

# RNA-DEPENDENT RNA POLYMERASE FROM SARS-COV-2. MECHANISM OF REACTION AND INHIBITION BY REMDESIVIR

Juan Aranda<sup>1</sup> & Modesto Orozco<sup>1,2\*</sup>

We combine sequence analysis, molecular dynamics and hybrid quantum mechanics/molecular mechanics simulations to obtain the first description of the mechanism of reaction of SARS-CoV-2 RNA-dependent RNA polymerase (RdRp) and of the inhibition of the enzyme by Remdesivir. Despite its evolutionary youth, the enzyme is highly optimized to have good fidelity in nucleotide incorporation and a good catalytic efficiency. Our simulations strongly suggest that Remdesivir triphosphate (the active form of drug) is incorporated into the nascent RNA replacing ATP, leading to a duplex RNA which is structurally very similar to an unmodified one. We did not detect any reason to explain the inhibitory activity of Remdesivir at the active site. Displacement of the nascent Remdesivir-containing RNA duplex along the exit channel of the enzyme can occur without evident steric clashes which would justify delayed inhibition. However, after the incorporation of three more nucleotides we found a hydrated Serine which is placed in a perfect arrangement to react through a Pinner's reaction with the nitrile group of Remdesivir. Kinetic barriers for crosslinking and polymerization are similar suggesting a competition between polymerization and inhibition. Analysis of SARS-CoV-2 mutational landscape and structural analysis of polymerases across different species support the proposed mechanism and suggest that virus has not explored yet resistance to Remdesivir inhibition.

---

<sup>1</sup> Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Baldiri Reixac 10, 08028 Barcelona, Spain

<sup>2</sup> Departament de Bioquímica i Biomedicina, Universitat de Barcelona, Universitat de Barcelona, Avinguda Diagonal 645, 08028 Barcelona, Spain

\* Correspondence to M.Orozco. [modesto.orozco@irbbarcelona.org](mailto:modesto.orozco@irbbarcelona.org)

## INTRODUCTION

SARS-CoV-2 (Covid19) emerged in China in December 2019 and rapidly spread over the world, causing a world-wide health threat. At the beginning of June 2020 Covid19 pandemic has caused around 400,000 fatalities and at least 6 million people have been infected, producing an unprecedented stress to the health-care systems.<sup>1</sup> From a phylogenetic point of view SARS-CoV-2 belongs to the  $\beta$ -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).<sup>2</sup> SARS-CoV-2 has a large (30 kb of positive RNA) genome, which forces it to a high fidelity in replication and accordingly to a relatively low mutational rate.<sup>3,4</sup> A high-fidelity RNA-dependent RNA polymerase (RdRp) and a proof-reading exonuclease<sup>5</sup> are the crucial elements for maintaining the stability of the viral genome.

RdRp is the core of the replication machinery of the virus, and the largest protein in the viral genome (932 residues). It binds to nsp7 and nsp8 to form an active complex,<sup>6</sup> that using as template the sense RNA generates a negative copy, which in a second cycle is used to generate 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.<sup>7</sup> A simple BLAST<sup>8</sup> query shows that SARS-CoV-2 RdRp is highly conserved within the Coronavirus family, but no homologs are found out of it. This would suggest that we are in front of a quite new protein, which, in principle, did not have enough evolutionary time to optimize its function. Unfortunately, the extremely high infectivity of the virus suggests that despite its youth, RdRp is very effective. How the enzyme achieves a good specificity and an excellent catalytic power is still unclear, generating an important gap in our knowledge on SARS-CoV-2 replicative cycle.

Due to its central role in the viral infective cycle, RNA polymerases are a major target for fighting RNA-viruses.<sup>6,9</sup> As today, the only FDA-approved drug for treatment of SARS-CoV-2 infection<sup>10</sup> is a RdRp inhibitor Remdesivir (R), a C-nucleoside (see Supplementary Fig. 1) which was in clinical trials for the treatment of Ebola,<sup>11</sup> a negative single strand RNA virus very distant from SARS-CoV-2, but whose replication is also dependent on the action of a RdRp. The mechanism of inhibition of RdRp by Remdesivir is mostly unknown and conflicting hypothesis have emerged the last months. This lack of knowledge dramatically hampers our ability to develop new and more active derivatives.

We present here a comprehensive theoretical study on the mechanism of action of SARS-CoV-2 RdRp. MD simulations show that despite its youth the enzyme has a discriminant active site which should guarantee high replication fidelity. MD and QM/MM simulations demonstrate that the enzyme follows a canonical 2-ions reaction mechanism with a catalytic efficiency similar to that of the highly evolved eukaryotic polymerases. The same type of calculations demonstrates that Remdesivir triphosphate (RTP; the expected bioactive form of Remdesivir) can be recognized and incorporated into a nascent RNA with an efficiency only slightly lower than natural nucleotides, i.e. R 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 which would justify inhibitory properties. Furthermore, no steric clashes were detected when the nascent RNA duplex was displaced along the exit channel (see below),

which argue against the hypothesis that steric clashes could explain a delayed inhibition. However, the same analysis suggests that when three more nucleotides are incorporated to the RNA after Remdesivir, the C1' nitrile group of Remdesivir reaches the proximity to Ser<sub>861</sub> favoring an unexpected crosslinking reaction. Activation barriers of polymerization and crosslinking are comparable, raising an interesting scenario of competition between polymerization and inhibition which can explain the unique inhibitory properties of Remdesivir. In summary, our results provide then, not only the first atomistic description of the mechanism of action of SARS-CoV-2 RNA polymerase, but also a plausible explanation for the inhibitory mechanism of Remdesivir.

## METHODS

**RdRp complex set-up.** The cryo-EM structure of the RdRp (nsp12) of SARS-CoV-2 complexed to nsp8 and nsp7<sup>12</sup> cofactors was used as our starting model. We aligned it with the RdRp from Hepatitis-C virus X-ray structure<sup>13</sup> from which RNA, the two catalytic metal ions and the diphosphate group of an NTP analogue were extracted (see Supplementary Methods for details). During the preparation of this manuscript a cryo-EM structure of RdRp in complex with a full RNA strand appeared<sup>14</sup>, confirming the accuracy of our model (see Supplementary Fig. 2). Afterwards, we built a UTP or RTP molecule inside the active site in the expected orientation required for incorporation into the nascent RNA strand. The system was solvated and neutralized prior to optimization, thermalization (T=298K) and equilibration (see Supplementary Methods for details). The final equilibrated structure was the starting point for further MD simulations. 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+4 position keeping constant the reactive alignments at the i-position (see Supplementary Fig. 3). 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.

