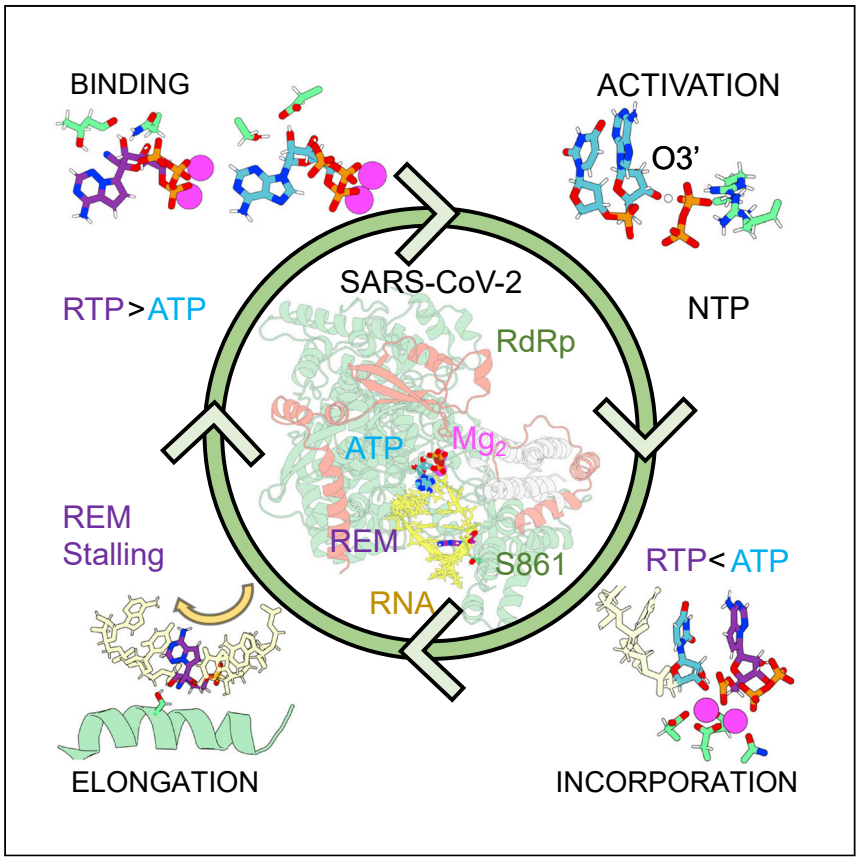
Chem Catalysis

CellPress
OPEN ACCESS

Article

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



Aranda et al. reveal the binding, activation, incorporation, and elongation of natural and remdesivir nucleotides inside SARS-CoV-2 RNA-dependent RNA polymerase. Reaction mechanisms are characterized at atomic resolutions and compared with human RNA polymerase II. During RNA polymerization, remdesivir is stalled in a stabilizing trap that prevents further translocation.

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

juan.aranda@irbbarcelona.org (J.A.)
modesto.orozco@irbbarcelona.org (M.O.)

Highlights

SARS-CoV-2 RdRp and human RNA Pol II display opposite ATP and RTP binding preferences

Viral polymerase makes use of a self-activated mechanism

RTP is incorporated slower that its natural counterpart, ATP, inside RdRp

During RNA polymerization, remdesivir is stalled in a stabilizing trap

Aranda et al., Chem Catalysis 2, 1084–1099
May 19, 2022
https://doi.org/10.1016/j.checat.2022.03.019

CellPress
OPEN ACCESS

Chem Catalysis

Article

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

Juan Aranda,[1,*] Milosz Wieczór,[1,2] Montserrat Terrazas,[1,3] Isabelle Brun-Heath,[1] and Modesto Orozco[1,4,5,*]

SUMMARY

We combine molecular dynamics, statistical mechanics, and hybrid quantum mechanics/molecular mechanics simulations to describe mechanistically the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) RNA-dependent RNA polymerase (RdRp). Our study analyzes the binding mode of both natural triphosphate substrates as well as remdesivir triphosphate (the active form of drug), which is bound preferentially over ATP by RdRp while being poorly recognized by human RNA polymerase II (RNA Pol II). A comparison of incorporation rates between natural and antiviral nucleotides shows that remdesivir is incorporated more slowly into the nascent RNA compared with ATP, leading to an RNA duplex that is structurally very similar to an unmodified one, arguing against the hypothesis that remdesivir is a competitive inhibitor of ATP. We characterize the entire mechanism of reaction, finding that viral RdRp is highly processive and displays a higher catalytic rate of incorporation than human RNA Pol II. Overall, our study provides the first detailed explanation of the replication mechanism of RdRp.

INTRODUCTION

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the virus that caused coronavirus disease 2019 (COVID-19), emerged in China in December 2019 and rapidly spread over the world, causing a worldwide health threat with more than 5 million fatalities.[1] From a phylogenetic point of view, SARS-CoV-2 belongs to the β genus of the coronavirus family, which includes other highly infective pathogens such as the Middle East respiratory syndrome CoV (MERS-CoV) or the severe acute respiratory syndrome CoV (SARS-CoV).[2] SARS-CoV-2 has a large (30 kb of positive RNA) genome, which forces it to strike a balance between high replication fidelity and genetic diversity.[3–6] A highly processive RNA-dependent RNA polymerase (RdRp) and a proof-reading exonuclease are the crucial elements for maintaining the stability of the viral genome, while at the same time enabling its mutation to adapt to new environments.

RdRp is the core of the replication machinery of the virus and one of the largest proteins in the viral genome (932 residues). It binds to nsp7 and nsp8 to form an active complex,[7–9] one that first uses sense RNA as a template to generate a negative copy and then, in a second cycle, generates new copies of genomic and sub-genomic RNAs. At least two other proteins are involved in the replication process: a 601-residue helicase and a 527-residue proof-reading exonuclease.[10,11] A simple BLAST[12] query shows that SARS-CoV-2 RdRp is highly conserved within the coronavirus family, but homologs out of this family show a quite low identity (less than 25%), which suggests that we are faced with a quite new protein, but with a surprisingly high functional efficiency.

The bigger picture

RNA-dependent RNA polymerase (RdRp) from severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is an attractive target to attack the viral replication. Many efforts have been directed toward the design of effective inhibitors of RdRp. However, the mechanism of nucleoside-triphosphate binding, as well as nucleotide activation and incorporation at atomic resolution, has not been deciphered. Also, the molecular mechanism of the promising antiviral remdesivir, a nucleotide analog, is still under study. In this work, binding preferences between natural and remdesivir-TP molecules have been analyzed inside RdRp and human RNA polymerase II active sites. Also, the detailed reaction mechanism of nucleotide activation and incorporation inside the RdRp are characterized. Afterward, during RNA polymerization, remdesivir is stalled in a stabilizing trap. The characterization of the replication mechanism of SARS-CoV-2 RdRp provided by this study can help guide the design of next-generation antivirals.

Due to its central role in the viral infection cycle, RNA polymerases are a major target for fighting RNA viruses.[13–15] Although vaccines began to be available as of November 2021, 50% of world's population still remains without any dose of it, and the use of effective antivirals is needed to prevent future fast-spreading coronavirus outbreaks due to ineffective immunizations or the emergence of new variants. As of today, the only FDA-approved drug for the treatment of SARS-CoV-2 infection[16] is a RdRp-inhibitor, remdesivir (R), a C-nucleoside (see Figure S1) that was approved for the treatment of Ebola.[17] Being a negative single-strand RNA virus, Ebola is very distant from SARS-CoV-2, but its replication is also dependent on the action of an RdRp.

Several SARS-CoV-2 RdRp cryoelectron microscopy (cryo-EM) structures have been reported since the beginning of the pandemic. The first structure comprised the RdRp complex.[8] Afterward, structures including different RNA duplexes were resolved showing that RNA binding causes no drastic rearrangements in RdRp.[9,18,19] W. Yin et al. were able to capture a structure where remdesivir has already been incorporated in a nascent RNA strand. Also, a pyrophosphate (PPi) molecule departing from the active site and two Mg^{2+} near it were found.[18] Recently, structures of RdRp in complex with favipiravir triphosphate (TP) provided new insights of the pre-catalytic state. However, these structures were resolved with the drug in a non-productive conformation[20] or without Mg^{2+} ions in the active site.[21] So far, however, the limited amount of atomistic data on the mode of binding and reaction mechanism of incorporation of both natural substrates and remdesivir hampers our ability to develop new and more active compounds.

We present here a comprehensive study on the mechanism of action of SARS-CoV-2 RdRp. Atomistic simulations characterize binding interactions and substrate preferences in the active site of the viral polymerase. While remdesivir binds to the viral active site more strongly than its natural counterpart, the opposite is found for human RNA polymerase II. Molecular dynamics (MD) and quantum mechanics/molecular mechanics (QM/MM) simulations demonstrate that the enzyme follows a canonical 2-ion reaction mechanism with a catalytic efficiency higher than that of the highly evolved human RNA polymerase II (RNA Pol II). Calculations and biochemical experiments demonstrate that remdesivir triphosphate (RTP; the expected bioactive form of remdesivir) can be recognized and incorporated into nascent RNA with an efficiency only slightly lower than natural nucleotides, i.e., remdesivir is not an inhibitor of nucleotide incorporation. Extended MD simulations failed to detect any dramatic distortion in the RNA duplex due to the presence of remdesivir that would account for its inhibitory properties. Furthermore, no steric clashes were detected when the nascent RNA duplex was displaced along the exit channel (see below), which argues against the hypothesis that steric clashes are responsible for delayed inhibition, suggesting a more specific inhibitory mechanism related perhaps to a transient covalent bond with the enzyme. In summary, our results, obtained through thorough computational simulations, provide the first atomistic description of the mechanism of action for SARS-CoV-2 RNA polymerase and provide clues on the mysterious inhibition mechanism of remdesivir.

RESULTS AND DISCUSSION

SARS-CoV-2 RdRp active site architecture

The viral protein (see Figure 1A) has a canonical active site, quite similar to those of other polymerases. The two essential Mg^{2+} are coordinated by the α and β phosphate groups of the incoming triphosphate nucleotide, as well as Asp_{618}, Asp_{760}, Asp_{761}, Tyr_{619}, and the O3′ terminal of the negative RNA strand (see Figures 1B

[1]Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Baldiri Reixac 10, 08028 Barcelona, Spain

[2]Department of Physical Chemistry, Gdansk University of Technology, Narutowicza 11/12, 80-233 Gdańsk, Poland

[3]Department of Inorganic and Organic Chemistry, Section of Organic Chemistry, IBUB, University of Barcelona, Martí i Franquès 1-11, 08028 Barcelona, Spain

[4]Departament de Bioquímica i Biomedicine, Universitat de Barcelona, Universitat de Barcelona, Avinguda Diagonal 645, 08028 Barcelona, Spain

[5]Lead contact

*Correspondence: juan.aranda@irbbarcelona.org (J.A.), modesto.orozco@irbbarcelona.org (M.O.)

https://doi.org/10.1016/j.checat.2022.03.019

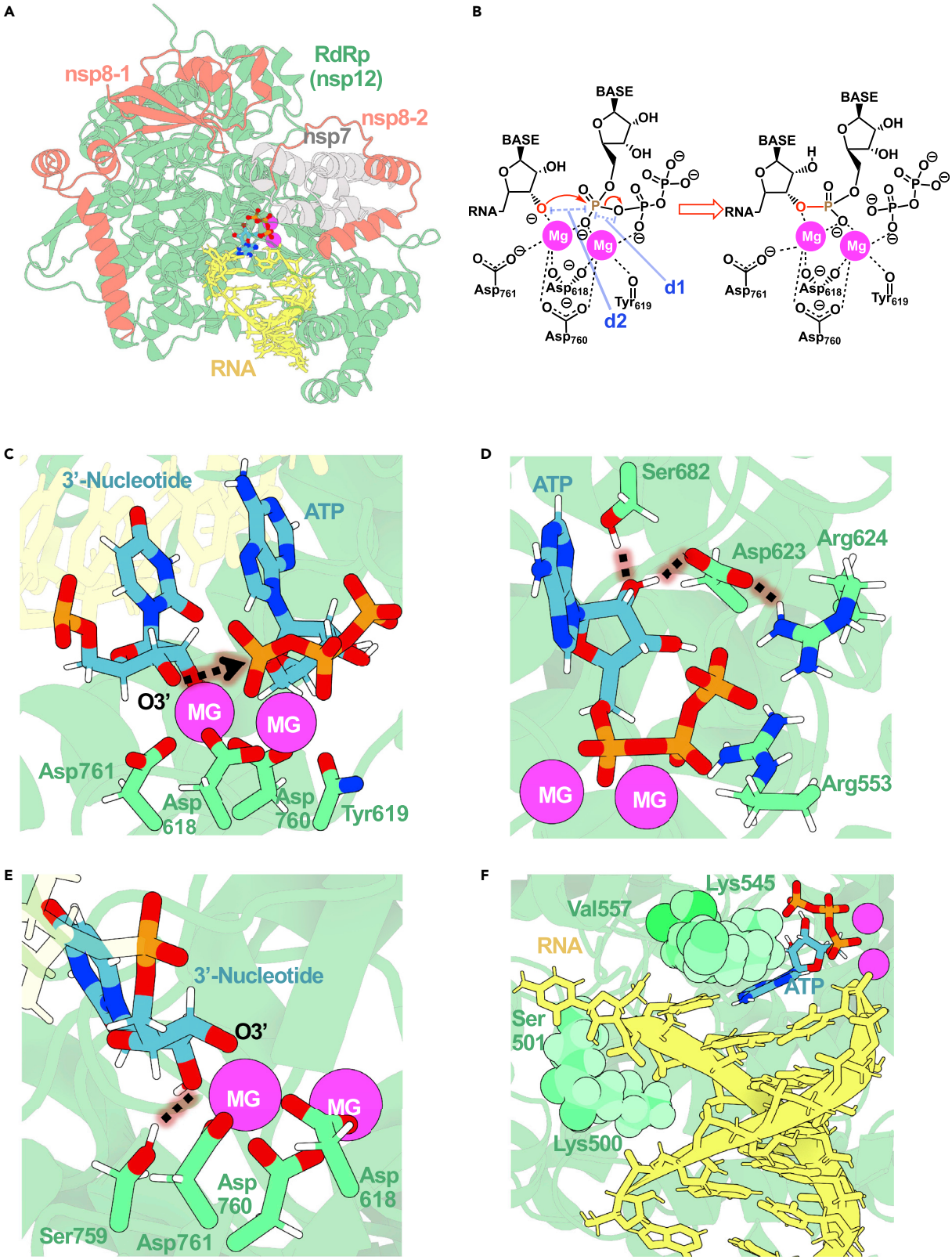

A
nsp8-1
RdRp
(nsp12)
nsp8-2
nsp7
RNA
B
BASE
OH
RNA
Mg
Asp761
Asp618
Asp760
Tyr619
d1
d2
C
3'-Nucleotide
ATP
O3'
MG
Asp761
Asp
618
Asp
760
Tyr619
D
ATP
Ser682
Asp623
Arg624
MG
Arg553
E
3'-Nucleotide
O3'
MG
Asp
760
Asp
618
Ser759
Asp761
F
Lys545
Val557
RNA
ATP
Ser
501
Lys500

Figure 1. Active site of SARS-CoV-2 RdRp makes it an efficient polymerase

(A) The replication complex of SARS-CoV-2 formed by the nsp12 RdRp enzyme (in green), the nsp8 and nsp7 cofactors (orange and gray, respectively), and an RNA template and nascent strands (yellow).

(B) Scheme depicting the two-metal-ion mechanism used by SARS-CoV-2 RdRp.

(C) The active site contains a well-defined coordination sphere of the two Mg^{2+} ions.

(D) NTP substrate recognition in the active site of RdRp is mediated by a pair of arginines, an aspartate, and a serine.

(E) The deprotonated 3′ terminal nucleotide is stabilized by one of the catalytic ions in the RdRp's active site.

(F) RdRp active site pocket enables base specificity between the incoming nucleotide and the template.

