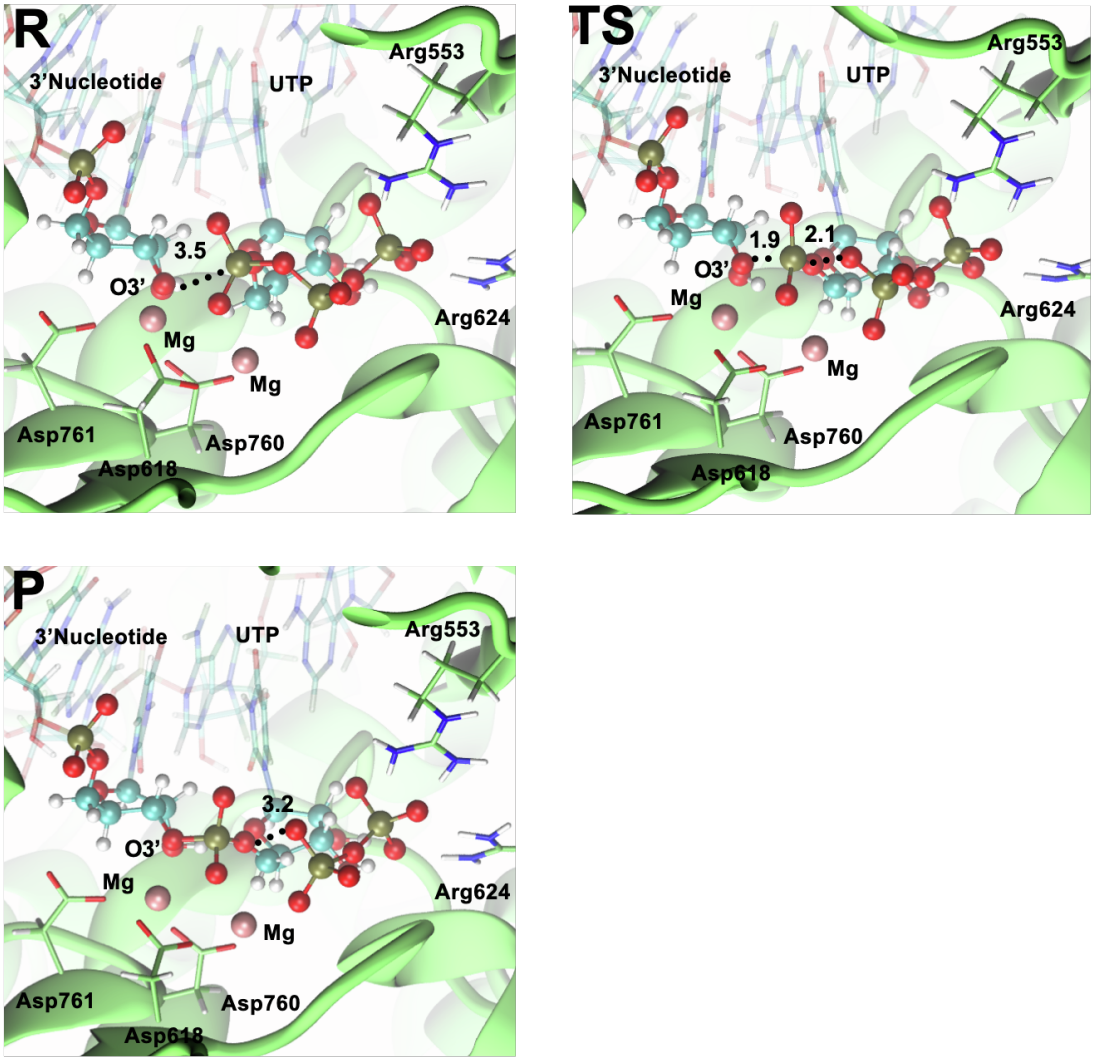

Figure S6. Active site insights of RdRp when a U is being incorporated to a nascent viral RNA strand. Reactant, Transition and Product states representative structures are depicted. Average distances involved in the phosphoryl transfer are shown in Å and depicted as dotted lines.

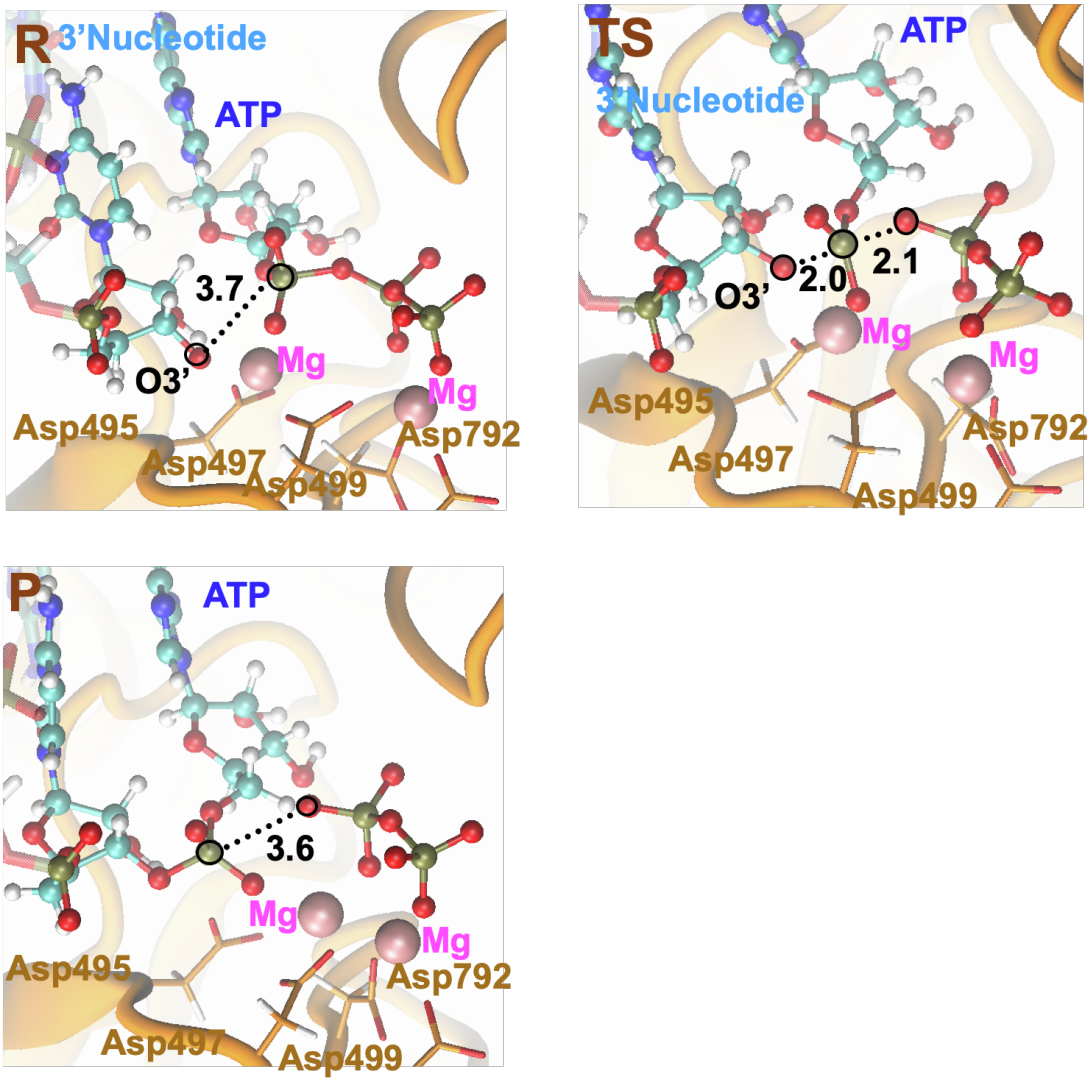

Figure S7. Active site insights of human RNA Pol II (shown in orange) when an A is being incorporated to a nascent RNA strand. Reactant, Transition and Product states representative structures are depicted. Average distances involved in the phosphoryl transfer are shown in Å and depicted as dotted lines.