**MD simulations on RdRp.** Classical trajectories were used to refine and check the stability of complexes prior to running QM/MM simulations, as well as to determine the impact of R-incorporation in RNA. Minimization, thermalization and equilibration were performed using standard procedures as described in Supplementary Methods. Production simulations were carried out using the AMBER 18 program<sup>15</sup> and state-of-the art conditions for a total time of at least 0.5  $\mu$ s. Water molecules were described through the TIP3P<sup>16</sup> model, parameters of Magnesium ions were taken from Allner et al.,<sup>17</sup> Carlson et al.<sup>18</sup> parameters were used for triphosphate groups, PARMBSC0-chiOL3 for RNA,<sup>19–22</sup> and ff14SB<sup>23</sup> for the proteins. Parameters and charges of UTP, RTP, Remdesivir nucleotide and 3' terminal U and R were derived to be compatible with the force fields making use of the RED server.<sup>24</sup> Additional details of the simulation setups can be found in Supplementary Methods.

**QM/MM systems.** Two families of systems were generated to explore enzyme reactivity. The first was created to follow nucleotide incorporation, while the second was intended to study a potential crosslinking between Ser<sub>861</sub> and the nitrile group of RNA-incorporated Remdesivir.

QM subsystems used in reactivity calculations are shown in Supplementary Fig. 4. The QM part was described at the B3LYP/6-31G\*\* level of theory, using the link-atom to join QM and MM regions.<sup>15</sup> 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 the B3LYP/6-31G\*\*/MM level of theory. A spherical droplet was extracted to compute two-dimensional B3LYP/6-31G\*\*/MM potential energy surfaces (PES) in both forward and reverse directions. Selected reaction coordinates for the phosphoryl transfer are shown in Fig. 2 B and those for the crosslink of Ser861 with Remdesivir are shown Fig. 3 B, (see also Supplementary Fig. 4 and Supplementary Methods). The AMBER program interfaced with Terachem 1.9<sup>25,26</sup> was used for the QM/MM calculations. Electrostatic embedding was used in all hybrid calculations.

### **QM/MM Exploration of the Minimum Free Energy Paths and Potential of Mean Force.**

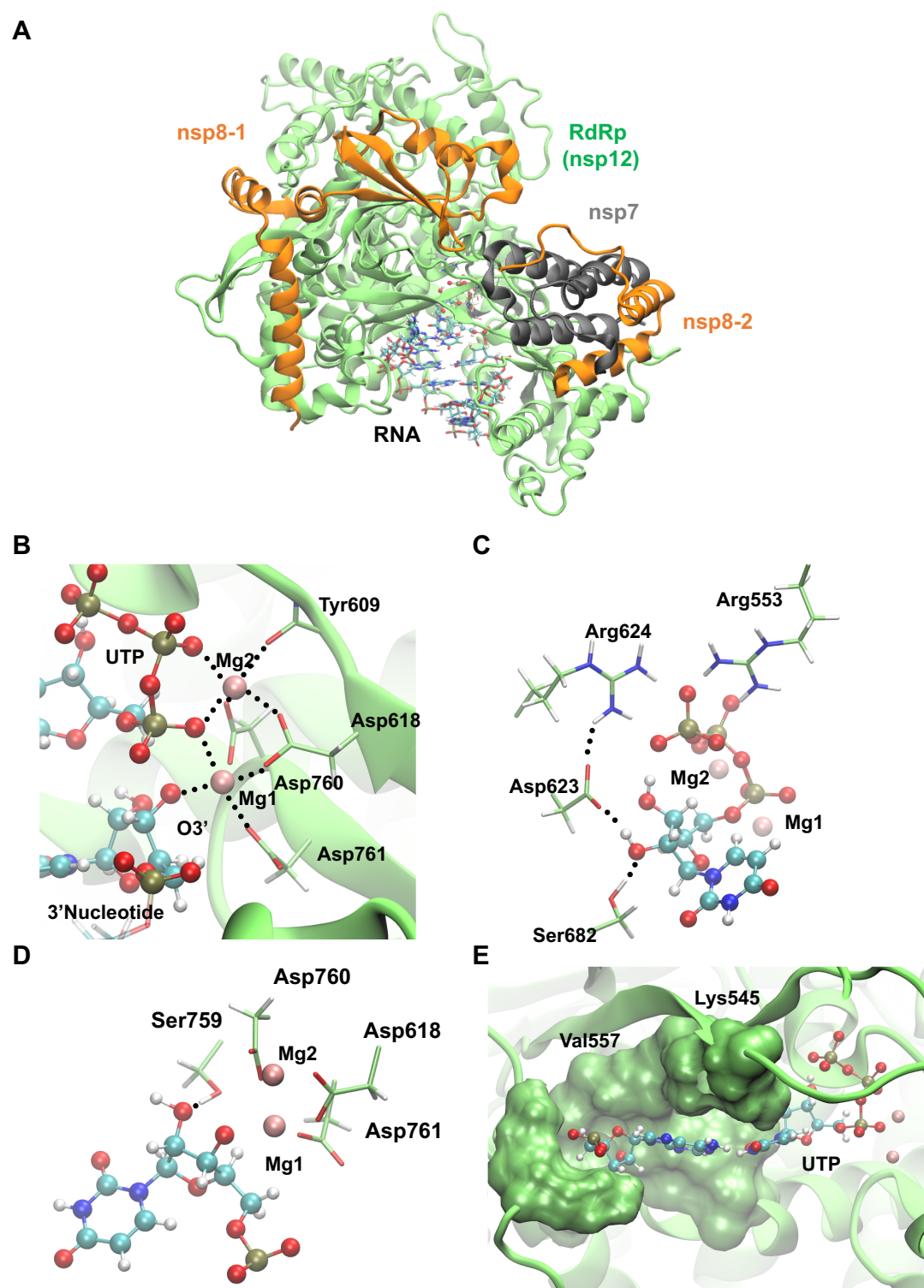
QM/MM-MD simulations were performed to obtain minimum free energy paths (MFEPs) by means of the string method.<sup>27,28</sup> This method allowed us to explore different reaction mechanisms and select the preferred one in terms of free energy. From 60 to 120 string nodes were used. Afterwards, a path collective variable (CV)<sup>28</sup> was defined to obtain the potential of mean force (PMF) using from 60 to 120 umbrella sampling<sup>29</sup> windows. The chosen set of CVs that followed the progress of the reactions, the breaking and forming bonds, are shown in Fig. 2 B, Fig. 3 B and Supplementary Methods (see Supplementary Fig. 4). MFEPs were obtained at the DFTB3<sup>30,31</sup>/MM level and the PMFs were corrected at the B3LYP/6-311++G\*\* level. The error of the free energy barriers was calculated as 95% confidence intervals and reached values within  $\pm 1$  kcal·mol<sup>-1</sup>. Each of the sampling windows consisted on 20ps of equilibration followed by 200ps of production. The AMBER program<sup>15</sup> with electrostatic embedding was used for the QM/MM calculations. Corrections at the high level of theory were made with Gaussian16<sup>32</sup> (see Supplementary Methods).