Figures were prepared with 3D Protein Imager.[24] A 3D structure representation can be accessed through https://mmb.irbbarcelona.org/3dRS/s/Y17cub.

and 1C), which—according to the circular reaction mechanism for polymerases by de Vivo and coworkers—is expected to be ionized.[22] Classical molecular-interaction-potential calculations,[23] using Mg^{2+} as the binding cation, further corroborate the two Mg^{2+} binding sites (see supplemental information and Figure S2). Compared with available structural data,[8,18] our catalytically competent active site unveils the slight movement of the Asp_{618}-Tyr_{619} loop (Figure S2) in order to create well-defined coordination spheres for the two Mg^{2+} and to be able to proceed with a phosphoryl transfer with high efficiency. Two arginine residues (Arg_{624} and Arg_{553}) bind the phosphates of the incoming nucleotide, aligning the gamma phosphate for an effective transfer (see Figure 1D). The preferential affinity for ribonucleotide triphosphate substrates (as opposed to deoxyribonucleotide ones) can be explained by the need for north puckering that controls the alignment of the reactive groups, as well as the presence of specific H bonds between the 2′OH group of the nucleoside triphosphate (NTP) and the side chains of Asp_{623} and Ser_{683} in the catalytic site (see Figure 1D). Additional hydrogen bonds are found between i+1 2′OH and Ser_{759} (see Figure 1E). The base specificity is controlled by the complementarity of hydrogen bonding with the template nucleobase (see Figures 1F and S3) and by phosphate coordination, as well as by the residues surrounding the active site that mechanically introduce strong isosteric requirements, altogether making a non-Watson-Crick pairing scheme very unlikely (see Figure 1F). Overall, the structural picture of the active site emerging from EM structures[8,9,18,19] and atomistic simulations strongly suggests that despite a short evolutionary history, SARS-CoV-2 RdRp has all the structural requirements to be an efficient RNA polymerase both in terms of catalysis rate and substrate specificity.

We also analyzed the SARS-CoV RdRp active site through MD simulations (see experimental procedures and supplemental experimental procedures). SARS-CoV and SARS-CoV-2 RdRp (nsp12) share more than a 96% sequence similarity, and their structures' root-mean-square deviation (RMSD) is 0.84 Å.[8] In addition, the loops containing the catalytic residues Asp_{618}, Asp_{760}, Asp_{761}, and Tyr_{619}, involved in the coordination of the Mg^{2+}s, are identical in sequence. During our simulations, SARS-CoV RdRp and SARS-CoV-2 RdRp Mg^{2+} coordination spheres remained almost identical (see Figure S4). The average distances between magnesium ions were found to be 3.9 $\pm$ 0.2 and 3.7 $\pm$ 0.1 Å for SARS-CoV and SARS-CoV-2, respectively. Also, the distance involved in nucleotide incorporation between Pα an O3′ atoms remained very similar at 3.6 $\pm$ 0.2 and 3.5 $\pm$ 0.1 Å, respectively. Thus, the observed difference in activity[7] between them should be attributed to other factors.

SARS-CoV-2 RdRp and human RNA Pol II display opposite binding preferences with respect to RTP and ATP

To directly compare the affinity of the binding sites of SARS-CoV-2 RdRp and human RNA Pol II, we performed alchemical free energy simulations, transforming ATP into RTP either in the binding pocket or in free solution. The obtained free-energy differences were directly translated into affinity differences (ΔΔG) using standard

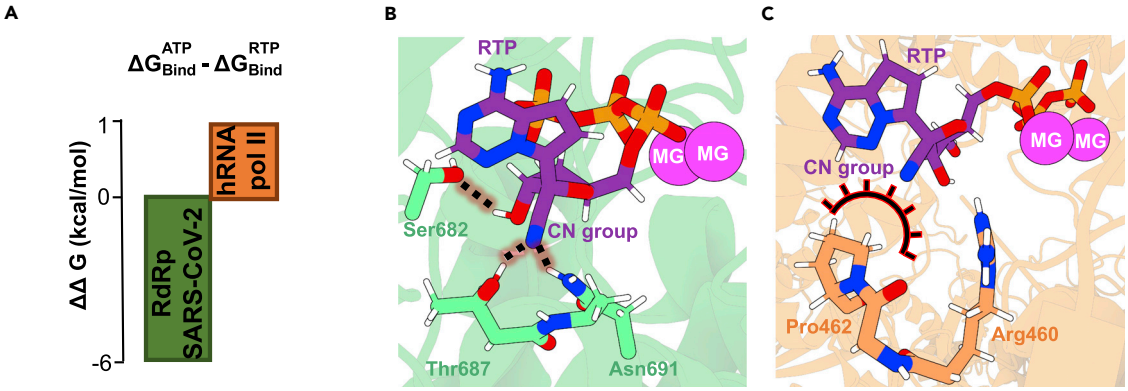


Figure 2. Binding preferences in viral and human RNA polymerases

(A) Binding free energies of ATP and RTP inside SARS-CoV-2 RdRp and human RNA Pol II (negative values corresponding to a preference for RTP over ATP).

(B) Most important interactions between RTP nitrile group and RdRp polymerase.

(C) RTP inside human RNA Pol II active site.

thermodynamic cycles as described above. As shown in Figure 2, the viral RdRp is characterized by a sizable preference for RTP, in agreement with available biochemical data,[14] which is a serendipitous fact given that it was originally designed to block an evolutionarily distinct polymerase in the Ebola virus. On the other hand, the lower affinity for the active site of human RNA Pol II suggests that remdesivir will rarely be incorporated into nascent human mRNA, partially explaining the drug's low toxicity in humans.[25] The observed preferences can be rationalized by inspecting stabilizing interactions inside enzymes' active sites. The nitrile group of RTP can accept two hydrogen bonds with Thr_{687} and Asn_{691} sidechains of RdRp of SARS-CoV-2 (Figure 2B). Similar interactions were observed in previous MD studies.[26,27] On the contrary, inside RNA Pol II, no hydrogen-bond donors able to stabilize the nitrile group are found (Figure 2C).

The SARS-CoV-2 RdRp activation mechanism

Before investigating the detailed mechanism of nucleotide incorporation, we explored the mechanism of O3′ deprotonation of the 3′-end ribose inside RdRp. This is a mandatory step of nucleophile formation that precedes nucleotide incorporation. We explored whether RdRp is able to activate the 3′-end nucleotide through a self-activated mechanism as has been proposed for other polymerases.[22] As shown in Figure 3 (see also Video S1), once a nucleotide has been incorporated inside RdRp, the newly created PPi molecule is found in perfect arrangement to abstract the O3′ hydrogen atom. Our simulations confirm that one non-bridging oxygen atom of PPi's γ phosphate group can abstract the proton of the 3′ hydroxyl with a free energy of activation of 8.1 kcal/mol. This barrier is ~3 kcal/mol higher than the reported value for O3′ deprotonation inside DNA polymerase eta (Pol-η).[28] Interestingly, the arrangement inside RdRp enables the direct transfer of the proton from the O3′ hydroxyl group to the γ phosphate group of the PPi, while in other polymerases, like the aforementioned DNA Pol-η, the β phosphate group firstly deprotonates the O3′ atom and then donates the proton to the γ group.[22] O3′ atom's deprotonation inside the RdRp step is found to be slightly endothermic as in DNA Pol's two-metal mechanism,[28] which, as described by us and others, is overcome by subsequent PPi–H release from the polymerase's active site.[29–33] Finally, following the replication cycle, nucleic-acid translocation and PPi–H departure from RdRp's active site enable the binding of the subsequent NTP molecule.[22]

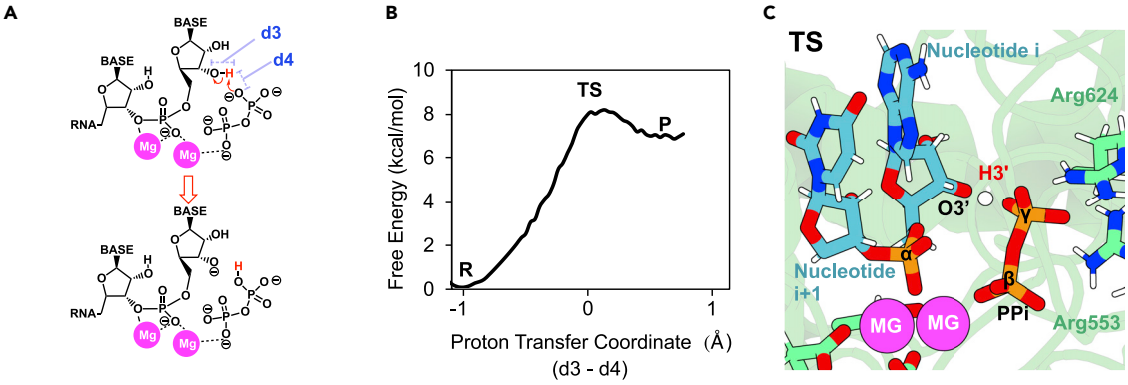


Figure 3. Mechanism of activation through O3′ deprotonation inside RdRp

(A) Scheme depicting the activation mechanism used by SARS-CoV-2 RdRp.

(B) Free-energy profile as a function of the proton transfer coordinate (d3-d4), obtained for the proton transfer reaction between the O3′ of the just-incorporated nucleotide and the γ phosphate group of the newly formed PPi.

(C) Transition state (TS) where the H3′ proton belonging to O3′ atom of nucleotide "i" is halfway to being transferred to γ phosphate group of PPi. Reaction coordinate (RC) consisted of the distances between O3′ or Oγ atoms to the H atom to be transferred, displayed as d3 and d4, respectively.

The SARS-CoV-2 RdRp reaction mechanism of nucleotide incorporation

QM/MM simulations were used (see experimental procedures and supplemental experimental procedures) to study the ability of SARS-CoV-2 RdRp to incorporate either a natural triphosphate (NTP; exemplified here by uridine-5′-triphosphate [UTP] and ATP) or RTP (remdesivir-TP) into nascent RNA. In the QM/MM reactant state, the Pα atom of the ATP is 3.6 ± 0.2 Å far from O3′ atom of the terminal nucleotide (see Figure S5). In addition, the O3′ atom is in line to perform the nucleophilic attack on the Pα atom of the ATP, displaying an angle of 173° ± 5° between Oα-Pα-O3′ atoms. While one Mg^{2+} cation is coordinated by the O3′ atom of the 3′ terminal, the Pα atom of the ATP, and carboxylate atoms of Asp_{618}, Asp_{760}, and Asp_{761}, the other Mg^{2+} cation is coordinated by the carbonyl group of Tyr_{619} and carboxylate atoms of Asp_{618} and Asp_{760}. In addition, the Mg^{2+}-Mg^{2+} distance is 3.6 ± 0.1 Å. Following the two-metal-ion mechanism scheme, one Mg^{2+} activates the O3′ atom toward the attack of Pα of NTP, and the other Mg^{2+} stabilizes the upcoming PPi leaving group, while both are in perfect disposition to stabilize the negatively charged transition state (TS). In the TS, the phosphoryl group is halfway to being transferred to the O3′ atom of the terminal nucleotide (see Figures 4B and S5). The O3′-Pα distance is 1.9 ± 0.1 Å, while the Pα-Oα distance is 2.1 ± 0.1 Å. Also, the Mg^{2+}-Mg^{2+} distance is slightly reduced by 0.2 ± 0.1 Å, while Mg^{2+}s maintained the same interactions with their respective coordination spheres. Metal-aided nucleotidyl transfer reactions can proceed through associative or dissociative TSs (S_N2 or S_N1, respectively) depending on the specific enzyme catalyzing the process.[29] It has been observed that while CRISPR-Cas9 proceeds through a concerted associative mechanism,[34] group II introns display a dissociative one.[35] Thus, in light of our results, the reaction inside SARS-CoV-2 RdRp proceeds through a concerted associative TS where the breaking and forming bond lengths are found to be similarly extended. Finally, in the product state, the nucleotide has been fully transfer reflected in a Pα-Oα distance of 3.5 Å, and a PPi molecule is formed (see Figure S5).

A careful analysis of trajectories and the free-energy profiles shows that the formation of a phosphodiester bond proceeds through a single free-energy maximum (TS) corresponding to an activation barrier of 15.8 kcal mol^{-1} for UTP and 16.2 kcal mol^{-1} for ATP (see Figures 4, S5, and S6 and Videos S2 and S3) and

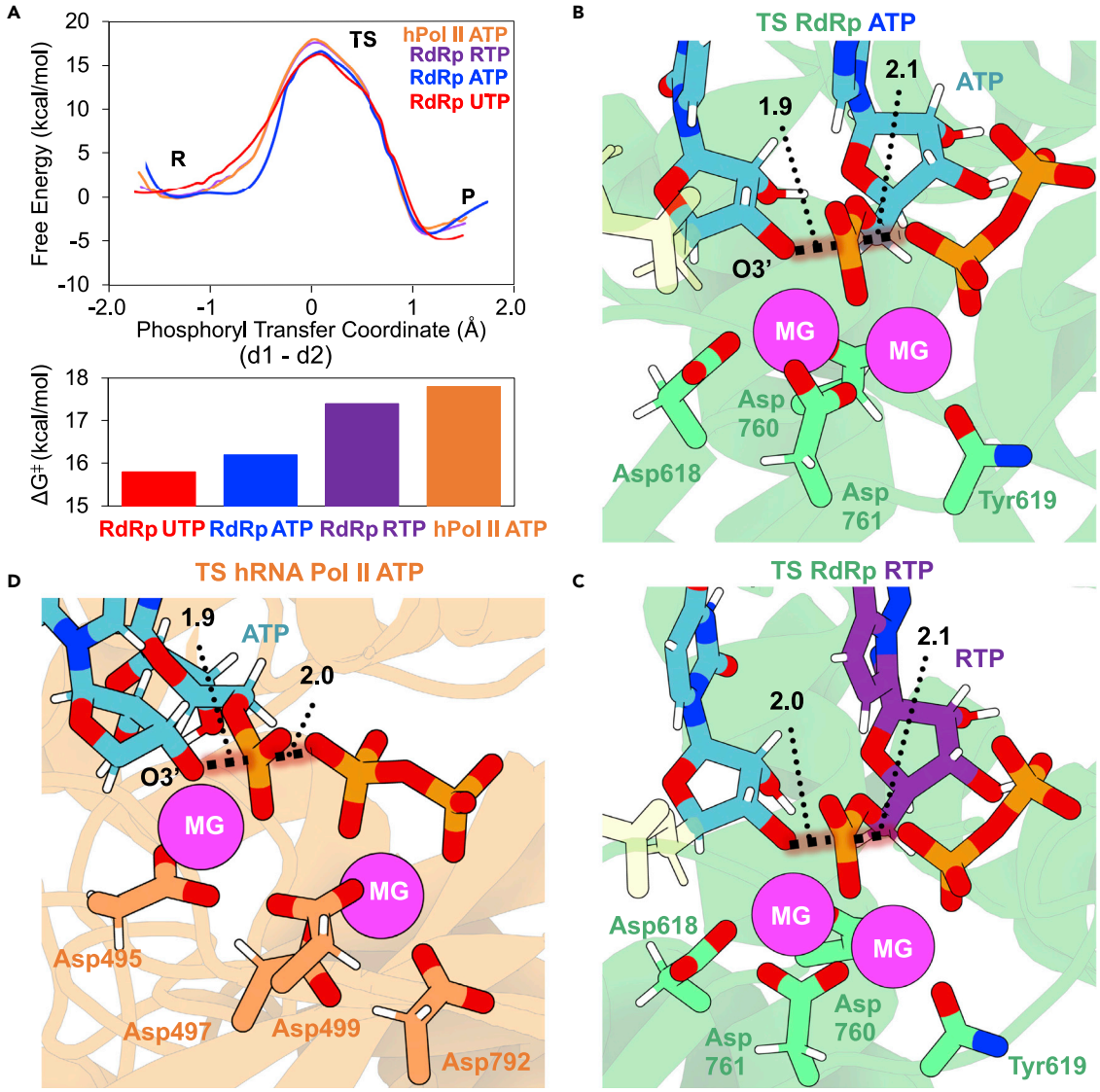


Figure 4. RNA elongation inside RdRp of SARS-CoV-2 and human RNA Pol II

(A) Top, free-energy profiles as a function of the phosphoryl transfer coordinate (d1-d2) for the incorporation of a U, an A, or an R to a nascent viral RNA strand. Profile for the incorporation of the inside human RNA Pol II is also shown. The phosphorylation reaction consists of a nucleophilic attack of the O3′ of the terminal nucleotide on the Pα of the triphosphate nucleotide. Bottom, bar plot displaying free energies of activation for the process for different NTPs and enzymes.