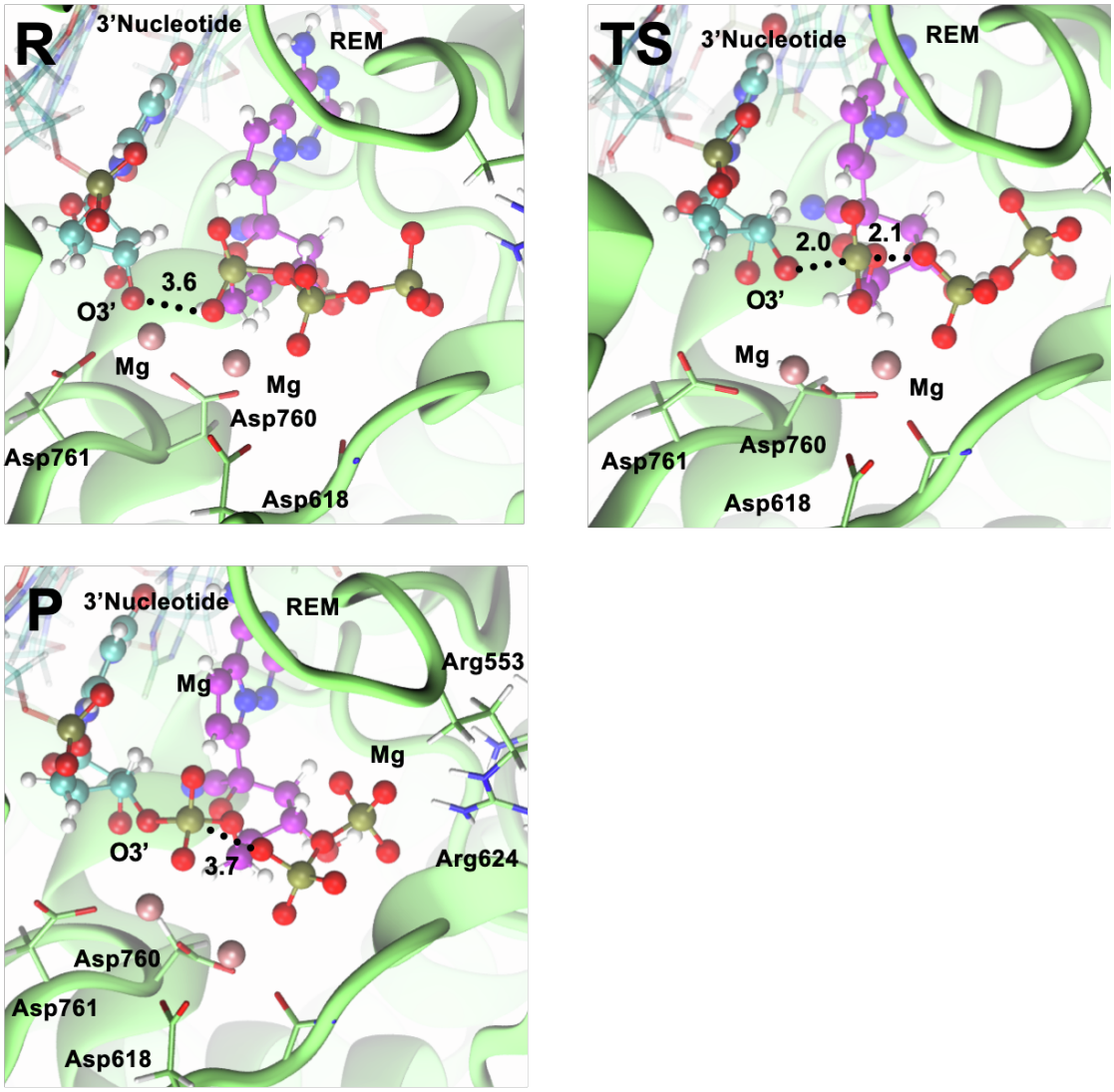

Figure S8. Active site insights of RdRp when a remdesivir is being incorporated to a nascent viral RNA strand. Reactant, Transition and Product states representative structures are depicted. Average distances involved in the phosphoryl transfer are shown in Å and depicted as dotted lines.

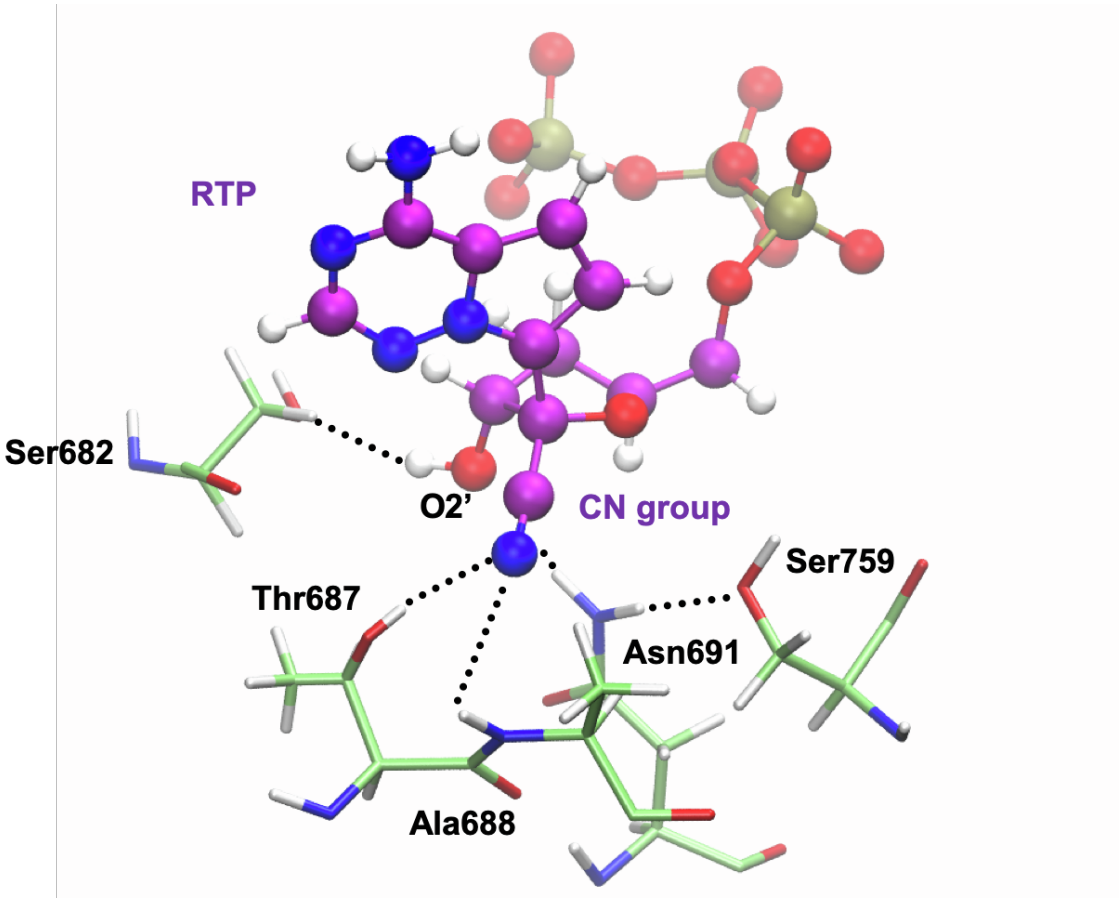

Figure S9. Interactions that are formed between the nitrile group of remdesivir-TP (RTP) and the active site of RdRp of SARS-CoV-2 during our simulations. remdesivir C atoms are shown in purple. Most important interactions are depicted as dotted lines. These interactions place remdesivir nitrile group towards the cavity and in a slightly different arrangement than UTP. Same pattern of recognition of the O2' group of RTP is achieved through the Ser_{682} residue.