**MD simulations on nascent RNA.** We performed MD simulations on two duplexes: r(CGCGAAUUGCGC)·r(CGCGAAUUGCGC) and r(CGCGARUUGCGC)·r(CGCGAAUUGCGC) 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<sup>33,34</sup> before 1  $\mu$ s long MD trajectories at constant temperature (T=298K) and pressure (P=1 atm). Details of simulations are shown in Supplementary Methods.

**Sequence analysis.** Mutations affecting RdRp were reconstructed using the augur pipeline to infer nucleotide changes at the internal nodes<sup>35</sup> applied to 12397 GISAID samples obtained on 30/05/2020. The consequence type of the RdRp mutations was annotated using a customized implementation of the Ensembl Variant Effect Predictor (VEP version 92) using the first SARS-CoV-2 sequenced genome (NCBI ID: NC\_045512v2) as a reference. Afterwards, mutations were mapped on the MD-refined structure of the RdRp-RNA complex to locate their proximity to crucial functional regions. The aggressiveness of each individual mutation was evaluated from BLOSUM80 matrices.<sup>36</sup> Homolog search were performed with BLASTP<sup>8</sup> as implemented in the ncbi web server (<https://blast.ncbi.nlm.nih.gov/>). Alignments were performed using EMBOSS program<sup>37</sup> as implemented in the EBI webserver (<https://www.ebi.ac.uk/Tools/psa/>).

## RESULTS AND DISCUSSION

**SARS-CoV-2 RdRp active site suggests good fidelity and activity.** The viral protein (see Fig. 1 A) has a quite canonical active site similar to those of other polymerases. Two essential  $Mg^{2+}$  coordinate the  $\beta$  and  $\gamma$  phosphate groups of the incoming triphosphate nucleotide as well as Asp<sub>618</sub>, Asp<sub>760</sub>, Tyr<sub>609</sub>, and the O3' terminal of the negative RNA strand (see Fig. 1 B), which according to the circular reaction mechanism for polymerases by de Vivo and coworkers is expected to be ionized.<sup>38</sup> Two Arginine residues (Arg<sub>624</sub> and Arg<sub>553</sub>) coordinate the phosphates of the incoming nucleotide aligning the gamma phosphate for an effective transfer (see Fig. 1 C). The affinity for ribonucleotides triphosphate substrates can be explained by the need for North puckering (otherwise reactive groups are not well aligned), as well as by a myriad of specific H-bonds between the 2'OH group of the NTP and the catalytic site Asp<sub>623</sub> and Ser<sub>683</sub> side chains (see Fig. 1 C). Additional hydrogen bonds are found between i+1 2'OH and Ser<sub>759</sub> (see Fig. 1 D). The base-specificity is controlled by the complementarity of hydrogen bonding with the template nucleobase (see Fig. 1 E, Supplementary Fig. 5), by the phosphate coordination, as well as by the residues surrounding the active site that embrace the pairs introducing strong isosteric requirements that make very unlikely a non-Watson-Crick pairing scheme (see Fig. 1 E). Overall the active site emerging from EM structures<sup>12,14,39,40</sup> and atomistic simulation strongly suggest that despite the reduced evolutionary refinement, SARS-CoV-2 RdRp has all the structural requirements to be an efficient RNA polymerase, both in terms of catalysis and substrate specificity.