(B and C) Active site views of the TSs were found when ATP (B) and RTP (C) (purple C atoms) are the substrates for the elongation reaction inside RdRp. The phosphoryl group is half-way to being transferred. One Mg^{2+} activates the O3′ toward nucleophilic attack and stabilizes the negatively charged TS. The other Mg^{2+} stabilizes the charged TS, as well as the nascent negatively charged PPi molecule. Distances involved in the reaction are shown as dotted lines with their average values in Å.

(D) TS insight of the elongation reaction catalyzed by human RNA pol II.

is characterized by a negative free energy of ~5 kcal/mol (prior to PPi release). To obtain an estimate of the relative efficiency of the viral enzyme, we studied the RNA polymerization reaction catalyzed by human RNA Pol II (see Figures 4D and S7 and Video S4). Recently published kinetic rate constants[36] of nucleotide incorporation by SARS-CoV-2's RdRp provide support for the quantitative accuracy of our estimates, giving confidence to the suggested reaction mechanism. To evaluate the degree of fitness of the enzyme, we used the same procedure to predict

the RNA Pol II activation barrier (see Figure 4), getting a value of 17.8 kcal mol^{-1} for ATP, in good agreement with experimental estimates.[37,38] This value is ~1.5 kcal/mol higher than that of the viral RdRp (see Figure 4A), demonstrating that despite its short evolutionary trajectory, the efficiency of SARS-Cov-2 RdRp is at least similar to those of highly evolved eukaryotic polymerases.[39–41]

Our equilibrium trajectories demonstrate that RTP fits very well into the active site, showing strong canonical Watson-Crick interactions with the uridine in the template RNA. Being isosteric to adenosine, it achieves a perfect shape complementarity and arrangement of reactive groups in the binding site, predicting that it will act as a substrate rather than an inhibitor. This hypothesis is confirmed by QM/MM simulations showing that incorporation of RTP can happen with a free-energy barrier only slightly larger than that of a natural substrate (10% up to 17.4 kcal mol^{-1}; see Figures 4C and S6 and Video S5). Such an increase is mostly due to a slight misalignment of the Oα-Pα-O3′ attack angle (161° ± 8° for RTP, 173° ± 5° for ATP, and 172° ± 5° for UTP) due to RTP nitrile's group interactions (see Figures S8 and S9). Thus, we can rule out the possibility of remdesivir inhibiting RdRp by blocking the ATP-binding site of RNA polymerase. On the contrary, our simulation strongly suggests that RTP can be efficiently incorporated in front of uridine in a nascent RNA strand. Again, the order of incorporation predicted by our theoretical calculations, UTP > ATP > RTP, agrees perfectly with recent accurate pre-steady-state kinetic experiments,[36] providing additional support to our calculations (see comparison in Figure S10). Dangerfield et al. also observed that although RTP is incorporated more slowly than ATP inside RdRp, it is incorporated more efficiently than its counterpart due to a higher specificity constant (k_{cat}/K_m),[36] which also agrees with our estimates.

Remdesivir is well tolerated in an RNA duplex

Remdesivir is a C-nucleoside with an extra C1′ cyano group, and as such, it might distort the helix, causing delayed inhibition of the enzyme due to hampered displacement of the nascent duplex along the exit channel. To explore this possibility, we performed MD simulations of two RNA duplexes differing only in the substitution of a central r(A·U) pair by r(R·U) one (see experimental procedures). The results, summarized in Figure 5, strongly suggest that remdesivir is well tolerated in an RNA duplex and does not introduce any major structural distortion that would justify termination of RNA synthesis. In particular, there are no significant differences between the hydrogen-bonding stability of r(A·U) and r(R·U) pairs (see Figure 5C), and helical parameters of the duplexes are insensitive to the presence of remdesivir. In summary, our results strongly argue against the idea that a dramatic structural alteration of the RNA duplex is the key factor responsible for remdesivir-induced termination of RNA synthesis. Our results agree with recent cryo-EM structures where remdesivir was incorporated at different or multiple positions of the nascent RNA strand without altering RNA-duplex structure.[42,43]

Remdesivir does not block nascent-strand elongations through steric hindrance

Trying to explore alternative reasons for the inhibitory properties of remdesivir, we slid the nascent RNA duplex along the exit tunnel of RdRp to make the r(R·U) pair by simulating the addition of extra nucleotides, which allowed us to scan interactions of the RNA at several positions along the exit tunnel. After R was incorporated, we were not able to detect any point of steric clash that could justify stopping the polymerase progression (see Figure S12). Interestingly, when three more nucleotides were incorporated, we found Ser_{861}, whose sidechain is located at around 3.7 ± 0.3 Å from the nitrile group of remdesivir (see Figure S13A and S13B). However, considering the flexibility of the Ser sidechain, especially in a well-solvated

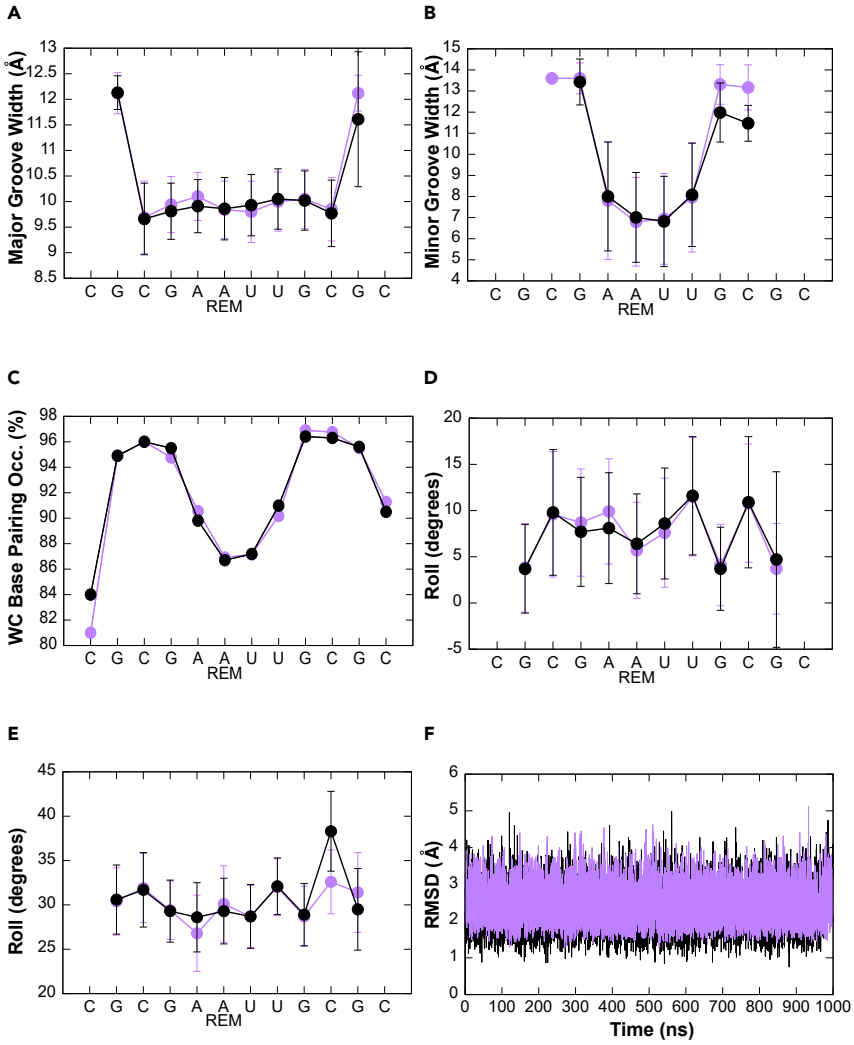


Figure 5. Remdesivir does not distort RNA structure

(A–F) Major (A) and minor (B) groove width, Watson-Crick base pairing (C), roll (D), twist (E), and RMSD (F) of the double helix are not affected when remdesivir is present (purple) with respect to control sequence (black) during MD simulations. The average values across the simulations are shown in black and purple dots for the control and remdesivir-containing sequences, respectively. Average standard deviations are shown as black and purple bars. The RMSD of one of the five replicas is shown in (F) (see also Figure S11).

microenvironment (see Figures S13A, S13C, and S13D), this strongly argues against the hypothesis that nitrile-Ser_{861} steric clash can explain the inhibitory properties of remdesivir.[14,42,43] To further discard the steric-clash hypothesis, we performed an alchemical mutational (A→R) scan along the nascent strand, looking at relative preferences of adenosine and remdesivir at positions from (i+2) to (i+5) in a poly-A RNA duplex embedded in the RdRp channel (see Figure 6). Very interestingly, A/R ΔG differences were very similar for positions i+2, i+3, and i+5, suggesting that there are not specific interactions between the incorporated nucleotide and the tunnel residues (as expected for an RNA polymerase designed to have a continuum output flux). On the contrary, significant differences are found at position (i+4), where a local free-energy minimum would hinder the movement of the nascent RNA toward the exit of the tunnel. The stabilization of remdesivir at position i+4 seems related to

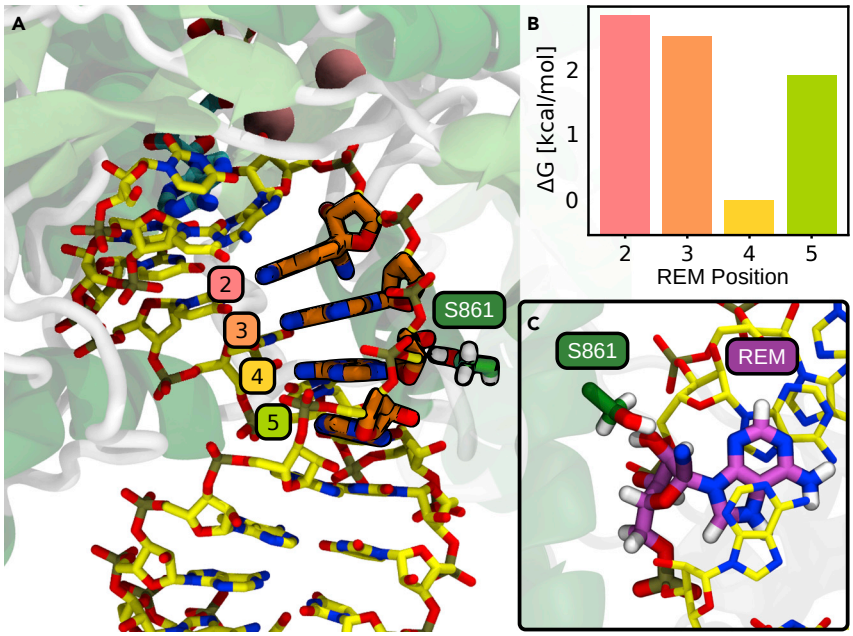


Figure 6. Remdesivir elongation along RdRp's exit channel

(A and B) Alchemically slided remdesivir along the i+2 to i+5 positions of the nascent RNA strand (A) and its associated free energies (B). In the i+4 position, remdesivir is energetically preferred with respect to neighboring positions.

(C) Insight displaying a stabilizing interaction between Ser861 and remdesivir is shown. Alchemically mutated adenines to remdesvir molecules are depicted in cyan, and the two Mg^{2+} cations of the active site are depicted as pink balls. The Ser861 residue, which is hypothesized to be involved in a steric clash, is shown.

the formation of direct or water-mediated hydrogen bonds between the cyano-group nitrogen and side-chain hydroxyl of Ser_{861} (see Figure S13).

Current biochemical and structural data show that remdesivir is stalled after 3 or 4 nucleotides are incorporated, but the reasons are unclear.[14,42–45] Also, biochemical experiments have showed that inhibition can be overcome with higher NTP concentrations.[14,42] In order to dissect the mechanism, structures of RdRp in post- and pre-translocated states with remdesivir in different positions of the nascent RNA strand have been resolved.[42,43] A previous MD study also suggested that the delayed termination mechanism could be due to remdesivir's destabilization of base pair interactions as well as of the nucleotide located in the active site.[26] Other experiments have suggested that when remdesivir is present in the template RNA strand, it would also inhibit RdRp through a secondary mechanism.[44] Some authors have noted as key for main inhibition a steric clash of the nitrile group with Ser_{861},[14,43–45] but our calculations failed to detect any steric clashes, suggesting the opposite, i.e., the vicinities of Ser_{861} are quite stabilizing for remdesivir. This would suggest that rather than being sterically unable to reach the position, remdesivir-containing RNA might be trapped in this position, hampering further sliding. Analysis of the recognition site (see Figures 6, S13C, and S13D) suggests that Ser_{861} could even be involved in a water-catalyzed Pinner's reaction,[46] which would lead to a transient covalent bond of RNA to the enzyme. The lack of stability of the resulting complex[47–50] precludes its experimental detection, but a transient covalent bond could contribute to trap remdesivir at the i+4 position, requiring an extra addition of NTP to escape from this stalling situation.[14]

To further explore the role of Ser_{861} as a stalling element in the sliding of remdesivir-containing RNA, we explored the exit channels of human RNA polymerases and CoV RdRps (see Figures S14 and S15). In the case of human RNA polymerases, no serine (or similar residue) was found in the expected displacement path of the remdesivir nitrile group (see Figure S14), suggesting that, if incorporated by human polymerases, remdesivir will not stop sliding of the nascent RNA. On the contrary, the same exercise made with other CoV RdRps (see Figure S15) found an amphipathic α-helix with a conserved Ser, which can play the same role of Ser_{861} in other coronaviruses. We can speculate that the same could happen for Ebola virus, the original targeted virus for remdesivir, but lack of structural information on Ebola RdRp precludes a detailed analysis.

Conclusions

SARS-CoV-2 RdRp is a protein common to other coronaviruses but shows little homology outside the family. As CoVs are just a few thousand years old, the protein has had a limited evolutionary period, and we could expect low efficiency. However, and quite surprisingly, our calculations demonstrate that the enzyme is very efficient, even more than eukaryotic polymerases. The viral enzyme follows a mechanism that is similar to that of bacterial or eukaryotic polymerases with the transferred phosphate being stabilized by 2 Mg^{2+} ions exquisitely coordinated by acidic residues of the catalytic site, while the phosphates of the incoming nucleotide are stabilized by a network of basic residues. The SARS-Cov-2 RdRp makes use of a self-activated mechanism where the gamma-phosphate group of PPi molecule deprotonates the hydroxylic 3′ terminal, generating the nucleophile that participates in the subsequent incorporation of a nucleotide.

Quite surprisingly, RTP is an excellent substitute of ATP, although it is not well recognized by human RNA Pol II. Simulations also show that after being bound to the viral active site, RTP does not block RdRp, but it is incorporated into the nascent RNA in front of a uridine. The resulting duplex does not show dramatic structural changes, which would hinder displacement of the nascent duplex along the exit channel. In fact, analysis of the displacement of the RNA-containing remdesivir along the exit tunnel of RdRp fails to detect points of steric clashes. Moreover, free-energy calculations show that the i+4 position is quite a favorable site for remdesivir. This suggests that the protein environment around i+4, particularly a hydrated serine, can act as a trap for the nascent RNA duplex, stalling the displacement of the helix by stabilizing remdesivir by either non-covalent or transient-covalent contacts.

EXPERIMENTAL PROCEDURES

Resource availability

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Modesto Orozco (modesto.orozco@irbbarcelona.org).

Materials availability

This study did not generate new unique reagents.

Data and code availability

The data that support the plots within this paper are available from the corresponding authors upon request. Trajectories will be deposited in the BIOEXCEL-COVID-19 database: https://bioexcel-cv19.bsc.es and will be accesible upon publication.

RdRp complex setup

The cryo-EM structure of the RdRp (nsp12) of SARS-CoV-2 complexed to nsp8 and nsp7[8] cofactors and the cryo-EM structure also complexed with an RNA duplex[18]

were used as our starting models. The first model was obtained when no other structure was available[46] by aligning the SARS-CoV-2 RdRp with RdRp from a hepatitis C virus X-ray structure,[51] from which RNA, the two catalytic metal ions, and the diphosphate group of an NTP analog were extracted (see supplemental experimental procedures for details). The structure was refined by comparing the cryo-EM structure of RdRp in complex with a full RNA strand[18] (see Figure S16). Finally, we built an ATP, UTP, or RTP molecule inside the active site of RdRp in the expected orientation required for incorporation into the nascent RNA strand. The systems were solvated and neutralized prior to optimization, thermalization (T = 298 K), and equilibration (see supplemental experimental procedures for details). The final equilibrated structures were the starting points for further MD simulations. The SARS-CoV RdRp starting model, including an RNA duplex and an ATP molecule, was built from an available cryo-EM structure[52] and subjected to the same modeling and equilibration protocols as described above. The progression of the polymerization process was simulated by adding additional base pair steps to the RNA duplex, moving the r(R·U) pair from i+1 to i+5 position while keeping constant the reactive alignments at the i position. These structures allowed us to trace the sliding of the nascent RNA duplex along the exit path and check for potential reasons for the R-induced delayed termination of the polymerization reaction.