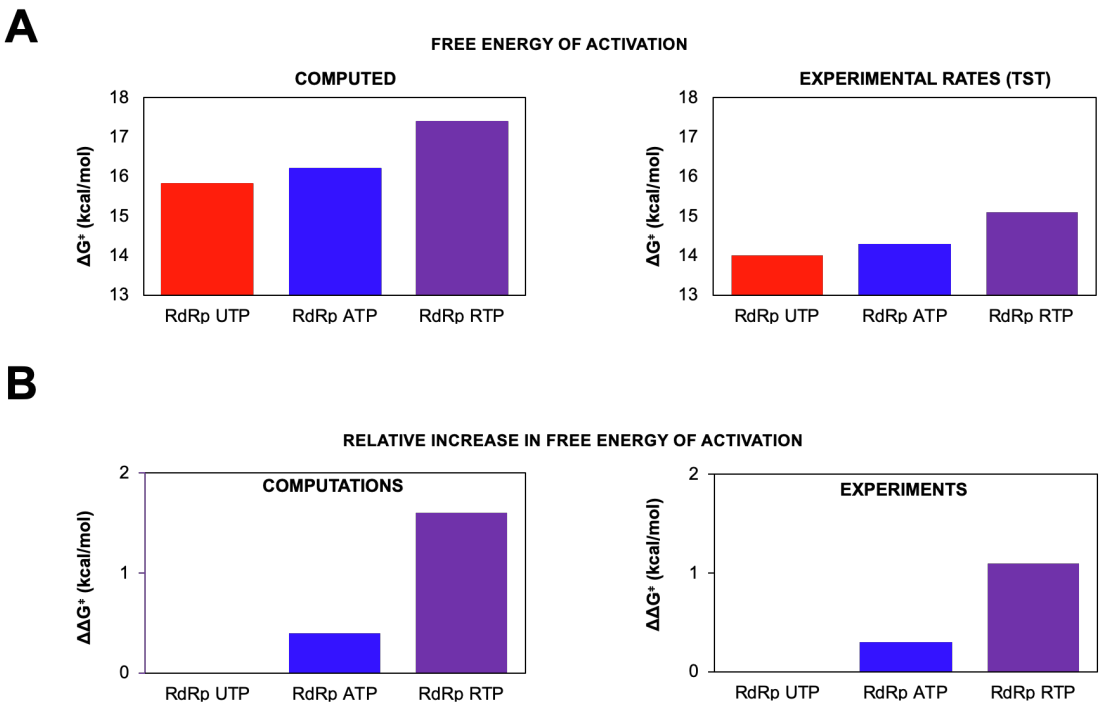

Figure S10. **A** Free energy of activation of UTP/ATP/RTP incorporation inside RdRp calculated (left panel) through QM/MM simulations in this study, and obtained through accurate pre-steady-state kinetic experiments[1] (right panel). **B** Increase in free energy of activation of incorporating an ATP or RTP inside RdRp, relative to UTP incorporation. Relative increase derived from computation and experiments are shown in left and right panels respectively. Rates were transformed to free energy of activation following Transition State Theory (TST) at room temperature.

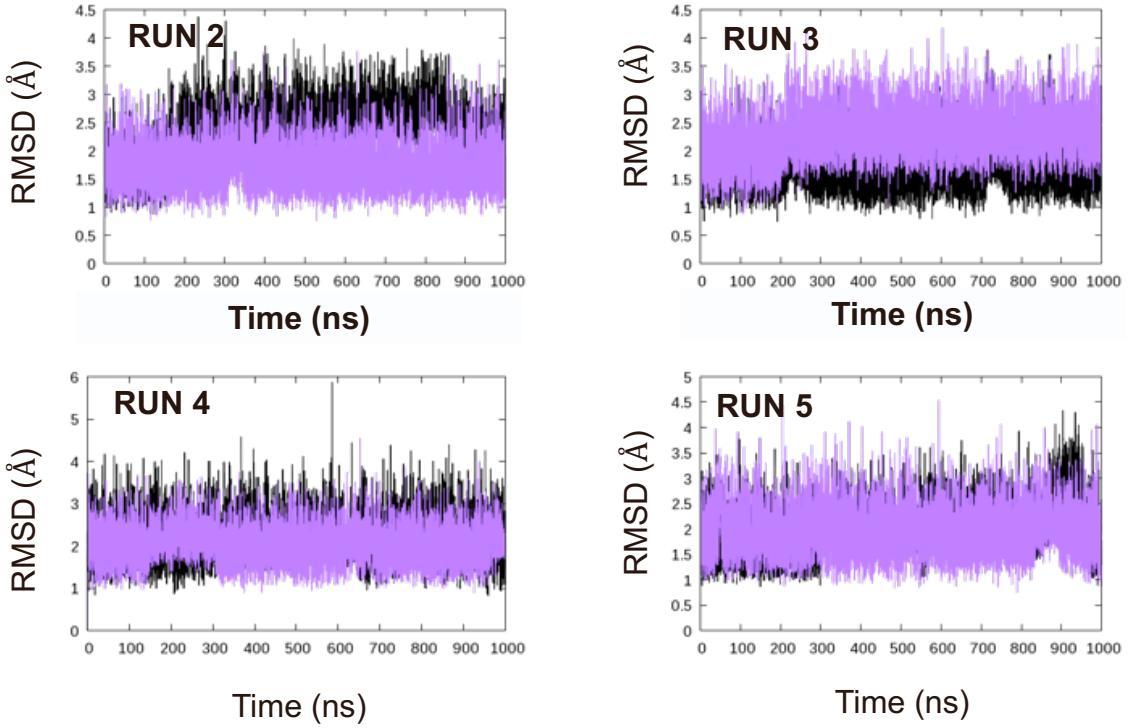

Figure S11. RMSD values in Å for the control sequence (in black) and the remdesivir-containing sequence (in purple) during 1 μs long MD simulations. 5 runs for each sequence were conducted including the run shown in Figure 5.