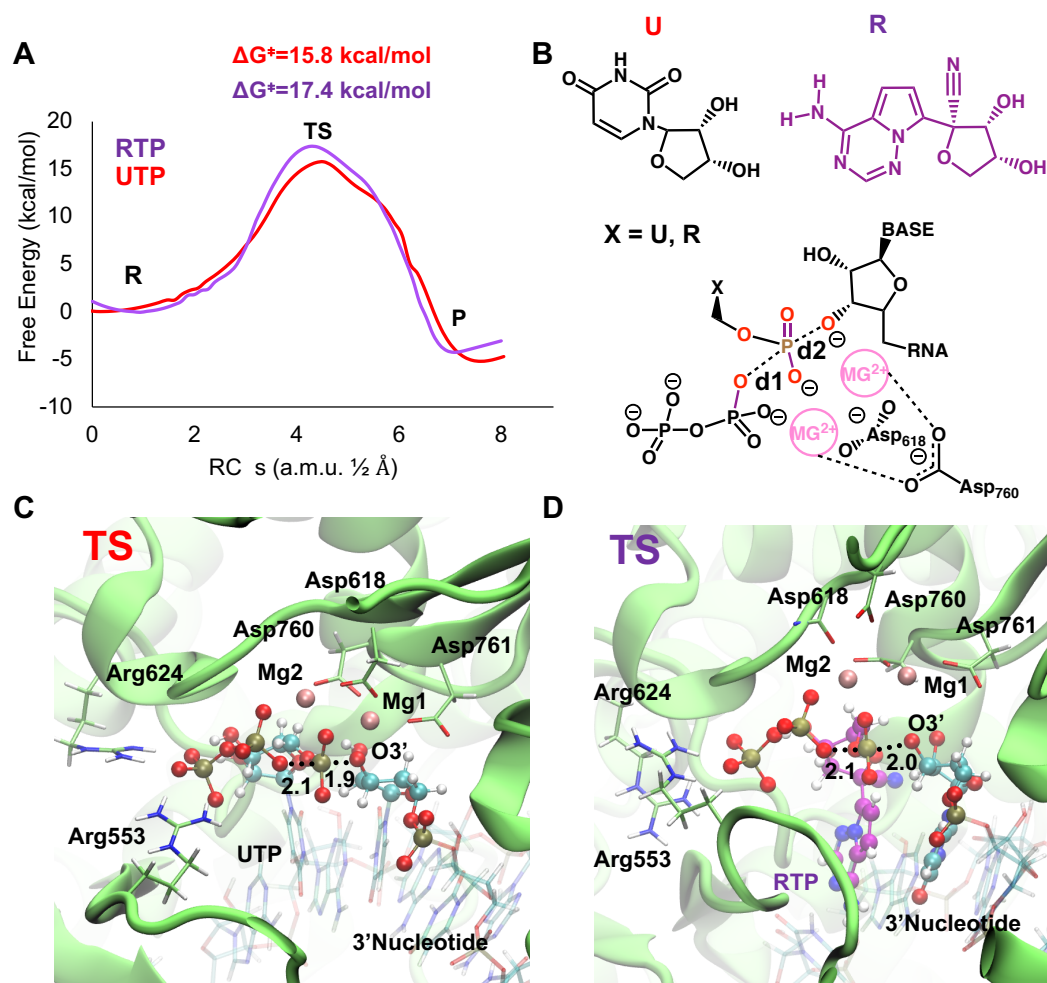


**Fig. 1 Active site of RdRp of SARS-CoV2 is an efficient polymerase** **A** Replication complex of SARS-CoV-2 used in this study is formed by the nsp12 RdRp enzyme, the nsp8 and nsp7 cofactors and RNA template and nascent strands in solution. **B** Active site insight showing the well-defined two metal ions coordination sphere. **C** Substrate recognition by active site residues of RdRp. **D** Interactions between the 3' terminal nucleotide and RdRp. **E** RdRp pocket enables base specificity between the incoming nucleotide and the template.



**The RNA-pol reaction mechanism.** QM/MM simulations were used (see Methods and Supplementary Methods) to study the ability of SARS-CoV-2 RdRp to incorporate a natural triphosphate (NTP, exemplified here by UTP), as well as RTP (Remdesivir-TP) into a nascent RNA. Careful analysis of trajectories and of the free energy profiles shows that the calculation is well converged leading to a smooth free energy profile with a single maximum (Transition State, TS), a free energy of activation of  $15.8 \text{ kcal}\cdot\text{mol}^{-1}$  for UTP (see Fig. 2, Supplementary Fig. 6 and Supplementary Video 1) and a negative free energy of reaction (around  $5 \text{ kcal/mol}$  prior to PPi release). Very interestingly, kinetic parameters for RdRp compare well to those associated to prokaryotic or eukaryotic polymerases,<sup>41–43</sup> which combined with the similarity in the reaction mechanism allow us to conclude that the enzyme can be quite novel in evolutionary terms, but it is as efficient as the highly evolved eukaryotic polymerases.

Trajectories show that RTP fits very well into the active site showing strong canonical Watson Crick interactions with the uridine in the template RNA. Its isosterism with adenosine allows a perfect shape complementarity at the binding site showing an arrangement of reactive groups that predicts that it will be a substrate rather than an inhibitor. This hypothesis is confirmed by accurate QM/MM simulation that show that incorporation of RTP can happen with a free energy barrier only slightly larger than that of a natural substrate (10% up to  $17.4 \text{ kcal}\cdot\text{mol}^{-1}$ ; see Fig. 2, Supplementary Fig. 7 and Supplementary Video 2). Such an increase is due mostly to a slight misalignment of the  $\text{O}\alpha\text{-P}\alpha\text{-O3'}$  attack angle ( $161 \pm 8^\circ$  for RTP and  $172 \pm 5^\circ$  for UTP; see also Supplementary Fig. 6-8). Thus, in agreement with recent EM experiments<sup>14</sup> 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.