Human RNA Pol II complex setup

The X-ray structure of human RNA Pol II (PDB: 5FLM)[53] consisting on a polymerase protein complex (composed on 12 subunits), a DNA template, and an RNA transcript were taken and submitted to further modeling. In order to obtain a fully processive protein enzyme, we aligned the aforementioned human RNA Pol II, whose trigger loop is in an opened conformation, to an RNA Pol II X-ray structure (PDB: 2E2H and 2E2J)[54] in the "closed" state. Finally, we extracted one Mg^{2+} cation and the triphosphate moiety, absents in the X-ray structure (PDB: 5FLM), from the closed-state X-ray structures (PDB: 2E2H and 2E2J) and introduced them to the final model of human RNA Pol II. The final structure was subjected to the same equilibration protocol as described above (see supplemental experimental procedures).

MD simulations on RdRp and human RNA Pol II

Classical trajectories were used to refine and check the stability of complexes prior to running QM/MM simulations, as well as to determine the binding mode and perform free-energy calculations. Minimization, thermalization, and equilibration were performed using standard procedures, as described in supplemental experimental procedures. Production simulations were carried out using the AMBER 19 program[55] and state-of-the-art conditions for a total time of at least 0.5 μs. Water molecules were described through the TIP3P[56] model, parameters of magnesium ions were taken from Allner et al.,[57] and Carlson et al.[58] parameters were used for triphosphate groups, PARMBSC1 for DNA,[59] PARMBSC0-chiOL3 for RNA,[60–63] and ff14SB[64] for the proteins. Parameters and charges of ATP, UTP, RTP, remdesivir nucleotide, and 3′ terminal nucleotides and R were derived to be compatible with the force fields making use of the RED server.[65] Additional details of the simulation setups can be found in supplemental experimental procedures and Figure S17.

QM/MM exploration of the minimum free-energy paths and potential of mean force

QM subsystems used in reactivity calculations for RdRp and RNA Pol II are shown in Figure S18. The link-atom method was employed to join QM and MM regions.[55] The hybrid QM/MM models were built using randomly selected snapshots obtained in

the last ns of unrestrained MD simulations, which were then minimized and re-equilibrated at a QM/MM-hybrid level of theory.

QM/MM-MD simulations were performed to obtain minimum free-energy paths (MFEPs) by means of the string method.[66,67] This method allowed us to explore different reaction mechanisms and select the preferred one in terms of free energy. Sixty to one-hundred twenty string nodes were used. Afterward, a path collective variable (CV)[67] was defined to obtain the potential of mean force (PMF) using 60 to 120 umbrella sampling[68] windows. The chosen set of CVs that followed the progress of the reactions and the breaking and forming bonds are shown in Figures 1B and 3A and supplemental experimental procedures (see Figure S18). MFEPs were obtained at the DFTB3[69,70]/MM level, and the PMFs were corrected at the B3LYP/6-311++G** level. Each of the sampling windows consisted on 20 ps of equilibration followed by 200 ps of production. We checked that the length of the production for the PMFs was sufficient to reduce the statistical error to the order of 1 kcal mol^{-1}. The error of all free-energy barriers and profiles was calculated as 95% confidence intervals and reached error values within ± 1 kcal mol^{-1}. The AMBER program[55] with electrostatic embedding was used for the QM/MM calculations. Corrections at the high level of theory were made with Gaussian16[71] (see supplemental experimental procedures).

Binding free-energy calculations

The difference binding free energy of ATP and RTP to polymerases was determined by computing the differences in the free energy associated with the ATP/RTP change in protein-complexed and isolated states.[72] The human and viral polymerases, complexed with an alchemical ATP/RTP residue in the binding site, were simulated in NPT conditions for 250 ns in both physical endpoints corresponding to ATP and RTP, and the initial 50 ns were considered equilibration. As a reference state, a single solvated ATP/RTP residue complexed with Mg^{2+} was simulated in the same manner. The resulting 200 ns trajectories were then used to extract seeding frames for the non-equilibrium free-energy protocol based on the Crooks theorem.[73] To obtain the binding free energies, 200 short 1 ns simulations were launched in each direction, with λ changing steadily from 0 to 1 or vice versa. Values of non-equilibrium work were computed for each run and converted to free energies using the Bennett acceptance ratio (BAR) method,[74] all using an in-house implementation of the protocol (gitlab.com/KomBioMol/crooks).

Free energy of remdesivir progression

To verify whether steric hindrance could be responsible for the stalling of RNA extension by RdRP, we performed alchemical free-energy simulations in which an adenine was mutated into remdesivir within the poly-A RNA duplex at four different positions (from i+2 to i+5). A non-equilibrium free-energy protocol virtually identical to the one outlined above was used to obtain relative free energies.

MD simulations on nascent RNA

We performed MD simulations on two duplexes, r(CGCGAAUUGCGC)·r(GCGCAAUUCGCG) and r(CGCGARUUGCGC)·r(GCGCAAUUCGCG), to determine the structural impact of the introduction of a remdesivir in a canonical RNA duplex. Starting structures were those expected for a canonical RNA duplex as implemented in AMBER. Systems were hydrated, minimized, thermalized, and equilibrated using standard protocols[75,76] before MD were performed at constant temperature (T = 298 K) and pressure (p = 1 atm). A total of 5 μs cumulative simulation time was

sampled for each system. Details of simulations are shown in supplemental experimental procedures.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.checat.2022.03.019.

ACKNOWLEDGMENTS

This work has been supported by the Spanish Ministry of Science (BFU2014-61670-EXP), the Catalan SGR, the Instituto Nacional de Bioinformática, the European Research Council (ERC SimDNA), the European Union's Horizon 2020 research and innovation program under grant agreement no. 676556, the Biomolecular and Bioinformatics Resources Platform (ISCIII PT 13/0001/0030) cofunded by the Fondo Europeo de Desarrollo Regional (FEDER), and the MINECO Severo Ochoa Award of Excellence (Government of Spain) (awarded to IRB Barcelona). M.O. is an ICREA academia researcher. J.A. acknowledges the Spanish Ministry of Science for a Juan de la Cierva post-doctoral grant. J.A. thanks Dr. K. Zinovjev for the string-method technical support. J.A. thanks the support team of the MareNostrum supercomputer and the support team at IRB Barcelona, specially R. Ramos. M.T. thanks the Instituto de Salud Carlos III for a Miguel Servet for a grant (CPII18/00032). Calculations were performed on the MareNostrum supercomputer at the Barcelona Supercomputer Center and in MMB cluster. M.O. and J.A. thank Prof. Núria López-Bigas and Dr. Francisco Martínez-Jiménez for their help in filtering data of mutations.

AUTHOR CONTRIBUTIONS

J.A. and M.O. conceived the project. J.A. conceptualized the goals and aims of the project. J.A. and M.W. performed simulations and analyzed the data. J.A. and M.O. supervised the project. M.O. acquired funding support for this project. J.A. prepared the initial manuscript. M.O., M.W., and J.A. reviewed and edited the initial manuscript and provided critical commentary and revisions. J.A. and M.O. prepared the final versions of the manuscript. M.T. and I.B.-H. commented on and discussed the experimental support of the project. All authors discussed, commented on, and revised the final version of the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: December 1, 2021
Revised: March 8, 2022
Accepted: March 24, 2022
Published: April 19, 2022

REFERENCES

1. WHO (2021). Novel Coronavirus (COVID-19) Situation (WHO).
2. De Wit, E., Van Doremalen, N., Falzarano, D., and Munster, V.J. (2016). SARS and MERS: recent insights into emerging coronaviruses. Nat. Rev. Microbiol. *14*, 523–534.
3. Robson, F., Khan, K.S., Le, T.K., Paris, C., Demirbag, S., Barfuss, P., Rocchi, P., and Ng, W.L. (2020). Coronavirus RNA proofreading: molecular basis and therapeutic targeting. Mol. Cell *79*, 710–727.
4. Hadfield, J., Megill, C., Bell, S.M., Huddleston, J., Potter, B., Callender, C., Sagulenko, P., Bedford, T., and Neher, R.A. (2020). Genomic epidemiology of novel coronavirus - global subsampling. Nextstrain real-time Track. Pathog. Evol. *26*, 2854.
5. Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., Wang, W., Song, H., Huang, B., Zhu, N., et al. (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet *395*, 565–574.
6. Rausch, J.W., Capoferri, A.A., Katusiime, M.G., Patro, S.C., and Kearney, M.F. (2020). Low genetic diversitymay be an Achilles heel of SARS-CoV-2. Proc. Natl. Acad. Sci. U S A *117*, 24614–24616.
7. Peng, Q., Peng, R., Yuan, B., Zhao, J., Wang, M., Wang, X., Wang, Q., Sun, Y., Fan, Z., Qi, J., et al. (2020). Structural and biochemical characterization of the nsp12-nsp7-nsp8 core

polymerase complex from SARS-CoV-2. Cell Rep. *31*, 107774.

8. 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 *368*, 779–782.

9. Hillen, H.S., Kokic, G., Farnung, L., Dienemann, C., Tegunov, D., and Cramer, P. (2020). Structure of replicating SARS-CoV-2 polymerase. Nature *584*, 154–156.

10. Romano, M., Ruggiero, A., Squeglia, F., Maga, G., and Berisio, R. (2020). A structural view of SARS-CoV-2 RNA replication machinery: RNA synthesis, proofreading and final capping. Cells *9*, 1267.

11. V'kovski, P., Kratzel, A., Steiner, S., Stalder, H., and Thiel, V. (2020). Coronavirus biology and replication: implications for SARS-CoV-2. Nat. Rev. Microbiol. *193*, 155–170.

12. McGinnis, S., and Madden, T.L. (2004). BLAST: At the core of a powerful and diverse set of sequence analysis tools. Nucleic Acids Res. *32*, W20–W25.

13. Gordon, C.J., Tchesnokov, E.P., Feng, J.Y., Porter, D.P., and Götte, M. (2020). The antiviral compound remdesivir potently inhibits RNAdependent RNA polymerase from Middle East respiratory syndrome coronavirus. J. Biol. Chem. *295*, 4773–4779.

14. Gordon, C.J., Tchesnokov, E.P., Woolner, E., Perry, J.K., Feng, J.Y., Porter, D.P., and Götte, M. (2020). 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. *295*, 6785–6797.

15. Yang, H., and Rao, Z. (2021). Structural biology of SARS-CoV-2 and implications for therapeutic development. Nat. Rev. Microbiol. *19*, 685–700.

16. U.S. National Library of Medicine (2020). Expanded Access Treatment Protocol: Remdesivir (RDV; GS-5734) for the Treatment of SARS-CoV2 (CoV) Infection (NCT04323761) (U.S. National Library of Medicine).

17. NCT03719586 (2018). Investigational Therapeutics for the Treatment of People with Ebola Virus Disease. https://clinicaltrials.gov/show/nct03719586.

18. 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 *368*, 1499–1504.

19. Wang, Q., Wu, J., Wang, H., Gao, Y., Liu, Q., Mu, A., Ji, W., Yan, L., Zhu, Y., Zhu, C., et al. (2020). Structural basis for RNA replication by the SARS-CoV-2 polymerase. Cell *182*, 417–428.e13.

20. 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. (2021). Structure of the SARS-CoV-2 RNA-dependent RNA polymerase in the presence of favipiravir-RTP. Proc. Natl. Acad. Sci. U S A *118*, e2021946118.

21. Peng, Q., Peng, R., Yuan, B., Wang, M., Zhao, J., Fu, L., Qi, J., and Shi, Y. (2021). Structural basis of SARS-CoV-2 polymerase inhibition by favipiravir. Innovation (N Y) *2*, 100080.

22. Genna, V., Vidossich, P., Ippoliti, E., Carloni, P., and Vivo, M.D. (2016). A self-activated mechanism for nucleic acid polymerization catalyzed by DNA/RNA polymerases. J. Am. Chem. Soc. *138*, 14592–14598.

23. 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.

24. Tomasello, G., Armenia, I., and Molla, G. (2020). The Protein Imager: a full-featured online molecular viewer interface with server-side HQ-rendering capabilities. Bioinformatics *36*, 2909–2911.

25. Warren, T.K., Jordan, R., Lo, M.K., Ray, A.S., Mackman, R.L., Soloveva, V., Siegel, D., Perron, M., Bannister, R., Hui, H.C., et al. (2016). Therapeutic efficacy of the small molecule GS-5734 against Ebola virus in rhesus monkeys. Nature *531*, 381–385.

26. Byléhn, F., Menéndez, C.A., Perez-Lemus, G.R., Alvarado, W., and De Pablo, J.J. (2021). Modeling the binding mechanism of remdesivir, favilavir, and ribavirin to SARS-CoV-2 RNA-dependent RNA polymerase. ACS Cent. Sci. *7*, 164–174.

27. Wakchaure, P.D., Ghosh, S., and Ganguly, B. (2020). Revealing the inhibition mechanism of RNA-dependent RNA polymerase (RdRp) of SARS-CoV-2 by remdesivir and nucleotide analogues: a molecular dynamics simulation study. J. Phys. Chem. B *124*, 10641–10652.

28. Stevens, D.R., and Hammes-Schiffer, S. (2018). Exploring the role of the third active site metal ion in DNA polymerase η with QM/MM free energy simulations. J. Am. Chem. Soc. *140*, 8965–8969.

29. Genna, V., Donati, E., and De Vivo, M. (2018). The catalytic mechanism of DNA and RNA polymerases. ACS Catal. *8*, 11103–11118.

30. Silva, D.A., Weiss, D.R., Avila, F.P., Da, L.T., Levitt, M., Wang, D., and Huang, X. (2014). Millisecond dynamics of RNA polymerase II translocation at atomic resolution. Proc. Natl. Acad. Sci. U S A *111*, 7665–7670.

31. Genna, V., Gaspari, R., Dal Peraro, M., and De Vivo, M. (2016). Cooperative motion of a key positively charged residue and metal ions for DNA replication catalyzed by human DNA polymerase-η. Nucleic Acids Res. *44*, 2827–2836.

32. Da, L.T., Wang, D., and Huang, X. (2012). Dynamics of pyrophosphate ion release and its coupled trigger loop motion from closed to open state in RNA polymerase II. J. Am. Chem. Soc. *134*, 2399–2406.

33. Aranda, J., Terrazas, M., Gómez, H., Villegas, N., and Orozco, M. (2019). An artificial DNAzyme RNA ligase shows a reaction mechanism resembling that of cellular polymerases. Nat. Catal. *2*, 544–552.

34. Casalino, L., Nierzwicki, Ł., Jinek, M., and Palermo, G. (2020). Catalytic mechanism of non-target DNA cleavage in CRISPR-cas9 revealed by ab initio molecular dynamics. ACS Catal. *10*, 13596–13605.

35. Casalino, L., Palermo, G., Rothlisberger, U., and Magistrato, A. (2016). Who activates the nucleophile in ribozyme catalysis? An answer from the splicing mechanism of group II introns. J. Am. Chem. Soc. *138*, 10374–10377.

36. 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.

37. Kellinger, M.W., Ulrich, S., Chong, J., Kool, E.T., and Wang, D. (2012). Dissecting chemical interactions governing RNA polymerase II transcriptional fidelity. J. Am. Chem. Soc. *134*, 8231–8240.

38. Geronimo, I., Vidossich, P., Donati, E., and De Vivo, M. (2021). Computational investigations of polymerase enzymes: structure, function, inhibition, and biotechnology. Wiley Interdiscip. Rev. Comput. Mol. Sci. *11*, e1534.