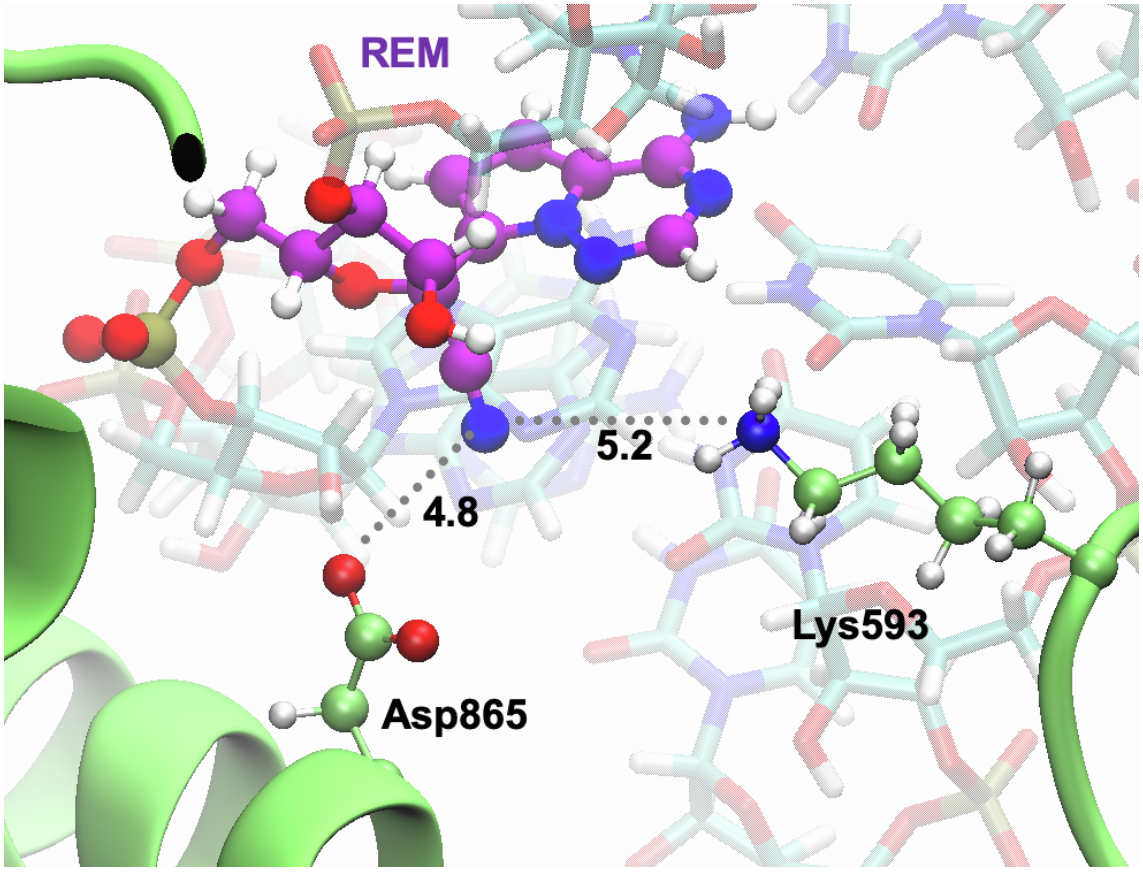

Figure S12. Closest residues of RdRp to remdesivir after two nucleotides have been incorporated. Residues are far to interact with any polar atom of remdesivir. Average distances during MD simulation between Asp_{865} and Lys_{593} sidechains to N atom of nitrile group of remdesivir are shown in Å.

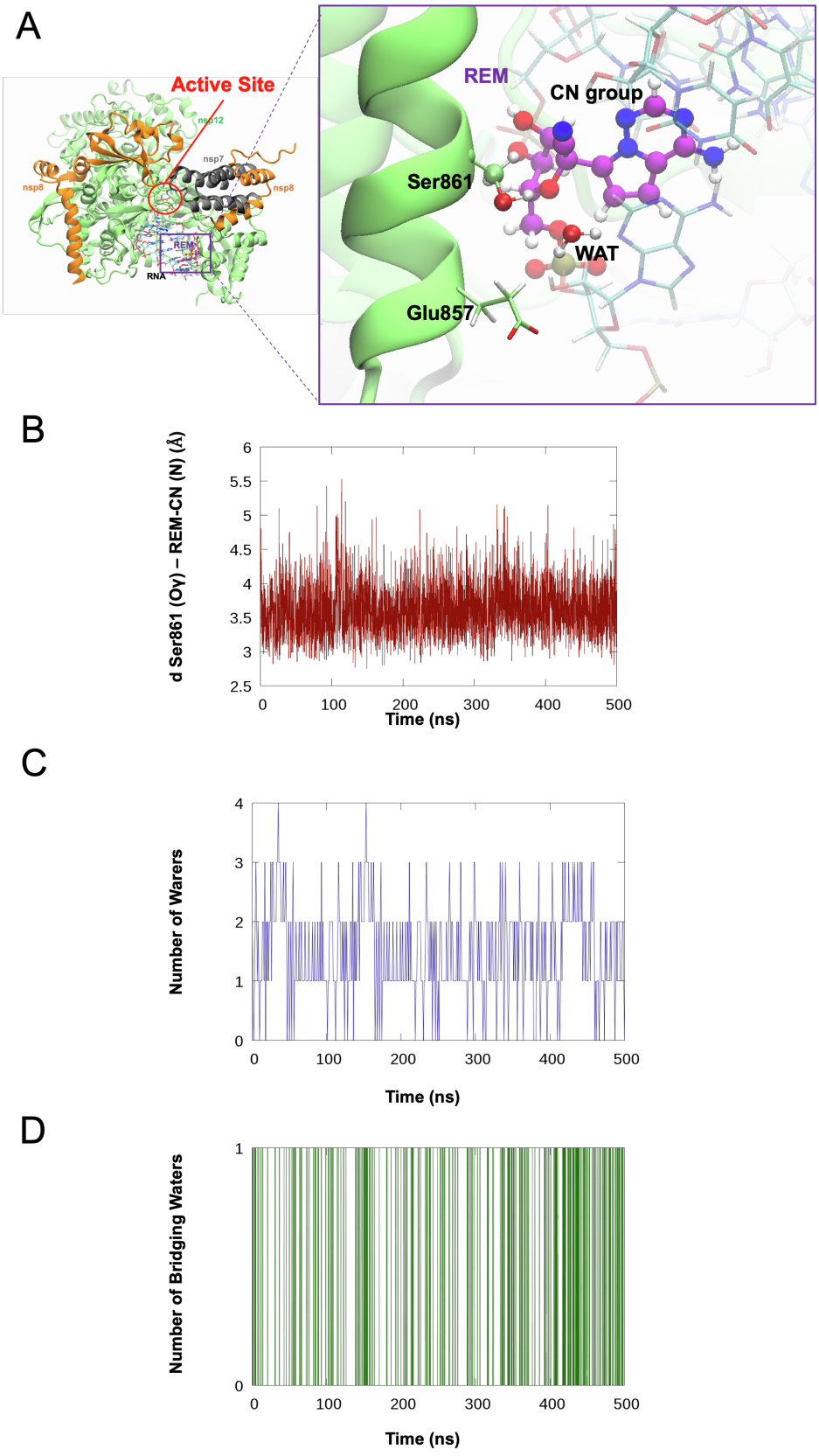

Figure S13. A Insight of the surroundings when remdesivir CN group is close to Ser_{861}. **B** Ser_{861} is found close to the nitrile group along 500 ns of MD simulation. **C** Number of waters within 3.5 Å distance from nitrile group. **D** Number of waters hydrogen bonded at the same time with the N atom of nitrile group of remdesivir and the sidechain Oγ atom of Ser_{861}.