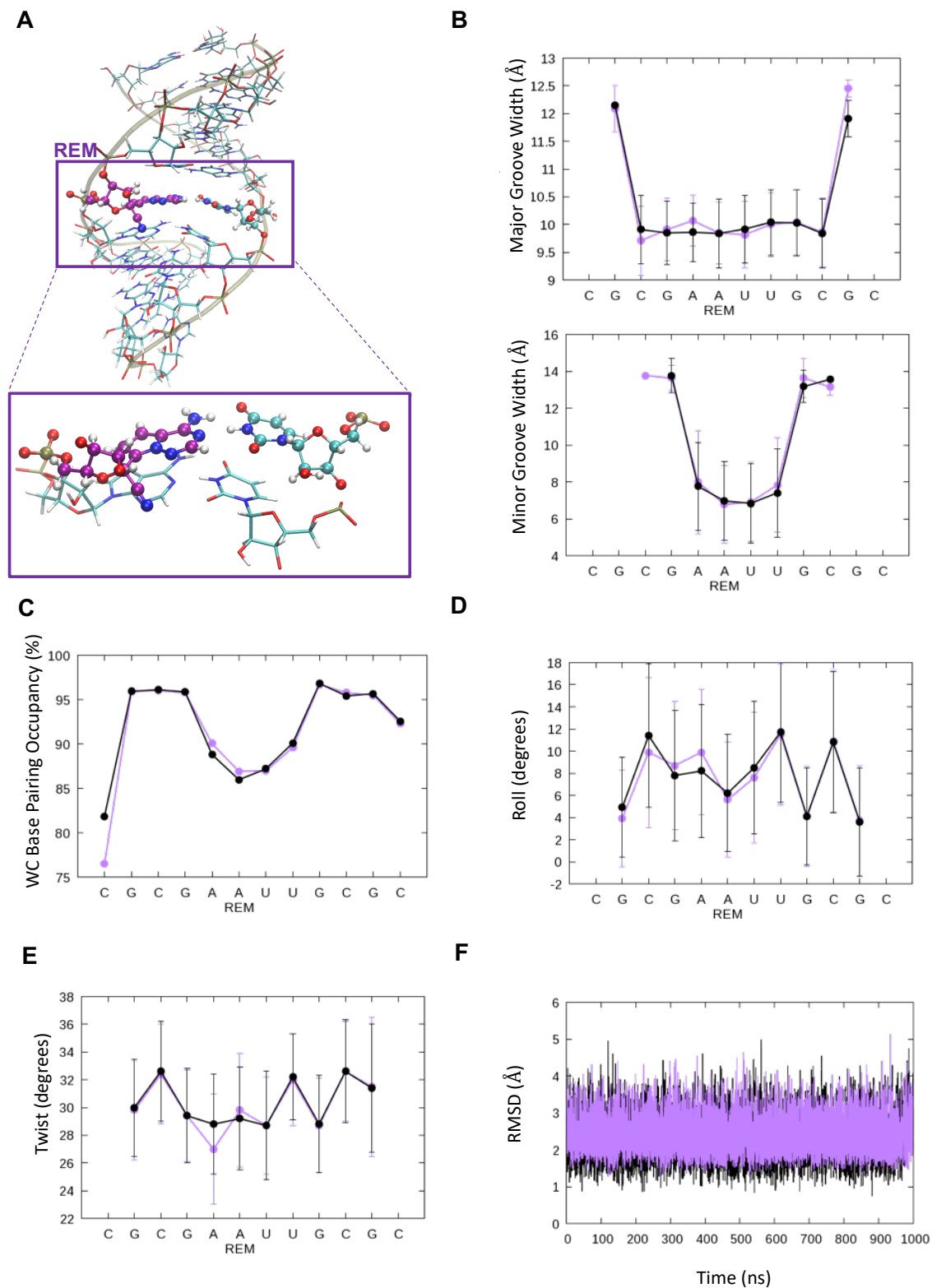


**Fig. 2 Viral RNA elongation inside RdRp of SARS-CoV-2.** **A** Free energy profiles for the incorporation of a U or a R to a nascent viral RNA strand. The reaction consists on a nucleophilic attack of the O3' of terminal nucleotide to the P $\alpha$  of the triphosphate nucleotide. **B** Schemes representing the transition states found for UTP or RTP substrates. Distances included in the definition of the Reaction Coordinate (s) are shown. **C** Active site insight of the TS found when UTP is the substrate for the elongation reaction. The phosphoryl group is half-way to be transferred. Mg1 activates the O3' towards the nucleophilic attack and stabilizes the negatively charged TS. Mg2 stabilizes the charged TS, and the newborn negatively charged PPi molecule. Distances involved in the reaction are shown as dotted lines with its average value in Å. **D** Active site insight of the TS found when RTP is incorporated (shown as ball and sticks with its C atoms shown in purple).

**Remdesivir is well tolerated in an RNA duplex.** Remdesivir is a C-nucleoside that may produce distortions in the helix which might result in a delayed inhibition of the enzyme due to incorrect 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 Methods). Results summarized in Fig. 3 strongly suggest that Remdesivir is well tolerated in an RNA duplex and does not introduce any major structural distortion which could justify termination of RNA synthesis. Particularly, there are not significant differences between the hydrogen bonding stability of r(A·U) and r(R·U) pairs (see



Fig. 3 C). With the required cautions derived from the well-known uncertainties of RNA force-field and of the limited sampling (1  $\mu$ s), our results argue against the idea that a dramatic structural alteration of the RNA duplex is the key responsible for Remdesivir-induced termination of the RNA synthesis.



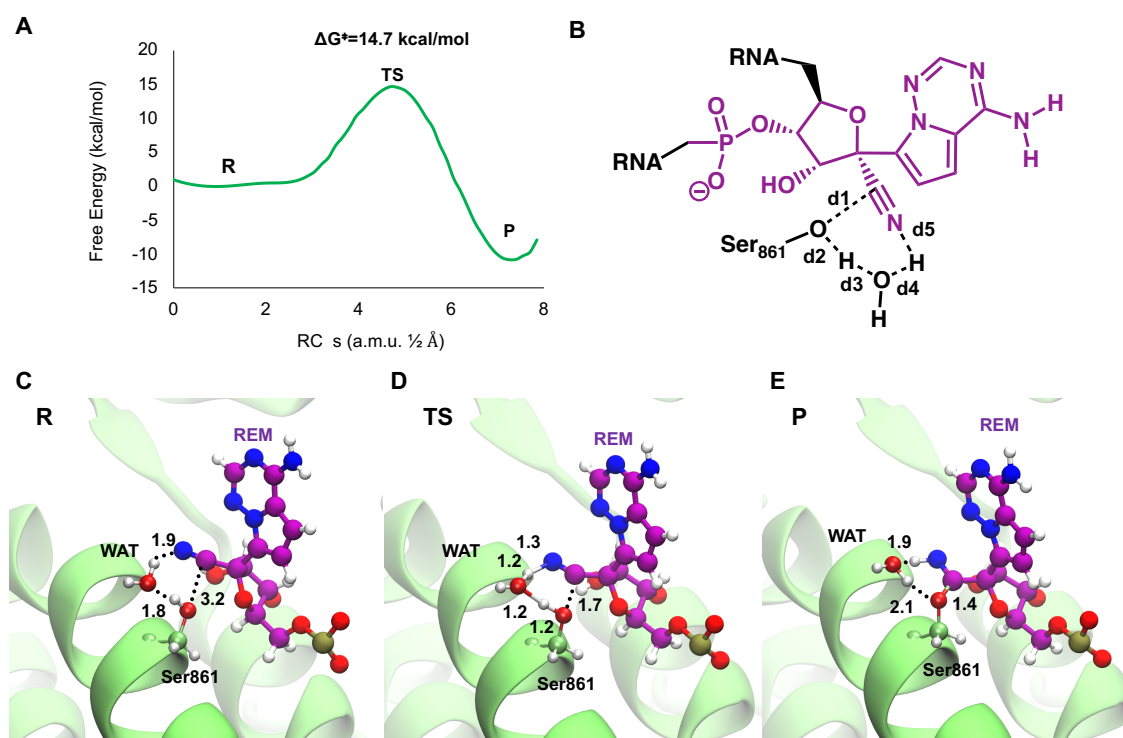
**Fig. 3 Remdesivir does not distort RNA structure.** **A** Overall view of an RNA duplex containing a Remdesivir nucleotide (shown as ball and sticks with its C atoms shown in purple). An insight is shown. Major and minor groove width **B**, WC base pairing **C**, Roll **D**, Twist **E**, and RMSD of the double helix **F** are not affected when Remdesivir is present (purple) in respect to control sequence (black).