39. Berdis, A.J. (2009). Mechanisms of DNA polymerases. Chem. Rev. *109*, 2862–2879.

40. Castro, C., Smidansky, E.D., Arnold, J.J., Maksimchuk, K.R., Moustafa, I., Uchida, A., Götte, M., Konigsberg, W., and Cameron, C.E. (2009). Nucleic acid polymerases use a general acid for nucleotidyl transfer. Nat. Struct. Mol. Biol. *16*, 212–218.

41. Wu, W.J., Yang, W., and Tsai, M.D. (2017). How DNA polymerases catalyse replication and repair with contrasting fidelity. Nat. Rev. Chem. *1*, 1–16.

42. Bravo, J.P.K., Dangerfield, T.L., Taylor, D.W., and Johnson, K.A. (2021). Remdesivir is a delayed translocation inhibitor of SARS-CoV-2 replication. Mol. Cell *81*, 1548–1552.e4.

43. Kokic, G., Hillen, H.S., Tegunov, D., Dienemann, C., Seitz, F., Schmitzova, J., Farnung, L., Siewert, A., Höbartner, C., and Cramer, P. (2021). Mechanism of SARS-CoV-2 polymerase stalling by remdesivir. Nat. Commun. *12*, 1–7.

44. Tchesnokov, E.P., Gordon, C.J., Woolner, E., Kocinkova, D., Perry, J.K., Feng, J.Y., Porter, D.P., and Götte, M. (2020). Template-dependent inhibition of coronavirus RNA-dependent RNA polymerase by remdesivir reveals a second mechanism of action. J. Biol. Chem. *295*, 16156–16165.

45. Seifert, M., Bera, S.C., van Nies, P., Kirchdoerfer, R.N., Shannon, A., Le, T.-T.-N., Meng, X., Xia, H., Wood, J.M., Harris, L.D., et al. (2021). Inhibition of SARS-CoV-2 polymerase by nucleotide analogs from a single-molecule perspective. Elife *10*, e70968.

46. Aranda, J., and Orozco, M. (2020). RNA-dependent RNA polymerase from SARS-CoV-2. Mechanism of reaction and inhibition by remdesivir. Preprint at bioRxiv. https://doi.org/10.1101/2020.06.21.163592.

47. Moon, J.B., Coleman, R.S., and Hanzlik, R.P. (1986). Reversible covalent inhibition of papain by a peptide nitrile. 13C NMR evidence for a thioimidate ester adduct. J. Am. Chem. Soc. *108*, 1350–1351.

48. Kalgutkar, A.S., and Dalvie, D.K. (2012). Drug discovery for a new generation of covalent drugs. Expert Opin. Drug Discov. *7*, 561–581.

49. Devkota, A.K., Edupuganti, R., Yan, C., Shi, Y., Jose, J., Wang, Q., Kaoud, T.S., Cho, E.J., Ren, P., and Dalby, K.N. (2014). Reversible covalent inhibition of eEF-2K by carbonitriles. ChemBioChem *15*, 2435–2442.

50. Berteotti, A., Vacondio, F., Lodola, A., Bassi, M., Silva, C., Mor, M., and Cavalli, A. (2014). Predicting the reactivity of nitrile-carrying compounds with cysteine: a combined computational and experimental study. ACS Med. Chem. Lett. *5*, 501–505.

51. 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 *347*, 771–775.

52. Kirchdoerfer, R.N., and Ward, A.B. (2019). Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. Nat. Commun. *10*, 1–9.

53. Bernecky, C., Herzog, F., Baumeister, W., Plitzko, J.M., and Cramer, P. (2016). Structure of transcribing mammalian RNA polymerase II. Nature *529*, 551–554.

54. Wang, D., Bushnell, D.A., Westover, K.D., Kaplan, C.D., and Kornberg, R.D. (2006). Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis. Cell *127*, 941–954.

55. 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 (University of California).

56. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., and Klein, M.L. (1983). Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. *79*, 926.

57. Allnér, O., Nilsson, L., and Villa, A. (2012). Magnesium ion-water coordination and exchange in biomolecular simulations. J. Chem. Theor. Comput. *8*, 1493–1502.

58. Meagher, K.L., Redman, L.T., and Carlson, H.A. (2003). Development of polyphosphate parameters for use with the AMBER force field. J. Comput. Chem. *24*, 1016–1025.

59. Ivani, I., Dans, P.D., Noy, A., Pérez, A., Faustino, I., Hospital, A., Walther, J., Andrio, P., Goñi, R., Balaceanu, A., et al. (2015). Parmbsc1: a refined force field for DNA simulations. Nat. Methods *13*, 55–58.

60. 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.

61. 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.

62. 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.

63. 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. Theor. Comput. *7*, 2886–2902.

64. 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. Theor. Comput. *11*, 3696–3713.

65. 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*, W511–W517.

66. 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*, 05B605.

67. Zinovjev, K., and Tuñón, I. (2017). Adaptive finite temperature string method in collective variables. J. Phys. Chem. A *121*, 9764–9772.

68. 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.

69. Gaus, M., Cui, Q., and Elstner, M. (2011). DFTB3: extension of the self-consistent-charge density-functional tight-binding method (SCC-DFTB). J. Chem. Theor. Comput. *7*, 931–948.

70. 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. Theor. Comput. *10*, 1518–1537.

71. 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., et al. (2016). Gaussian 16, Rev. B.01 (Gaussian, Inc.).

72. Lybrand, T.P., McCammon, J.A., and Wipff, G. (1986). Theoretical calculation of relative binding affinity in host-guest systems. Proc. Natl. Acad. Sci. U S A *83*, 833–835.

73. Crooks, G.E. (1999). Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. Phys. Rev. E *60*, 2721.

74. Bennett, C.H. (1976). Efficient estimation of free energy differences from Monte Carlo data. J. Comput. Phys. *22*, 245–268.

75. Dans, P.D., Walther, J., Gómez, H., and Orozco, M. (2016). Multiscale simulation of DNA. Curr. Opin. Struct. Biol. *37*, 29–45.

76. Dans, P.D., Ivani, I., Hospital, A., Portella, G., González, C., and Orozco, M. (2017). How accurate are accurate force-fields for B-DNA? Nucleic Acids Res. *45*, 4217–4230.

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

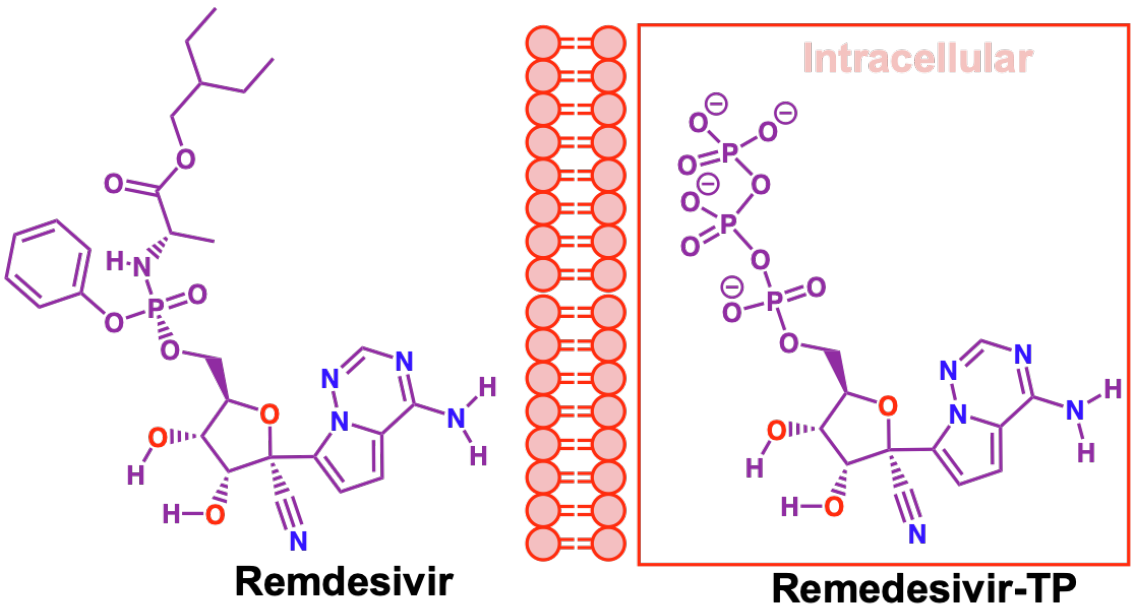


Figure S1. Chemical structures of Remdesirvir in its prodrug form and as a triphosphate nucleoside inside the cell.

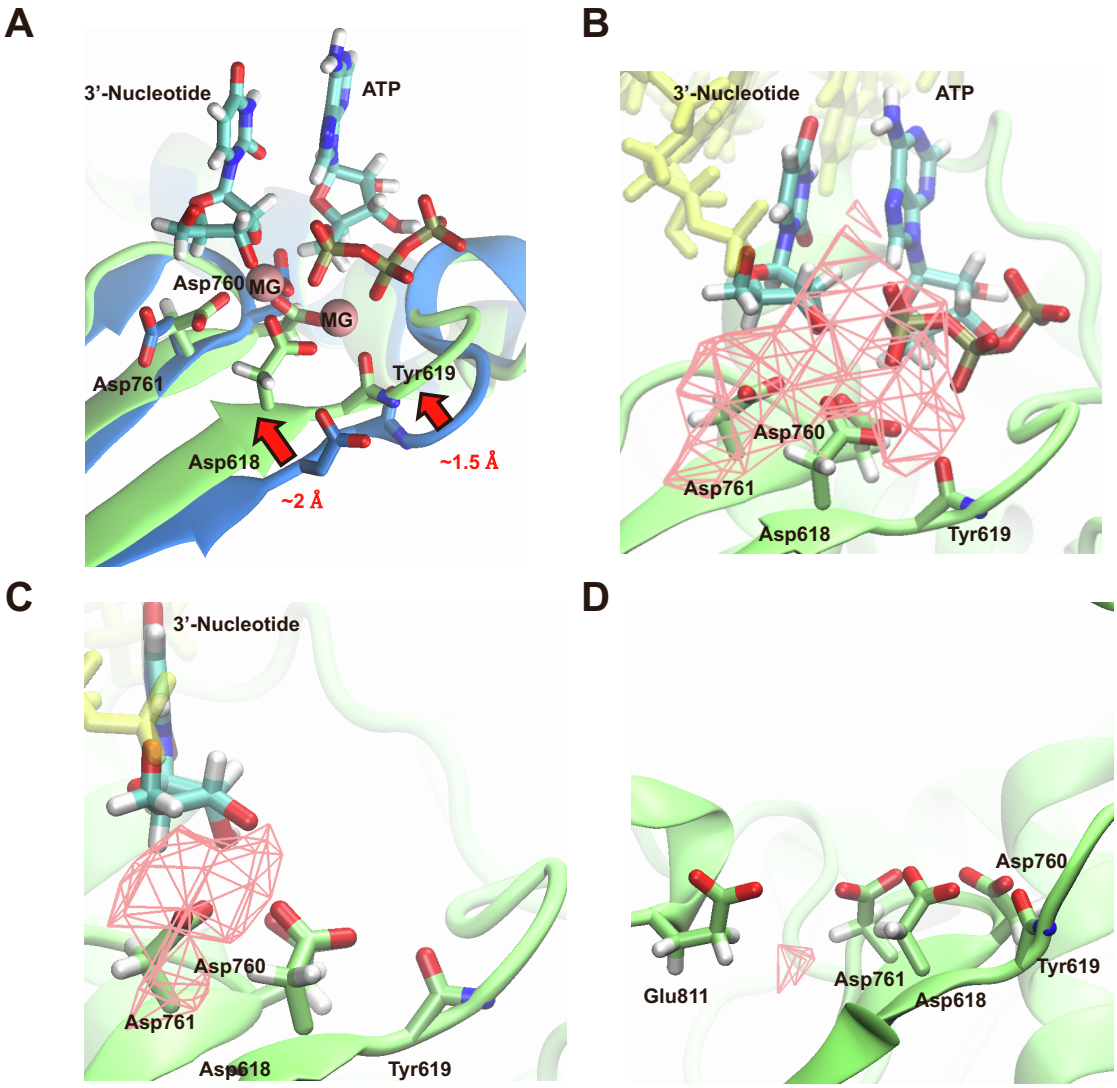


Figure S2. A Structural alignment of SARS-CoV-2 RdRp (PDB ID: 6M71, in blue) with the obtained catalytically competent structure of this study (in green). Values next to arrows indicate the difference (in Å) between structures, measured as the distance between Cα's corresponding to residues Asp_{618} 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}.

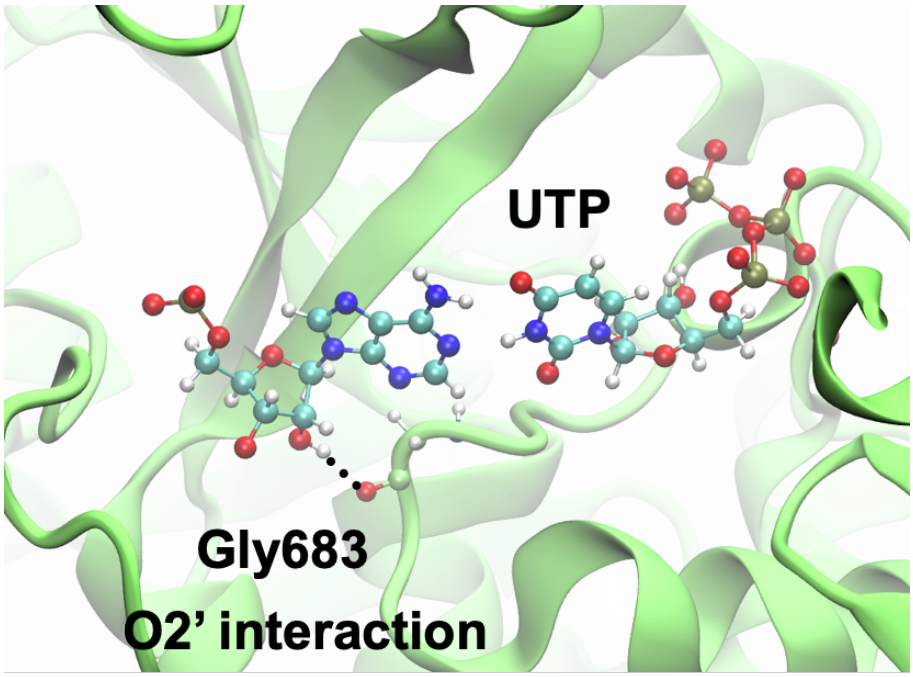


Figure S3. Gly_{683} forms a hydrogen bond interaction with the O2' hydroxyl group of the template nucleotide in "i" position, recognizing the entry of a ribonucleotide template.

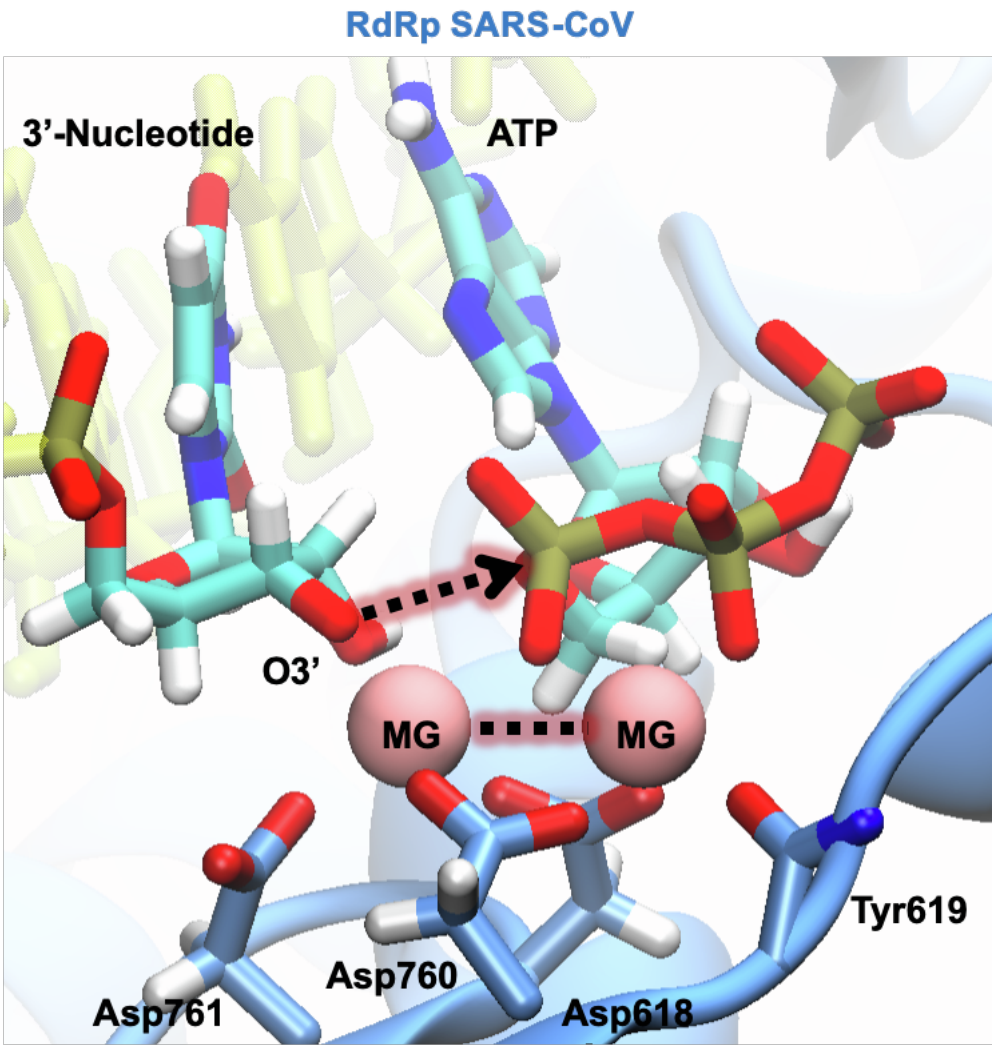


Figure S4. SARS-CoV RdRp's active site structure displays a very similar arrangement as compared to SARS-CoV-2 RdRp's during our 0.5 μs MD simulations.