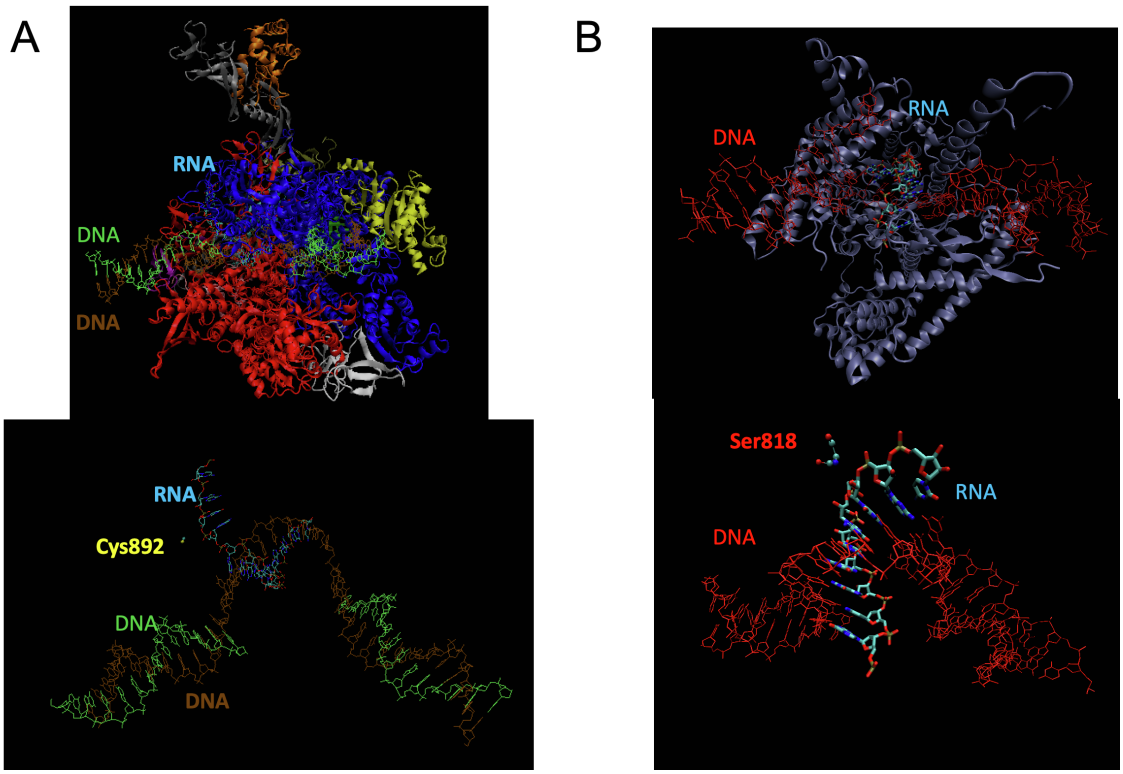

Figure S14. A Human RNA polymerase II-DNA/RNA complex (PDB ID: 5FLM[2]). Cys_{861} which is the closest Cys or Ser residue to a C1' atom of the nascent RNA strand, is placed to more than 8 Å distance. **B** Human mitochondrial RNA polymerase in complex with DNA and RNA, PDB ID: 4BOC.[3] Ser_{818}, the closest Cys or Ser residue to a C1' atom of the nascent RNA strand is placed at 6.3 Å distance.

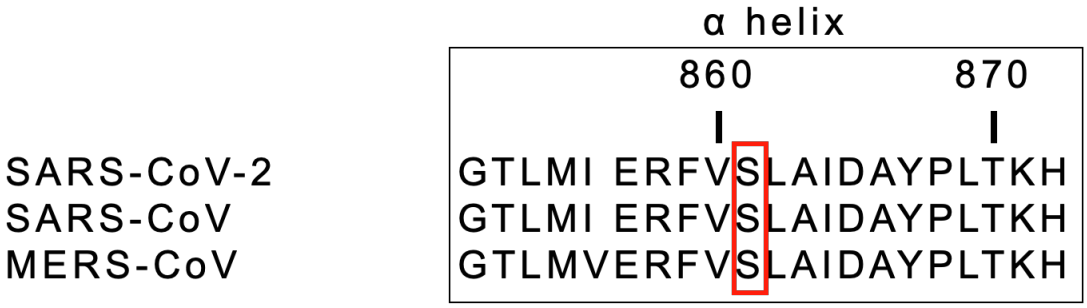

Figure S15. Alignment of SARS-CoV-2, SARS-CoV and MERS-CoV sequence for the alpha helix where Ser_{861} is placed.

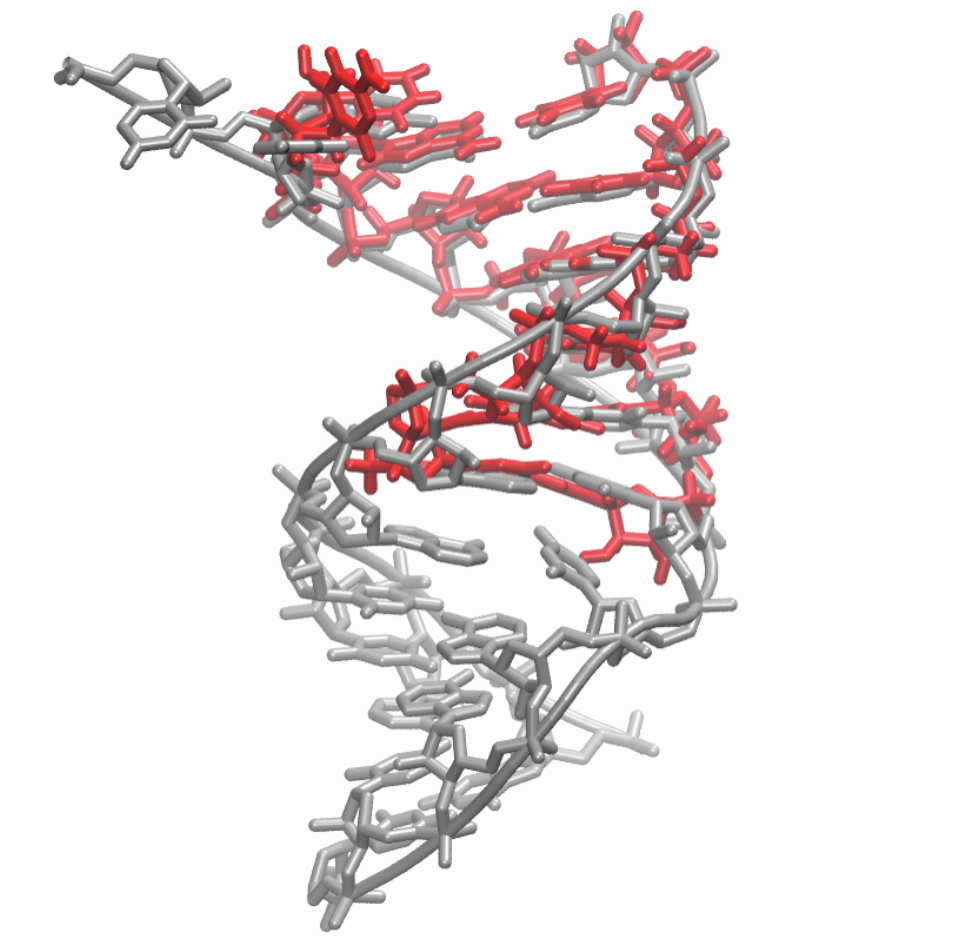

Figure S16. Alignment of RNA from cryo-EM structure with PDB ID:7bv2,[4] displayed in gray, and the double stranded RNA optimized and employed in our simulations and extracted from Hepatitis C virus, PDB ID: 4wtg,[5] displayed in red. RMSD between backbone atoms of the nucleic acids is 1.1 Å.

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

SARS-CoV-2 RdRp and human RNA Pol II Systems set up

Our starting point was the cryo-EM structure of the of the SARS-CoV-2 RdRp in complex with its cofactors.[6] as well as and the cryo-EM structure complexed with an RNA duplex.[4] The former structure was resolved without the Mg^{2+} cations needed for the catalysis and without the RNA template and nascent strand, we aligned it with the Hepatitis-C virus X-ray structure[5] which was crystallized with two Mn^{2+} cations, a nucleotide analog diphosphate molecule and a RNA template strand and a nascent RNA strand. We selected this structure as it showed the best alignment for both the cleft where RNA binds and the active site. In addition, it was resolved with two catalytic divalent cations and a diphosphate nucleotide analogue, which enabled us to build our selected substrates based on X-ray positions. Alignment was performed making use of the Pymol program, selecting a set of atoms that consisted on the atoms in the catalytic domains, and the residues placed in the cleft which are in charge of the RNA binding of both RdRp's. Thus, we then used the RNA molecule as well as the two cations and the diphosphate nucleotide molecule in our SARS-CoV-2 RdRp systems. The two cations were modeled as Mg^{2+} cations, and the nucleotide diphosphate was used as a template to build the UTP, ATP or remdesivir-TP molecules. In the case of the cryo-EM structure complexed with the RNA duplex[4] we overlapped its active site with aforementioned modeled active site. As in this structure the nucleotide has already been incorporated to the new RNA strand although it has not translocated we rebuilt the NTPs molecules based on the other model and the location of the PPi molecule present in the active site. Afterwards the systems were protonated making use of the LEAP module of the AMBER program.[7] The systems were solvated with LEAP module into a truncated octahedron box of TIP3P water molecules with a buffer of water molecules extending for 12 Å in every direction around the systems. Systems were neutralized by adding K+ ions. For the magnesium ions the parameters developed by Allner et al. were employed.[8] Proteins were described with ff14SB[9] AMBER ff. The RNA was simulated by combining ff99, the PARMBSC0 modifications and the chiOL3 modifications for RNA.[10–13] Charges

and parameters for the non-standard residues were derived to be compatible with the employed AMBER force field making use of the RED server.[14] Specifically, we derived parameters for a 3'-terminal uridine nucleotide deprotonated at its O3', a 3'-terminal remdesivir nucleotide deprotonated at its O3', a Uridine-TP, a remdesivir-TP and a remdesivir nucleotide. Same parameters and force field settings were employed in the simulation of human RNA Pol II and SARS-CoV RdRp complex.