**Remdesivir can crosslink RNA and RdRp.** Trying to explore alternative reasons for the inhibitory properties of Remdesivir we slide the nascent RNA duplex along the exit tunnel of RdRp to make the r(R-U) pair by simulating addition of extra nucleotides. This allows 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 Supplementary Fig. 9). Interestingly, when three more nucleotides were incorporated, we found a Ser<sub>861</sub> whose side chain locates at around  $3.7 \pm 0.3$  Å from the nitrile group of Remdesivir (see Supplementary Fig. 10 A B). The distance does not justify however a steric clash considering the large flexibility of the Ser sidechain, especially in a well-solvated microenvironment (see Supplementary Fig. 10 A), arguing against the hypothesis that nitrile-Ser<sub>861</sub> steric clash can explain the inhibitory properties of Remdesivir.<sup>10</sup> However, a close look to the trajectories shows that the Ser<sub>861</sub> side chain is frequently pointing towards the nitrile group of Remdesivir in an arrangement suitable for a nucleophilic attack and in the presence of water that spontaneously enters into the cavity (see Supplementary Fig. 10 C D). In summary, it is placed in an optimal position as to act as a general acid/base catalyst. Note that, while not suggested for this protein, attacks of serine and cysteine to nitrile groups have been characterized for other proteins,<sup>44,45</sup> which drives us to explore in more detail this potential reaction mechanism by using a QM/MM strategy similar to that used to describe the polymerization reaction (see Methods). Results (see Fig. 4, QM/MM PES in Supplementary Fig. 11, and Supplementary Videos 3 and 4) show a concerted reaction mechanism where, while hydroxylic group of Ser<sub>861</sub> attacks the carbon atom of the nitrile group, a water molecule (see Supplementary Fig. 10 D) abstracts one proton from the hydroxyl group of Ser<sub>861</sub>, donating it to the N atom of nitrile group (see Fig. 4 D, Supplementary Video 3). The structures sampled along the minimum free energy profile agree with the electronic flow for a Pinner reaction.<sup>46</sup> Our calculations suggest an exergonic ( $-10.7$  kcal·mol<sup>-1</sup>; see Fig. 4 A) process showing a single TS (as expected from a concerted mechanism; see Figure 4 and Supplementary Fig. 11) with an activation barrier of  $14.7$  kcal·mol<sup>-1</sup>; see Fig. 3 A), which compare with the activation barrier of polymerization of a UTP ( $15.8$  kcal·mol<sup>-1</sup>) and RTP ( $17.4$  kcal·mol<sup>-1</sup>).

QM/MM calculations suggest the existence of a kinetic competition between UTP incorporation and crosslinking leading to enzyme inhibition. In principle, barrier for UTP incorporation is slightly larger than that of crosslinking, which suggests that the RNA will be kinetically trapped stopping the extension of the RNA, and inhibiting then the activity of the enzyme. However, we can expect residual polymerization leading to longer transcripts, as experimentally observed.<sup>9</sup> A larger concentration of Remdesivir might favor the inhibition by a double mechanism: by increasing the probability of RTP incorporation into nascent RNA and also by enlarging the gap in relative activation free energy between polymerization and

crosslinking (from 1.1 kcal·mol<sup>-1</sup> to 2.7 kcal·mol<sup>-1</sup>), and driving then the kinetic balance towards the inhibition side.

The proposed covalent inhibition mechanism is consistent with the incorporation of three more nucleotides after incorporation of Remdesivir to RNA<sup>9</sup> and agrees with a myriad of indirect evidences other than the delay inhibitory properties of Remdesivir. For example, mutational analysis shows that Ser<sub>861</sub> has a crucial role in R-induced inhibition of RdRp.<sup>40</sup> Furthermore, large and systematic efforts to substitute the nitrile group by less problematic substituents (the nitrile group appears as a “probably undesired” consequence of the Remdesivir synthetic pathway) resulted in molecules with a lower inhibitory profile.<sup>47</sup> Finally, Pinner reaction has not been described for RdRp, but has been experimentally validated for many several nitrile-containing drugs designed to inhibit cysteine or serine proteases.<sup>48,49</sup> Altogether, results strongly suggest that Remdesivir inhibits SARS-CoV-2 by a covalent mechanism.



**Fig. 4 Remdesivir forms a covalent complex with RdRp of SARS-CoV-2.** **A** Free energy profile for the attack of Ser<sub>861</sub> to the C atom of nitrile group of Remdesivir (see the bidimensional PES in Supplementary Fig. 11). **B** Scheme representing the concerted TS where nucleophilic attack of Ser<sub>861</sub> occurs at the same time that a water mediated proton transfer to the nitrile group of Remdesivir. Distances included in the definition of the Reaction Coordinate (s) are shown. **C** Active site insight of Reactant state. Distances involved in the reaction are shown as dotted lines with its average value in Å. **D** Active site insight of the the TS. **E** Active site insight of the covalent complex formed by Remdesivir and RdRp.

**Mutational analysis of SARS-CoV-2 RNA-pol agrees with the proposed reaction mechanisms.**

Thousands of SARS-CoV-2 viruses have been sequenced, showing that RdRp accumulates a non-negligible number of mutations, which lead to viable virus, i.e., to active RdRp. We should then check whether some of these mutations invalidate our reaction mechanism. As shown in Supplementary Fig. 12, this is not the case, missense variant leading to amino-acid changes are typically mild according to BLOSUM80 matrices and are located in loops on the surface of the protein (Supplementary Fig. 12 A). Particularly, the active site and the exit tunnel are very well conserved during the viral evolution (see Supplementary Fig. 12 B, C). Moreover, virus has not explored yet mutational landscape in the region of the Ser<sub>861</sub> (see Supplementary Fig. 12 D).