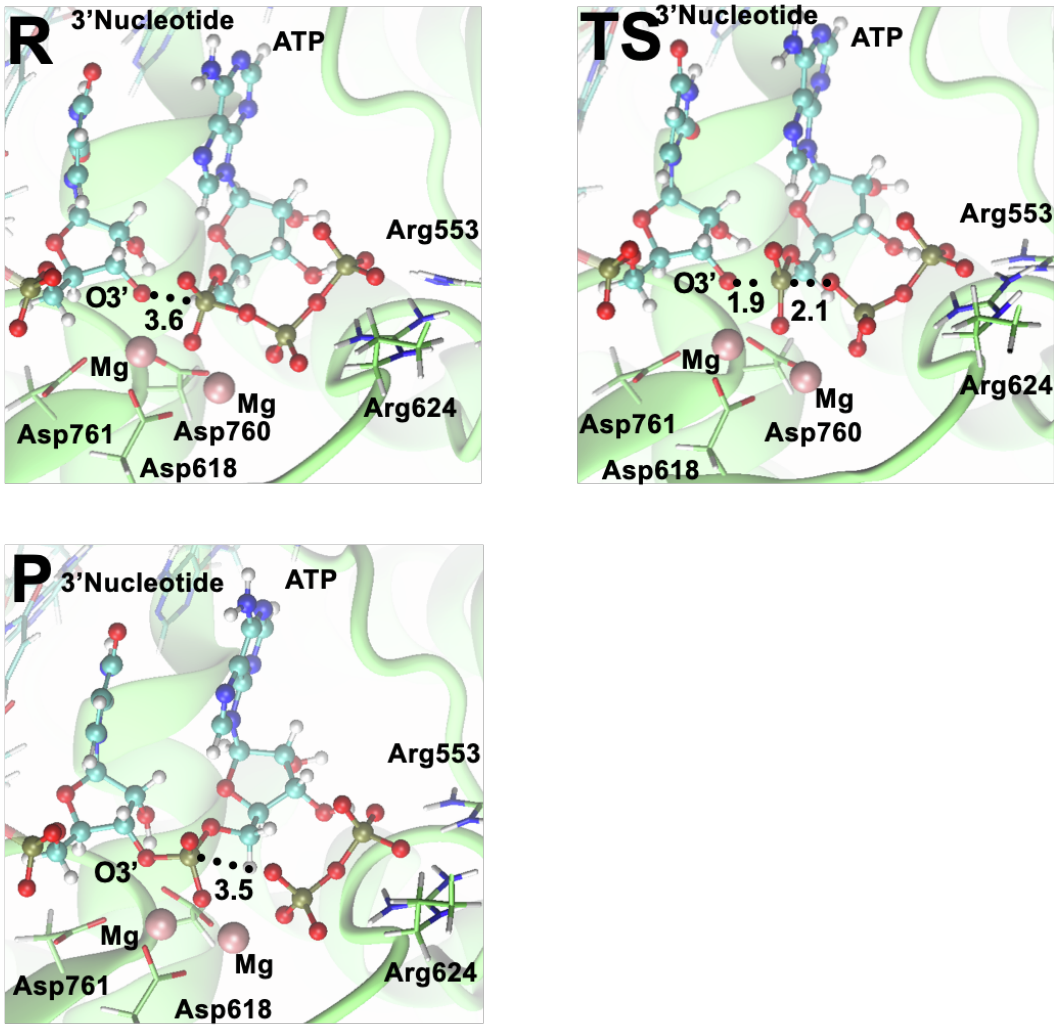


Figure S5. Active site insights of RdRp when a A is being incorporated to a nascent viral RNA strand. Reactant, Transition and Product states representative structures are depicted. Average distances involved in the phosphoryl transfer are shown in Å and depicted as dotted lines.

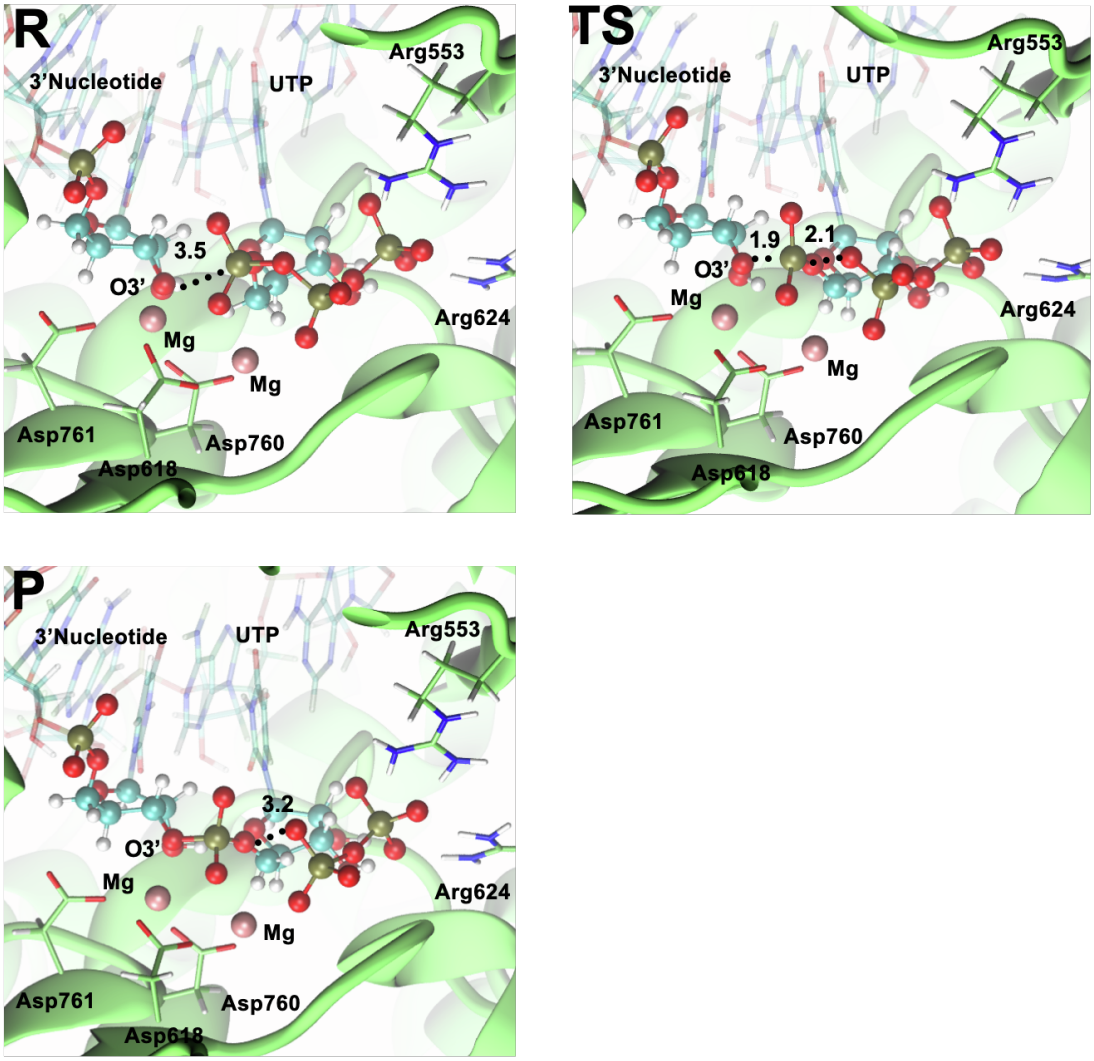


Figure S6. Active site insights of RdRp when a U is being incorporated to a nascent viral RNA strand. Reactant, Transition and Product states representative structures are depicted. Average distances involved in the phosphoryl transfer are shown in Å and depicted as dotted lines.

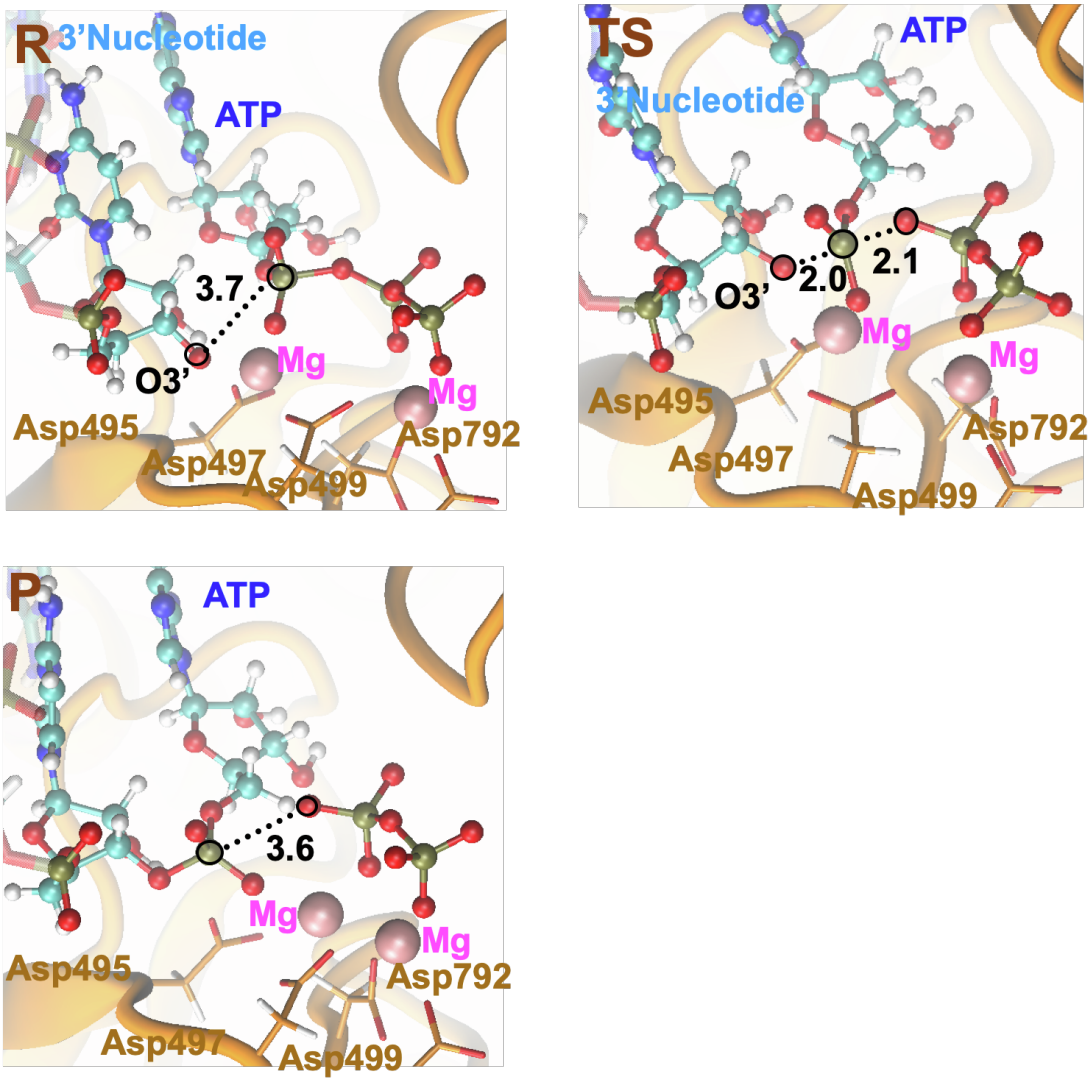


Figure S7. Active site insights of human RNA Pol II (shown in orange) when an A is being incorporated to a nascent RNA strand. Reactant, Transition and Product states representative structures are depicted. Average distances involved in the phosphoryl transfer are shown in Å and depicted as dotted lines.

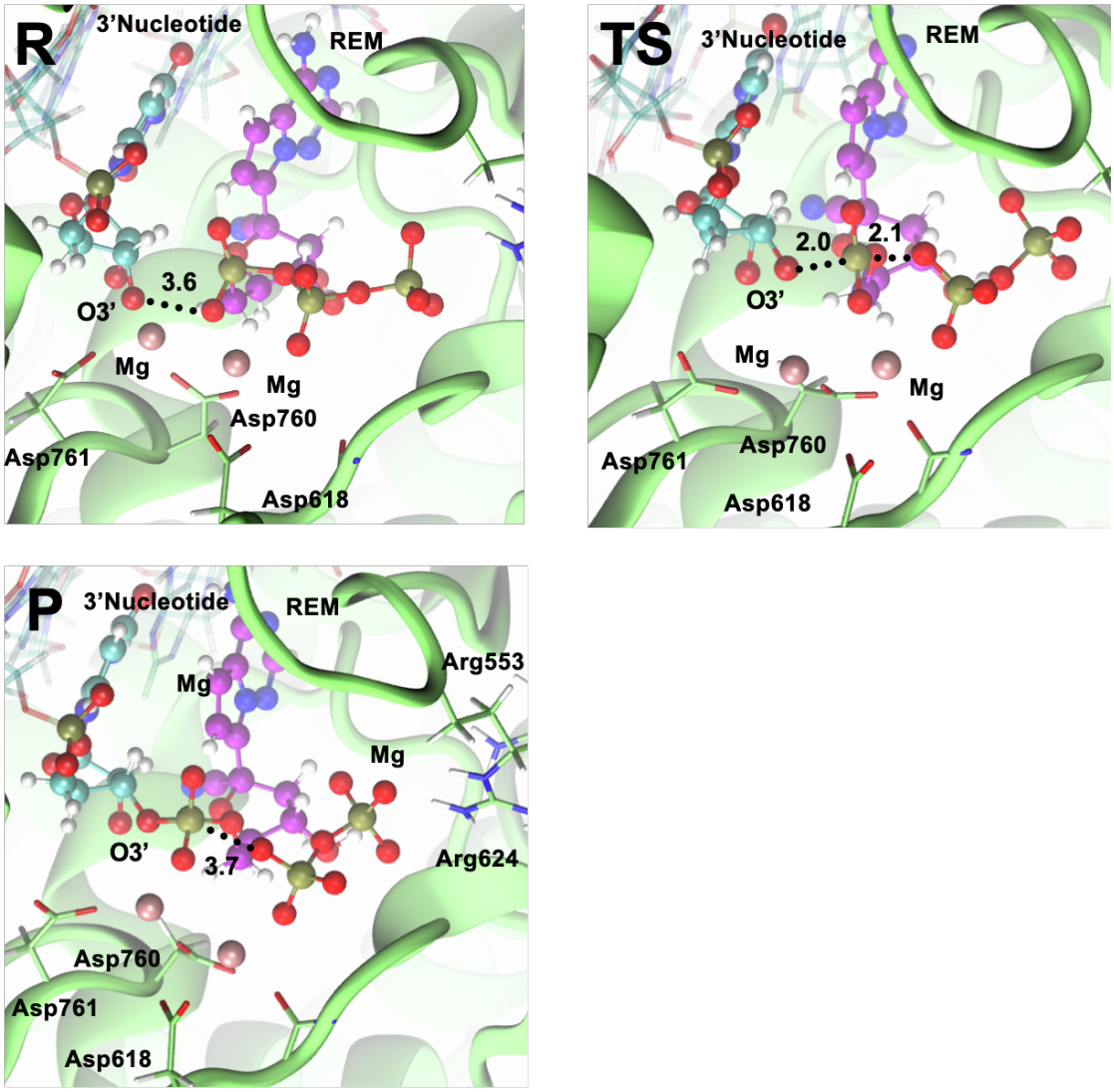


Figure S8. Active site insights of RdRp when a remdesivir is being incorporated to a nascent viral RNA strand. Reactant, Transition and Product states representative structures are depicted. Average distances involved in the phosphoryl transfer are shown in Å and depicted as dotted lines.