Molecular Dynamic Simulations

Classical MD simulations were carried out using the AMBER 18 program[7] with a time step of 2 fs and applying the SHAKE algorithm[15] to bond lengths involving hydrogen atoms. Simulations were carried out in the isothermal-isobaric ensemble with a pressure of 1 atm and a temperature of 298 K. The Berendsen algorithm[16] was applied to control the pressure and the temperature with a coupling constant of 5 ps. The Particle Mesh Ewald method[17] was used to compute long-range electrostatic interactions using standard defaults and a cutoff in the real-space of 10 Å. The systems were energy minimized, thermalized and pre-equilibrated for 100 ns before the production run was conducted. During this multi-step approach, we firstly equilibrated the water box and counterions, then released the side-chains of the protein residues and then released the nucleobases gradually by maintaining its backbone frozen. Afterwards we released the whole protein atoms by maintaining the active site residues (UTP, ATP or RTP, the Mg^{2+} cations and their coordination spheres) and the nucleic acid backbone frozen. We then performed an MD run imposing a restraint to the distance between the 3'-hydroxyl oxygen atom of the terminal nucleotide and the α-phosphate atom of UTP, ATP or RTP. Finally, we slowly released the positional restraints imposed to the system and the distance restraint. A total time of 500 ns of fully unrestrained MD simulations were performed for all the systems. Thus, we performed MD simulation in systems containing RdRp with its cofactors, RNA, two MG^{2+} cations and: a ATP molecule, a UTP molecule, a RTP molecule, a UTP and a remdesivir nucleotide placed in position i+1 to i+4. In the case of SARS-

CoV RdRp the system consisted in RdRp complex (nsp12, nsp7/8 cofactors), RNA duplex, two MG^{2+} cations and an ATP molecule. RMSD for the different systems during our MD simulations are shown below. Same equilibration protocol and settings were employed in the simulation of the human RNA Pol II and SARS-CoV RdRp complex system.

Cation binding and Classical Molecular Interaction Potential calculations

We used the classical molecular interaction potential (CMIP)[18] to explore the more probable regions where Mg^{2+} ions could bind. AMBER Lennard-Jones parameters were used to determine the van der Waals contribution, using the ones provided by Allner for Mg^{2+}, and the Poisson-Boltzmann equation[19] was used to determine the electrostatic interaction term. The ionic strength was set to 0.15 M, the dielectric constant of the reaction-field was 78.14 M. This calculation allowed us to predict the binding of the two Mg^{2+} ions as follows.

MD simulations and CMIP calculations unveiled the detailed mode of binding of the two Mg^{2+} inside SARS-CoV-2 RdRp active site. During classical MD simulations equilibration and optimization stages, when a NTP (ATP/UTP) molecule and two Mg^{2+} ions are included in the SARS-CoV-2 RdRp-RNA duplex model system, the loop containing Asp618 and Tyr619 residues approach the loop where Asp760 and Asp761 are located to form a well defined high interaction energy region (see Fig 1C, Fig S2 A). These regions correspond to two Mg^{2+} coordination spheres (see Fig 1C, Fig S2 B). Nevertheless, when no ATP molecule it is included in the model (now consisting on RdRp-RNA duplex) only a high interaction energy region corresponding to the binding of one Mg^{2+} ion is found (see Fig S2 C). It is accepted that the NTP molecule enters polymerases active site carrying a Mg^{2+} ion[20] what may account for our observations. Moreover, when CMIP calculations are performed in RdRp alone no specific nor high interaction energy region is found (see Fig S2 D). Thus, in SARS-CoV-2 RdRp the entry of an NTP-MG entity triggers the slight movement (~2 Å, see Fig S2) of the Asp618 and Tyr619 loop and creates two well defined coordination spheres to form a catalytically active conformation of the RdRp.

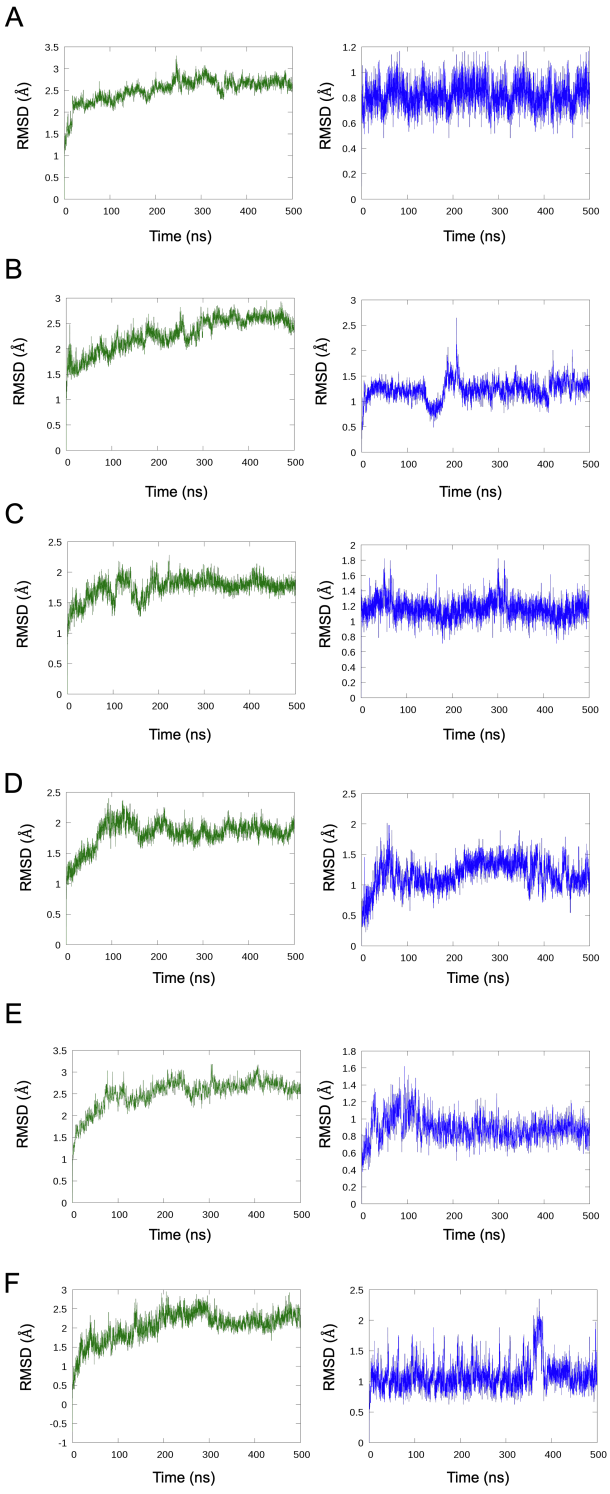

Figure S17. RMSD for the protein (green) and nucleic acid (blue) backbone atoms during MD simulations. **A** RMSDs for the system containing a ATP molecule. **B** RMSDs for the system containing a UTP molecule. **C** RMSD for the system containing an RTP molecule. **D** RMSD for the system containing a remdesivir incorporated to the nascent RNA strand. **E** RMSD for the system incorporating one more nucleotide after Remedesivir. **F** RMSD for the system incorporating two more nucleotides after Remedesivir. **G** RMSD for the system incorporating three more nucleotides after Remedesivir.