**Remdesivir is not likely to affect human RNA polymerases.** Remdesivir enters in human cells and is transformed into RTP also by human enzymes. Considering the similarity in the mechanism of reaction of SARS-CoV-2 RdRp and human polymerases we could expect Remdesivir being incorporated into RNA, which could rise to inhibition of the human enzyme with toxic consequences. To check this possibility we explored the extension of the nascent RNA along the exit channel to trace close interactions between the nitrile group and serine or cysteine residues in the two crystallized human RNA polymerases (RNA polymerase<sup>50</sup> and human RNA polymerase II<sup>51</sup>). We did not find any of these residues in the expected displacement path of Remdesivir once incorporated into the nascent RNA (see Supplementary Fig. 13). In other words, no significant inhibition of the human polymerases is expected. Note that this explains the reduced toxicity of Remdesivir in humans.<sup>52,53</sup>

**Inhibition mechanism can be common to other viruses.** Ser<sub>861</sub> is highly conserved in other CoVs RdRps (see Supplementary Fig. 14), placing a constant position in an alpha helix that adopts a highly conserved 3D arrangement (compare SARS-CoV-2 and SARS-CoV structures)<sup>12,14,39,40,54</sup> and this position is well preserved in those cases for which CoV RdRps structure is available. This would suggest that Remdesivir might be effective against many other CoV RdRps. The mechanism of inhibition of other distant RNA-viruses (Marburg virus, Ebola, Hepatitis C and many others) might be similar, but needs to be elucidated. It is tempting to believe that enzymes evolved to discriminate between DNA (2'deoxyribose) and RNA (ribose) might be decorated with a large number of serine, tyrosine or threonine in the exit channel, suggesting a general delay inhibitory mechanism like that suggested here. Note that this would agree with the delayed inhibition of Ebola RdRp by Remdesivir.<sup>55</sup> The lack of structural information and the poor sequence similarity between the RdRps precludes to confirm this hypothesis.

## CONCLUSIONS

SARS-CoV-2 RdRp is a protein common to other coronaviruses, but showing little homology out of 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, our calculations show that it has a well refined reaction site able to select the entering nucleotide and to catalyze its addition to a nascent RNA. 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<sup>2+</sup>

ions exquisitely coordinated by acidic residues of the catalytic site, while the phosphates of the incoming nucleotide being stabilized by a network of basic residues.

Our simulations very clearly suggest that RTP is not inhibiting the active site by competing with ATP at the catalytic site, but that 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 fail to detect points of steric clashes, but show the spontaneous formation of a catalytical arrangement that justifies a Pinner reaction (between the nitrile group of Remdesivir and Ser<sub>681</sub>), leading to the formation of a protein-RNA covalent bond. However, we cannot determine whether this inhibition will be reversible or not, as inspection of the free energy curve (Fig. 4D) and the behavior of similar inhibitors<sup>49</sup> for other proteins suggest that it may display slow reversibility.<sup>49,56</sup> This potential reversibility combined with the kinetic competition between crosslinking and polymerization might explain the unusual inhibitory properties of Remdesivir. Results presented here open then the possibility to design better inhibitors of CoVs RdRps and to a rethinking on the use of covalent inhibitors of pathological proteins.

## REFERENCES

1. WHO. Novel Coronavirus (COVID-19) Situation. *WHO (June 1)* (2020).
2. De Wit, E., Van Doremalen, N., Falzarano, D. & Munster, V. J. SARS and MERS: Recent insights into emerging coronaviruses. *Nature Reviews Microbiology* (2016). doi:10.1038/nrmicro.2016.81
3. Hadfield, J. *et al.* Genomic epidemiology of novel coronavirus - Global subsampling. *Nextstrain: real-time tracking of pathogen evolution* (2020).
4. Lu, R. *et al.* Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* (2020). doi:10.1016/S0140-6736(20)30251-8
5. Ferron, F. *et al.* Structural and molecular basis of mismatch correction and ribavirin excision from coronavirus RNA. *Proc. Natl. Acad. Sci. U. S. A.* (2017). doi:10.1073/pnas.1718806115
6. Gordon, C. J., Tchesnokov, E. P., Feng, J. Y., Porter, D. P. & Götte, M. The antiviral compound remdesivir potently inhibits RNA-dependent RNA polymerase from Middle East respiratory syndrome coronavirus. *Journal of Biological Chemistry* (2020). doi:10.1074/jbc.AC120.013056
7. Romano, M., Ruggiero, A., Squeglia, F., Maga, G. & Berisio, R. A Structural View of SARS-CoV-2 RNA Replication Machinery: RNA Synthesis, Proofreading and Final Capping. *Cells* (2020). doi:10.3390/cells9051267
8. McGinnis, S. & Madden, T. L. BLAST: At the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res.* (2004). doi:10.1093/nar/gkh435
9. Gordon, C. J. *et al.* Remdesivir is a direct-acting antiviral that inhibits RNA-dependent RNA polymerase from severe acute respiratory syndrome coronavirus 2 with high potency. *J. Biol. Chem.* (2020). doi:10.1074/jbc.ra120.013679
10. Expanded Access Treatment Protocol: Remdesivir (RDV; GS-5734) for the Treatment of SARS-CoV2 (CoV) Infection (NCT04323761). *U.S. National Library of Medicine* (2020).
11. Investigational Therapeutics for the Treatment of People With Ebola Virus Disease. <https://clinicaltrials.gov/show/nct03719586> (2018).
12. Gao, Y. *et al.* Structure of the RNA-dependent RNA polymerase from COVID-19 virus. *Science* (80-. ). (2020). doi:10.1126/science.abb7498
13. Appleby, T. C. *et al.* Structural basis for RNA replication by the hepatitis C virus polymerase. *Science* (80-. ). (2015). doi:10.1126/science.1259210
14. Yin, W. *et al.* Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir. *Science* (80-. ). (2020). doi:10.1126/science.abc1560
15. Case, D. A. *et al.* *Amber 18 University of California, San Francisco.* (2018).
16. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L.



Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926 (1983).

17. Allnér, O., Nilsson, L. & Villa, A. Magnesium ion-water coordination and exchange in biomolecular simulations. *J. Chem. Theory Comput.* **8**, 1493–1502 (2012).
18. Meagher, K. L., Redman, L. T. & Carlson, H. A. Development of polyphosphate parameters for use with the AMBER force field. *J. Comput. Chem.* **24**, 1016–1025 (2003).
19. Cheatham, T. E., Cieplak, P. & Kollman, P. A. 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 (1999).
20. Wang, J., Cieplak, P. & Kollman, P. A. 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 (2000).
21. Pérez, A. *et al.* Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of  $\alpha/\gamma$  Conformers. *Biophys. J.* **92**, 3817–3829 (2007).
22. Zgarbová, M. *et al.* 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 (2011).
23. Maier, J. A. *et al.* ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* (2015). doi:10.1021/acs.jctc.5b00255
24. Vanquelef, E. *et al.* 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.* (2011). doi:10.1093/nar/gkr288
25. *Petachem*. <http://www.petachem.com>. (accessed June 1, 2020).
26. Götz, A. W., Clark, M. A. & Walker, R. C. An extensible interface for QM/MM molecular dynamics simulations with AMBER. *J. Comput. Chem.* **35**, 95–108 (2014).
27. Vanden-Eijnden, E. & Venturoli, M. Revisiting the finite temperature string method for the calculation of reaction tubes and free energies. *J. Chem. Phys.* (2009). doi:10.1063/1.3130083
28. Zinovjev, K. & Tuñón, I. Adaptive Finite Temperature String Method in Collective Variables. *J. Phys. Chem. A* (2017). doi:10.1021/acs.jpca.7b10842
29. Torrie, G. M. & Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* (1977). doi:10.1016/0021-9991(77)90121-8
30. Gaus, M., Cui, Q. & Elstner, M. DFTB3: Extension of the self-consistent-charge density-functional tight-binding method (SCC-DFTB). *J. Chem. Theory Comput.* (2011). doi:10.1021/ct100684s
31. Gaus, M., Lu, X., Elstner, M. & Cui, Q. Parameterization of DFTB3/3OB for