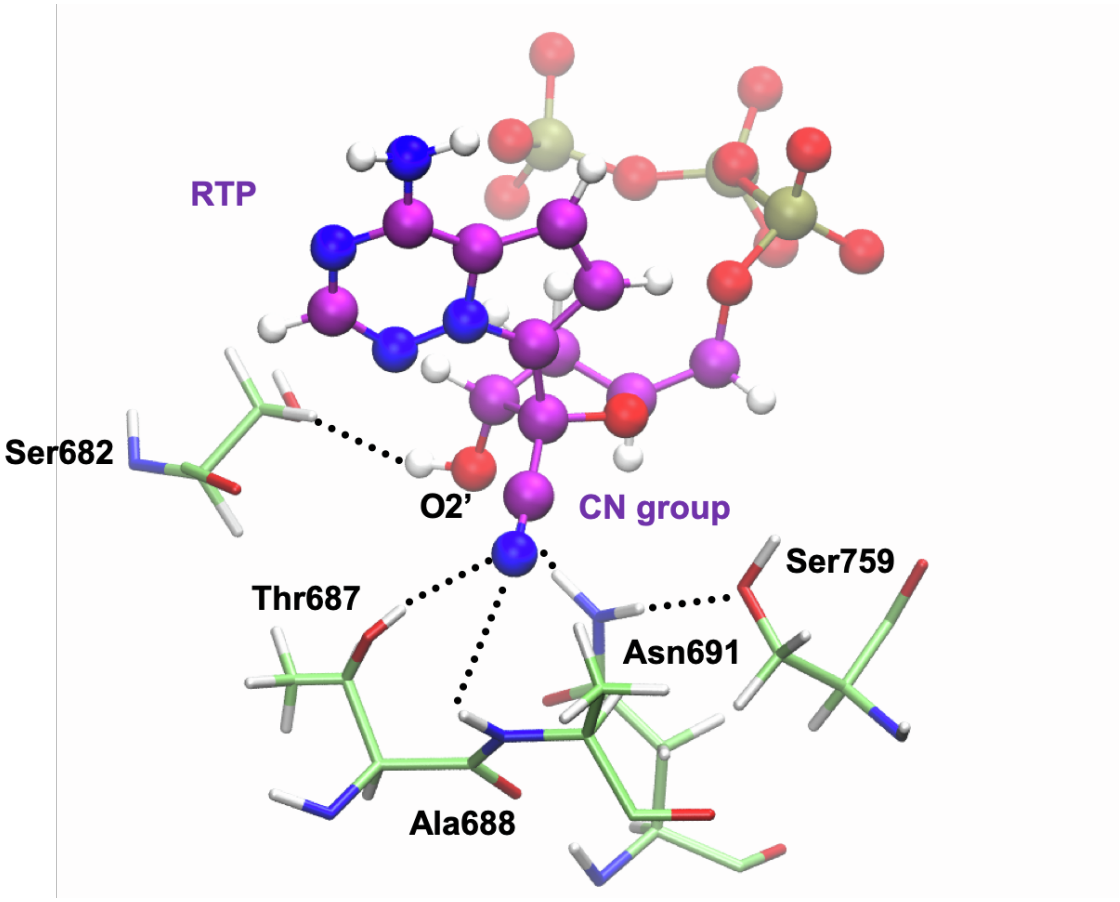


Figure S9. Interactions that are formed between the nitrile group of remdesivir-TP (RTP) and the active site of RdRp of SARS-CoV-2 during our simulations. remdesivir C atoms are shown in purple. Most important interactions are depicted as dotted lines. These interactions place remdesivir nitrile group 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.

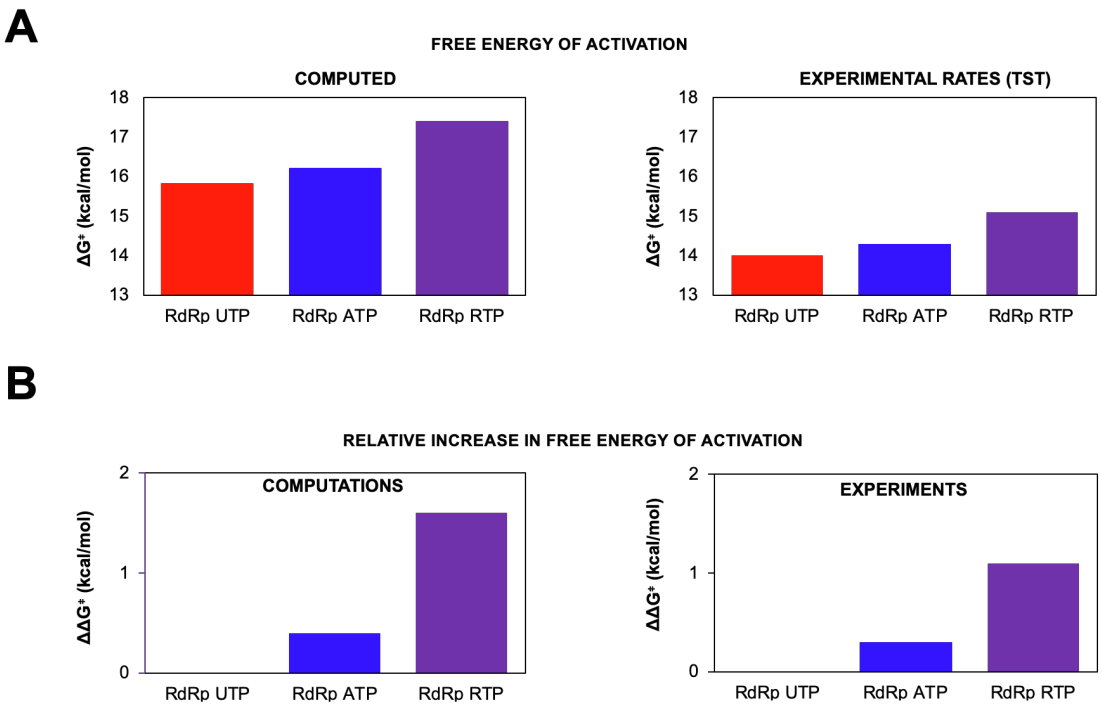


Figure S10. **A** Free energy of activation of UTP/ATP/RTP incorporation inside RdRp calculated (left panel) through QM/MM simulations in this study, and obtained through accurate pre-steady-state kinetic experiments[1] (right panel). **B** Increase in free energy of activation of incorporating an ATP 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.

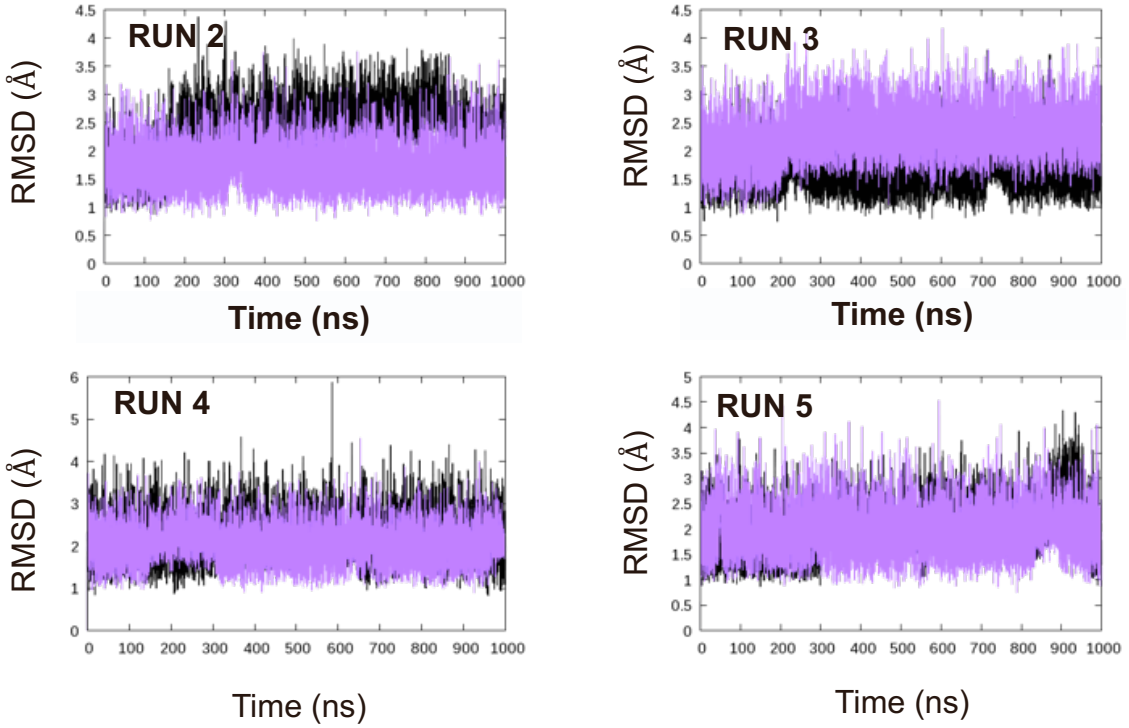


Figure S11. RMSD values in Å for the control sequence (in black) and the remdesivir-containing sequence (in purple) during 1 μs long MD simulations. 5 runs for each sequence were conducted including the run shown in Figure 5.

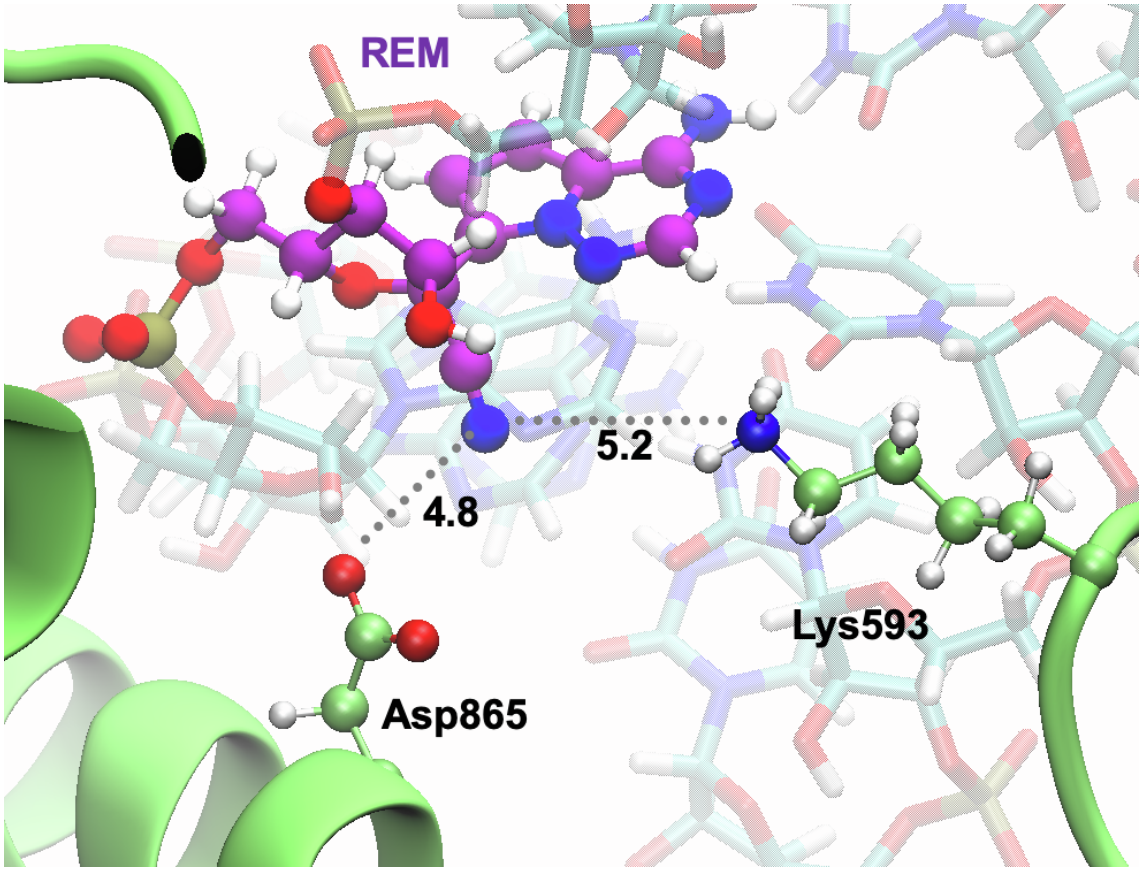


Figure S12. Closest residues of RdRp to remdesivir after two nucleotides have been incorporated. Residues are far to interact with any polar atom of remdesivir. Average distances during MD simulation between Asp_{865} and Lys_{593} sidechains to N atom of nitrile group of remdesivir are shown in Å.

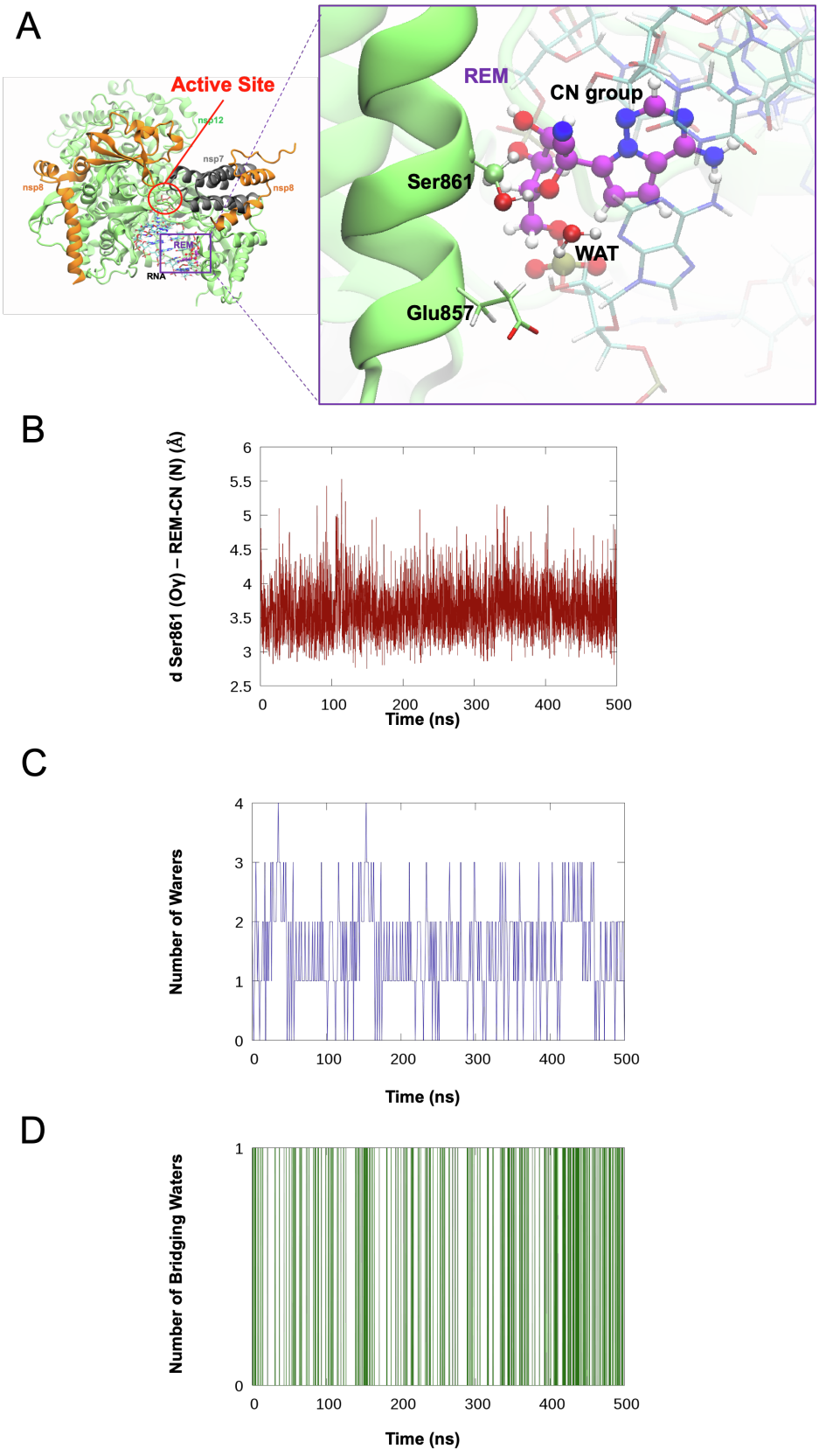


Figure S13. A Insight of the surroundings when remdesivir CN group is close to Ser_{861}. **B** Ser_{861} is found close to the nitrile group along 500 ns of MD simulation. **C** Number of waters within 3.5 Å distance from nitrile group. **D** Number of waters hydrogen bonded at the same time with the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