Structural analysis of remdesivir inside an RNA double helix

We performed MD simulations of 1 μs long in two double-stranded RNA dodecamers. One was used as a control a contained only natural-occurring nucleotides with the sequence r(CGCGAAUUGCGC)·r(GCGCAAUUCGCG) while in the other a remdesivir was placed in the central position r(CGCGARUUGCGC)·r(GCGCAAUUCGCG). The double-stranded RNA molecules where built making use of AMBER Nucleic Acid Builder module. Two A-RNA molecules were constructed. MD simulations were conducted with same protocol and parameters as the ones already explained in the previous section. RNA helical base-pair step parameters were calculated by making use of CURVES+ and CANAL programs.[21]

QM/MM Calculations

We selected snapshots of the last 50ns as our starting point to build our QM/MM models. The AMBER program making use of the interface with Terachem 1.9 program[22,23] or Gaussian16 program[24] were used. All calculations were performed with electrostatic embedding. For the ligation reaction, the QM subsystem consisted on the UTP, ATP or RTP molecule, the terminal nucleotide's sugar ring without the nucleobase, two magnesium ions, and both the side-chain of the protein residues and the waters involved in its coordination sphere (see Fig. S4 A). The total number of QM atoms were 117 including the link atoms when a UTP molecule was studied, 121 for ATP, and 122 when a RTP was present.

We used the link atoms procedure as implemented in the AMBER program to saturate the valence of the frontier between the QM and the MM subsystems. After the system was built the system was re-equilibrated at the QM/MM level by performing minimizations and a 10 ps long NPT QM/MM-MD simulation using periodic boundary conditions with an electrostatic cutoff of 12 Å for the QM/MM electrostatic interactions.

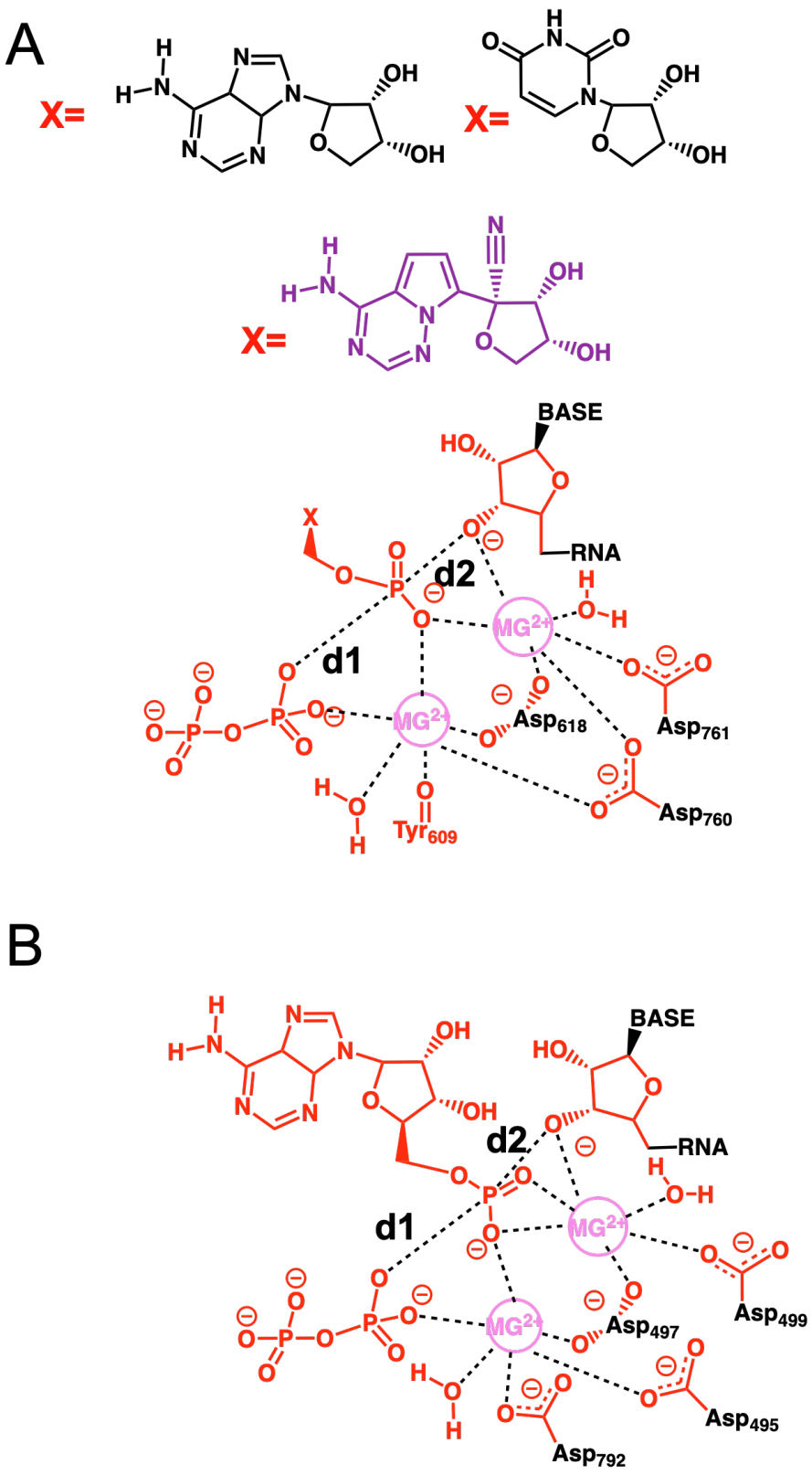

Figure S18. A Atoms described at QM level (in red and pink) in the hybrid QM/MM calculations during the ligation reaction step. Distances involved in the Reaction Coordinates employed are shown. **B** Atoms described at QM level (in red) in the hybrid QM/MM calculations for the Human RNA Pol II incorporation reaction. Distances involved in the Reaction Coordinates employed are shown.

Exploration of the Minimum Free Energy Paths and Potential of Mean Force

By means of the string method[25] we investigated the preferred minimum free energy paths (MFEP) by performing QM/MM-MD simulations. We selected snapshots of the last 50ns of the MD simulations as our starting point to build our QM/MM models. The QM subsystems are shown in Fig. S3 A for RdRp and Fig. S3 B for human RNA Pol II, and atoms were described at the DFTB3[26,27]/MM level, with corrections at B3lyp/6-311++G** to the electronic energy. In Fig. S3 A B is depicted the active space consisting on 2 (d1 and d2 in Fig. S3 A) distances that were selected to trace the MFEPs. Afterwards a collective variable was defined along the path[28,29] for a given reaction mechanism and was used to obtain the potential of mean force (PMF) using the umbrella sampling technique.[30] Each MFEP was computed by using 60 to 120 string nodes for the phosphoryl transfer reaction inside SARS-Cov-2 RdRp and human RNA Pol II. During the adaptive string optimization the positions and force constants of umbrella sampling windows were taken from the adjusted node parameters.[31] A time step of 1 fs was employed in all cases. Temperature was set to 298K. For the determination of MFEPs the averaged positions of the string nodes were determined in the last 20 ps after the string had converged. Different initial guesses were employed to explore all possible reaction mechanisms. Afterwards 120 points were interpolated for each MFEP between the converged string nodes. These points were used to define the collective variable (s)[28,29] which measures the advance of the system along the MFEP. Umbrella sampling windows were simulated during 20 ps for a relaxation run and during 200 ps during the production run. The time step employed was the same used in the calculation of the MFEP. The statistical uncertainties were calculated as 95% confidence intervals and reached error values within ± 1 kcal·mol^{-1} in the whole free energy profile. This was checked for the whole profiles and for each reaction mechanism studied. Finally, interpolated corrections[28,29] were made to the DFTB3/MM at the high level B3lyp/6-311++G**/MM in the following way. From the structures collected during the PMF production we performed minimizations in each of the nodes of the MFEP for 1000 minimization steps. Then, single point energy calculations at both the B3lyp/6-311++G**/MM and DFTB3/MM were performed. Finally, the corrections were applied as follows:

$$E = E_{QM}^{LL} + E_{QM/MM}^{LL} + E_{MM} + \mathrm{Spl}[\Delta E_{LL}^{HL}(s)]$$

where Spl is a one-dimensional cubic spline function and its argument, ΔE_{LL}^{HL} , is the correction term obtained as the difference between the single-point high-level (HL) energy of the QM system and the low level one (LL).