sulfur and phosphorus for chemical and biological applications. *J. Chem. Theory Comput.* (2014). doi:10.1021/ct401002w

32. 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. Gaussian 16, Rev. B.01. *Gaussian, Inc., Wallingford, CT* (2016).
33. Dans, P. D., Walther, J., Gómez, H. & Orozco, M. Multiscale simulation of DNA. *Current Opinion in Structural Biology* (2016). doi:10.1016/j.sbi.2015.11.011
34. Dans, P. D. *et al.* How accurate are accurate force-fields for B-DNA? *Nucleic Acids Res.* (2017). doi:10.1093/nar/gkw1355
35. Hadfield, J. *et al.* NextStrain: Real-time tracking of pathogen evolution. *Bioinformatics* (2018). doi:10.1093/bioinformatics/bty407
36. Henikoff, S. & Henikoff, J. G. Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. U. S. A.* (1992). doi:10.1073/pnas.89.22.10915
37. Rice, P., Longden, L. & Bleasby, A. EMBOSS: The European Molecular Biology Open Software Suite. *Trends in Genetics* (2000). doi:10.1016/S0168-9525(00)02024-2
38. Genna, V., Vidossich, P., Ippoliti, E., Carloni, P. & Vivo, M. De. A Self-Activated Mechanism for Nucleic Acid Polymerization Catalyzed by DNA/RNA Polymerases. *J. Am. Chem. Soc.* (2016). doi:10.1021/jacs.6b05475
39. Hillen, H. S. *et al.* Structure of replicating SARS-CoV-2 polymerase. *Nature* (2020). doi:10.1038/s41586-020-2368-8
40. Wang, Q. *et al.* Structural basis for RNA replication by the SARS-CoV-2 polymerase. *Cell* (2020). doi:https://doi.org/10.1016/j.cell.2020.05.034
41. Berdis, A. J. Mechanisms of DNA polymerases. *Chem. Rev.* **109**, 2862–2879 (2009).
42. Castro, C. *et al.* Nucleic acid polymerases use a general acid for nucleotidyl transfer. *Nat. Struct. Mol. Biol.* (2009). doi:10.1038/nsmb.1540
43. Wu, W. J., Yang, W. & Tsai, M. D. How DNA polymerases catalyse replication and repair with contrasting fidelity. *Nature Reviews Chemistry* (2017). doi:10.1038/s41570-017-0068
44. Nabeno, M. *et al.* A comparative study of the binding modes of recently launched dipeptidyl peptidase IV inhibitors in the active site. *Biochem. Biophys. Res. Commun.* (2013). doi:10.1016/j.bbrc.2013.03.010
45. Law, S. *et al.* Identification of mouse cathepsin K structural elements that regulate the potency of odanacatib. *Biochem. J.* (2017). doi:10.1042/BCJ20160985
46. Roger, R. & Neilson, D. G. The chemistry of imidates. *Chem. Rev.* (1961). doi:10.1021/cr60210a003
47. Cho, A. *et al.* Synthesis and antiviral activity of a series of 1'-substituted 4-aza-

- 7,9-dideazaadenosine C-nucleosides. *Bioorganic Med. Chem. Lett.* (2012). doi:10.1016/j.bmcl.2012.02.105
48. Turk, B. Targeting proteases: Successes, failures and future prospects. *Nature Reviews Drug Discovery* (2006). doi:10.1038/nrd2092
  49. Kalgutkar, A. S. & Dalvie, D. K. Drug discovery for a new generation of covalent drugs. *Expert Opinion on Drug Discovery* (2012). doi:10.1517/17460441.2012.688744
  50. Schwinghammer, K. *et al.* Structure of human mitochondrial RNA polymerase elongation complex. *Nat. Struct. Mol. Biol.* (2013). doi:10.1038/nsmb.2683
  51. Bernecky, C., Herzog, F., Baumeister, W., Plitzko, J. M. & Cramer, P. Structure of transcribing mammalian RNA polymerase II. *Nature* (2016). doi:10.1038/nature16482
  52. Warren, T. K. *et al.* Therapeutic efficacy of the small molecule GS-5734 against Ebola virus in rhesus monkeys. *Nature* (2016). doi:10.1038/nature17180
  53. Sheahan, T. P. *et al.* Broad-spectrum antiviral GS-5734 inhibits both epidemic and zoonotic coronaviruses. *Sci. Transl. Med.* (2017). doi:10.1126/scitranslmed.aal3653
  54. Kirchdoerfer, R. N. & Ward, A. B. Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. *Nat. Commun.* (2019). doi:10.1038/s41467-019-10280-3
  55. Tchesnokov, E. P., Feng, J. Y., Porter, D. P. & Götte, M. Mechanism of inhibition of ebola virus RNA-dependent RNA polymerase by remdesivir. *Viruses* (2019). doi:10.3390/v11040326
  56. Devkota, A. K. *et al.* Reversible covalent inhibition of eEF-2K by carbonitriles. *ChemBioChem* (2014). doi:10.1002/cbic.201402321

## **Acknowledgments**

This work has been supported by the Spanish Ministry of Science (BFU2014-61670-EXP), the Catalan SGR, the Instituto Nacional de Bioinformática, and 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 MareNostrum supercomputer and support team at IRB Barcelona. 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.

## **Additional information**

**Competing financial interests:** The authors declare no competing financial interests.