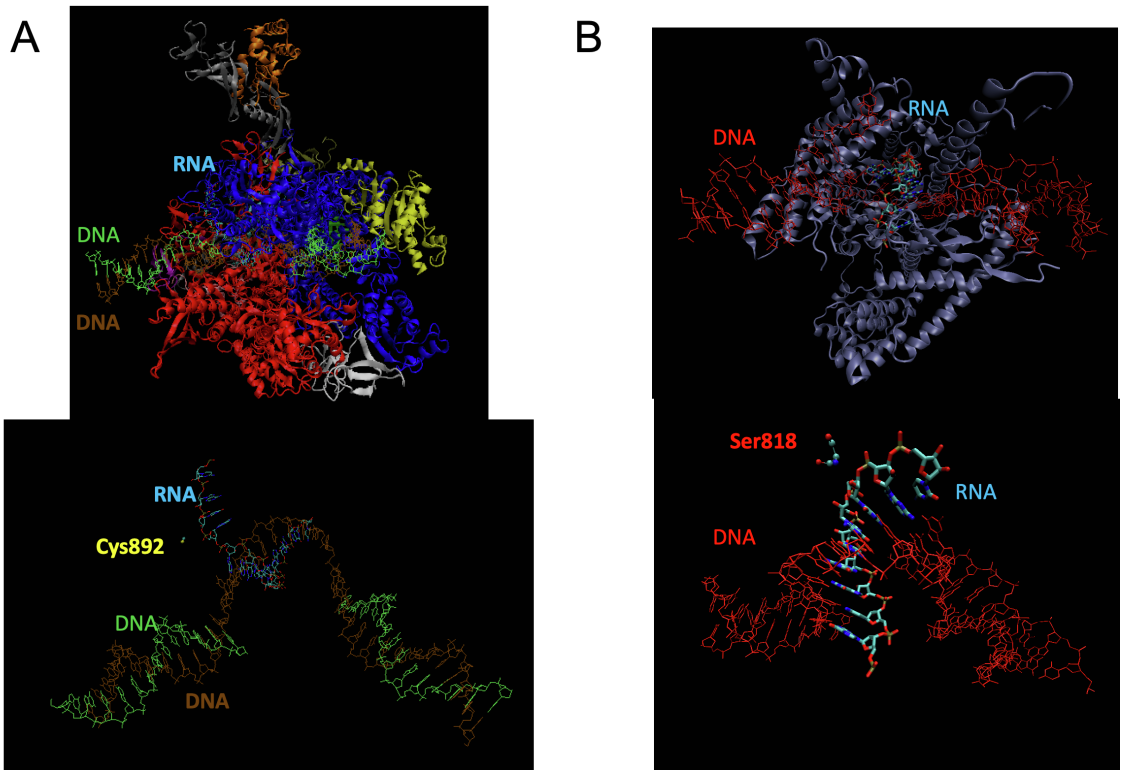


Figure S14. A Human RNA polymerase II-DNA/RNA complex (PDB ID: 5FLM[2]). Cys_{861} which is the closest Cys or Ser residue to a C1' atom of the nascent RNA strand, is placed to more than 8 Å distance. **B** Human mitochondrial RNA polymerase in complex with DNA and RNA, PDB ID: 4BOC.[3] Ser_{818}, the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

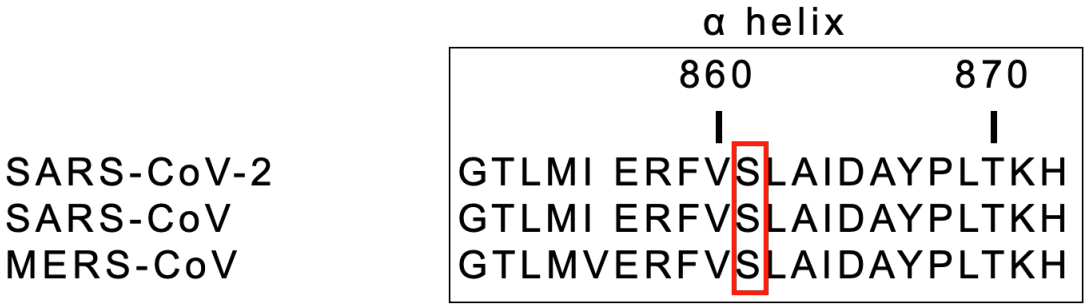


Figure S15. Alignment of SARS-CoV-2, SARS-CoV and MERS-CoV sequence for the alpha helix where Ser_{861} is placed.

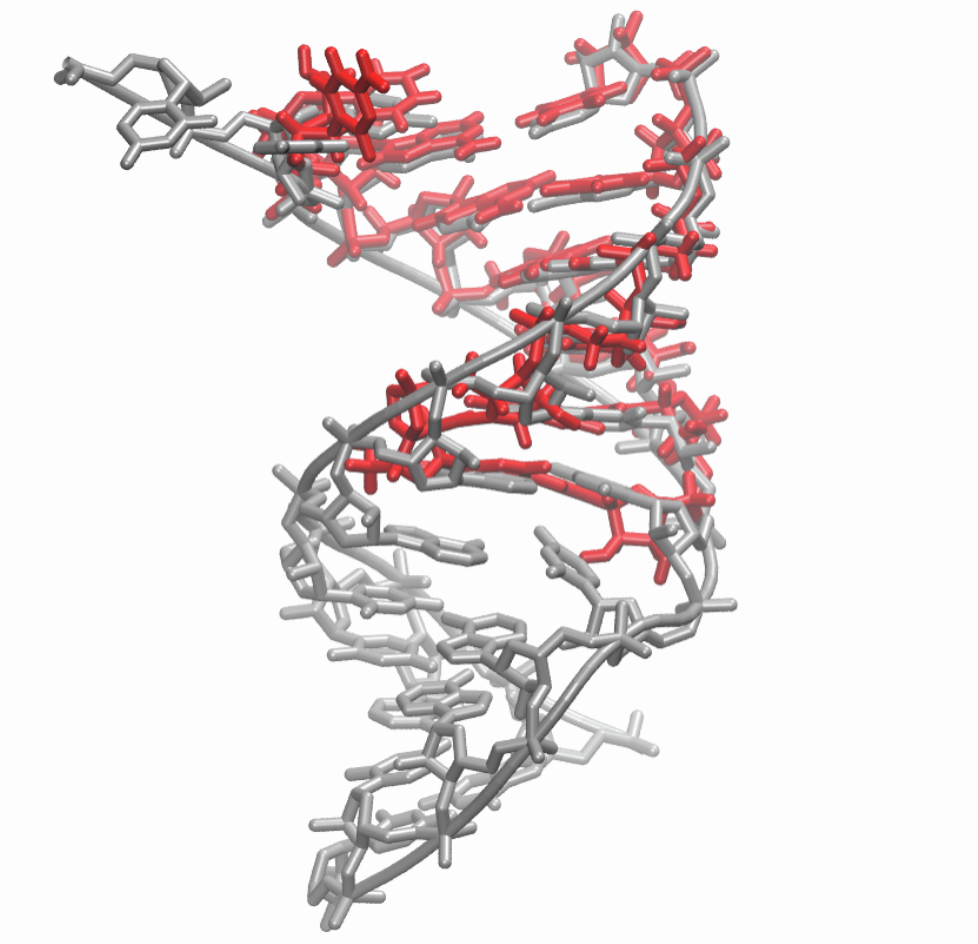


Figure S16. Alignment of RNA from cryo-EM structure with PDB ID:7bv2,[4] displayed in gray, and the double stranded RNA optimized and employed in our simulations and extracted from Hepatitis C virus, PDB ID: 4wtg,[5] displayed in red. RMSD between backbone atoms of the nucleic acids is 1.1 Å.

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.

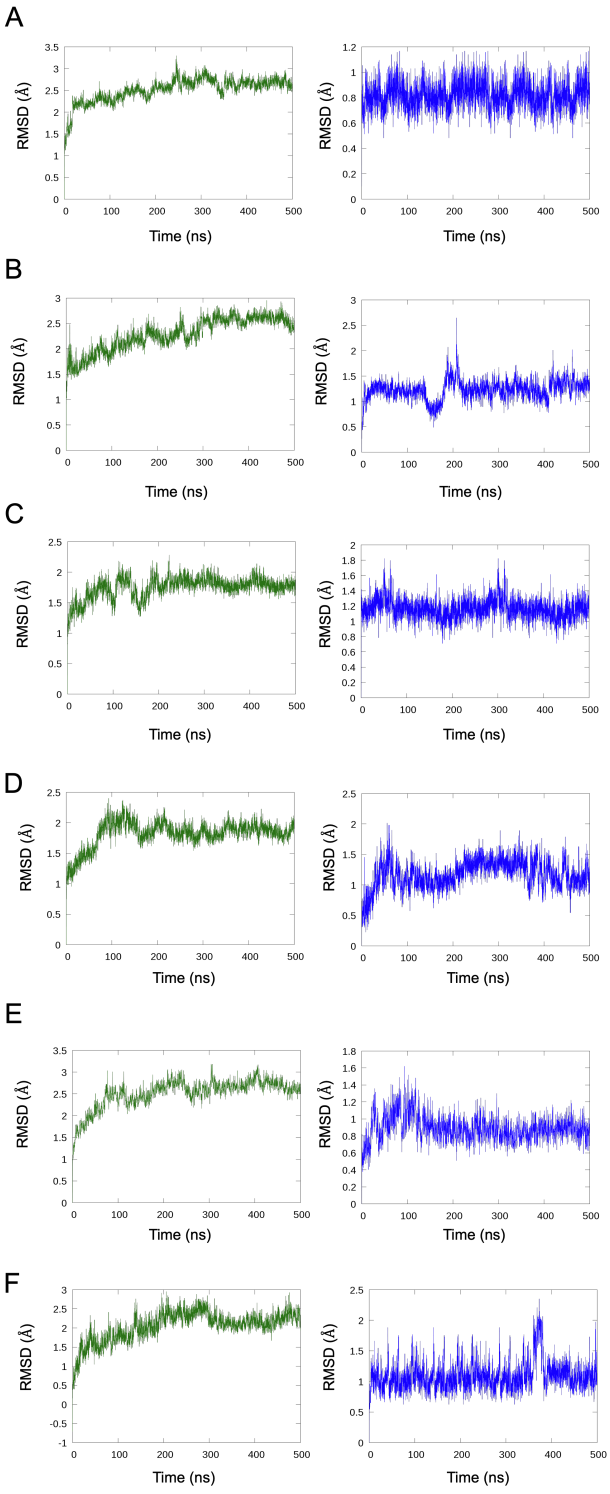


Figure S17. RMSD for the protein (green) and nucleic acid (blue) backbone atoms during MD simulations. **A** RMSDs for the system containing a ATP molecule. **B** RMSDs for the system containing a UTP molecule. **C** RMSD for the system containing an RTP molecule. **D** RMSD for the system containing 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.

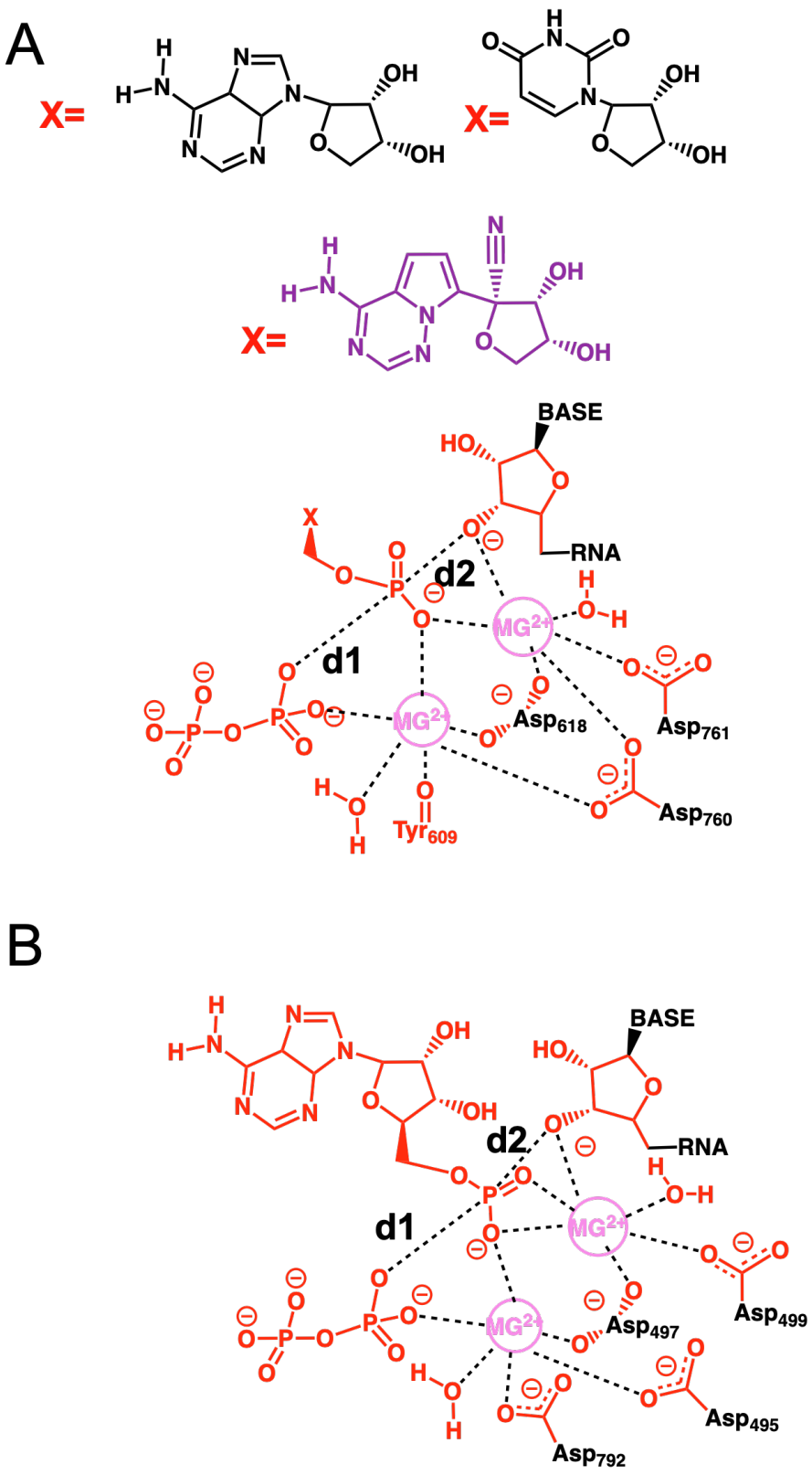


Figure S18. A Atoms described at QM level (in red and pink) in the hybrid QM/MM calculations during the ligation reaction step. Distances involved in the Reaction Coordinates employed are shown. **B** Atoms described at QM level (in red) in the hybrid QM/MM calculations for the Human RNA Pol II incorporation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