REFERENCES

1. Dangerfield, T.L., Huang, N.Z., and Johnson, K.A. (2020). Remdesivir Is Effective in Combating COVID-19 because It Is a Better Substrate than ATP for the Viral RNA-Dependent RNA Polymerase. iScience *23*, 101849.
2. Bernecky, C., Herzog, F., Baumeister, W., Plitzko, J.M., and Cramer, P. (2016). Structure of transcribing mammalian RNA polymerase II. Nature.
3. Schwinghammer, K., Cheung, A.C.M., Morozov, Y.I., Agaronyan, K., Temiakov, D., and Cramer, P. (2013). Structure of human mitochondrial RNA polymerase elongation complex. Nat. Struct. Mol. Biol.
4. Yin, W., Mao, C., Luan, X., Shen, D.D., Shen, Q., Su, H., Wang, X., Zhou, F., Zhao, W., Gao, M., et al. (2020). Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir. Science (80-.). *368*, 1499–1504.
5. Appleby, T.C., Perry, J.K., Murakami, E., Barauskas, O., Feng, J., Cho, A., Fox, D., Wetmore, D.R., McGrath, M.E., Ray, A.S., et al. (2015). Structural basis for RNA replication by the hepatitis C virus polymerase. Science (80-.). *347*, 771–775.
6. Gao, Y., Yan, L., Huang, Y., Liu, F., Zhao, Y., Cao, L., Wang, T., Sun, Q., Ming, Z., Zhang, L., et al. (2020). Structure of the RNA-dependent RNA polymerase from COVID-19 virus. Science (80-.). *368*, 779–782.
7. Case, D.A., Babin, V., Berryman, J.T., Betz, R.M., Cai, Q., Cerutti, D.S., Cheatham, T.E., Darden, T.A., Duke, R.E., Gohlke, H., et al. (2019). {Amber 19} OR - University of California, San Francisco.
8. Allnér, O., Nilsson, L., and Villa, A. (2012). Magnesium ion-water coordination and exchange in biomolecular simulations. J. Chem. Theory Comput. *8*, 1493–1502.
9. Maier, J.A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K.E., and Simmerling, C. (2015). ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. J. Chem. Theory Comput. *11*, 3696–3713.
10. Cheatham, T.E., Cieplak, P., and Kollman, P.A. (1999). A Modified Version of the Cornell *et al.* Force Field with Improved Sugar Pucker Phases and Helical Repeat. J. Biomol. Struct. Dyn. *16*, 845–862.
11. Wang, J., Cieplak, P., and Kollman, P.A. (2000). How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? J. Comput. Chem. *21*, 1049–1074.
12. Pérez, A., Marchán, I., Svozil, D., Sponer, J., Cheatham, T.E., Laughton, C.A., and Orozco, M. (2007). Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of α/γ Conformers. Biophys. J. *92*, 3817–3829.
13. Zgarbová, M., Otyepka, M., Šponer, J., Mládek, A., Banáš, P., Cheatham, T.E., and Jurečka, P. (2011). Refinement of the Cornell et al. Nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. J. Chem. Theory Comput. *7*, 2886–2902.
14. Vanquelef, E., Simon, S., Marquant, G., Garcia, E., Klimerak, G., Delepine, J.C., Cieplak, P., and Dupradeau, F.Y. (2011). R.E.D. Server: A web service for deriving RESP and ESP charges and building force field

libraries for new molecules and molecular fragments. Nucleic Acids Res. *39*.
15. Ryckaert, J.P., Ciccotti, G., and Berendsen, H.J.C. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J. Comput. Phys. *23*, 327–341.
16. Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A., and Haak, J.R. (1984). Molecular dynamics with coupling to an external bath. J. Chem. Phys. *81*, 3684–3690.
17. Darden, T., York, D., and Pedersen, L. (1993). Particle mesh Ewald: An N [center-dot] log(N) method for Ewald sums in large systems. J. Chem. Phys. *98*, 10089–10092.
18. Gelpí, J.L., Kalko, S.G., Barril, X., Cirera, J., De La Cruz, X., Luque, F.J., and Orozco, M. (2001). Classical molecular interaction potentials: Improved setup procedure in molecular dynamics simulations of proteins. Proteins Struct. Funct. Genet. *45*, 428–437.
19. Orozco, M., and Luque, F.J. (2001). Theoretical Methods for the Description of the Solvent Effect in Biomolecular Systems. (Chem. Rev. 2000, 100, 4187−4226. Published on the Web Oct 21, 2000.). Chem. Rev. *101*, 203–204.
20. Wu, S., Li, L., and Li, Q. (2017). Mechanism of NTP Binding to the Active Site of T7 RNA Polymerase Revealed by Free-Energy Simulation. Biophys. J. *112*, 2253–2260.
21. Lavery, R., Moakher, M., Maddocks, J.H., Petkeviciute, D., and Zakrzewska, K. (2009). Conformational analysis of nucleic acids revisited: Curves+. Nucleic Acids Res.
22. Ufimtsev, I.S., and Martinez, T.J. (2009). Quantum chemistry on graphical processing units. 3. Analytical energy gradients, geometry optimization, and first principles molecular dynamics. J. Chem. Theory Comput.
23. Götz, A.W., Clark, M.A., and Walker, R.C. (2014). An extensible interface for QM/MM molecular dynamics simulations with AMBER. J. Comput. Chem. *35*, 95–108.
24. Frisch G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratch, D. J., M.J.. T. (2016). Gaussian 16, Rev. B.01. Gaussian, Inc., Wallingford, CT.
25. Vanden-Eijnden, E., and Venturoli, M. (2009). Revisiting the finite temperature string method for the calculation of reaction tubes and free energies. J. Chem. Phys. *130*.
26. Gaus, M., Cui, Q., and Elstner, M. (2011). DFTB3: Extension of the self-consistent-charge density-functional tight-binding method (SCC-DFTB). J. Chem. Theory Comput. *7*, 931–948.
27. Gaus, M., Lu, X., Elstner, M., and Cui, Q. (2014). Parameterization of DFTB3/3OB for sulfur and phosphorus for chemical and biological applications. J. Chem. Theory Comput. *10*, 1518–1537.
28. Zinovjev, K., Martí, S., and Tuñón, I. (2012). A collective coordinate to obtain free energy profiles for complex reactions in condensed phases. J. Chem. Theory Comput.
29. Zinovjev, K., Ruiz-Pernía, J.J., and Tuñón, I. (2013). Toward an automatic

determination of enzymatic reaction mechanisms and their activation free energies. J. Chem. Theory Comput. *9*, 3740–3749.
30. Torrie, G.M., and Valleau, J.P. (1977). Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. J. Comput. Phys. *23*, 187–199.
31. Zinovjev, K., and Tuñón, I. (2017). Adaptive Finite Temperature String Method in Collective Variables. J. Phys. Chem. A *121*, 9764–9772.