

# Modeling the Enzymatic Mechanism of the SARS-CoV-2 RNA-Dependent RNA Polymerase by DFT/MM-MD: An Unusual Active Site Leading to High Replication Rates

Emmanuelle Bignon\* and Antonio Monari\*



Cite This: *J. Chem. Inf. Model.* 2022, 62, 4261–4269



Read Online

ACCESS |



Metrics & More

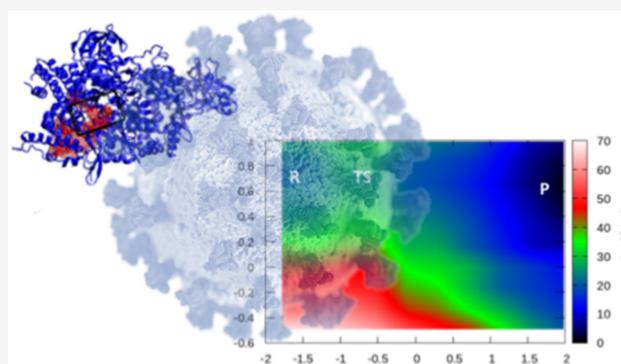


Article Recommendations



Supporting Information

**ABSTRACT:** Viral infection relies on the hijacking of cellular machineries to enforce the reproduction of the infecting virus and its subsequent diffusion. In this context, the replication of the viral genome is a key step performed by specific enzymes, i.e., polymerases. The replication of SARS-CoV-2, the causative agent of the COVID-19 pandemics, is based on the duplication of its RNA genome, an action performed by the viral RNA-dependent RNA polymerase. In this contribution, by using highly demanding DFT/MM-MD computations coupled to 2D-umbrella sampling techniques, we have determined the chemical mechanisms leading to the inclusion of a nucleotide in the nascent viral RNA strand. These results highlight the high efficiency of the polymerase, which lowers the activation free energy to less than 10 kcal/mol. Furthermore, the SARS-CoV-2 polymerase active site is slightly different from those usually found in other similar enzymes, and in particular, it lacks the possibility to enforce a proton shuttle via a nearby histidine. Our simulations show that this absence is partially compensated by lysine whose proton assists the reaction, opening up an alternative, but highly efficient, reactive channel. Our results present the first mechanistic resolution of SARS-CoV-2 genome replication at the DFT/MM-MD level and shed light on its unusual enzymatic reactivity paving the way for the future rational design of antivirals targeting emerging RNA viruses.



## INTRODUCTION

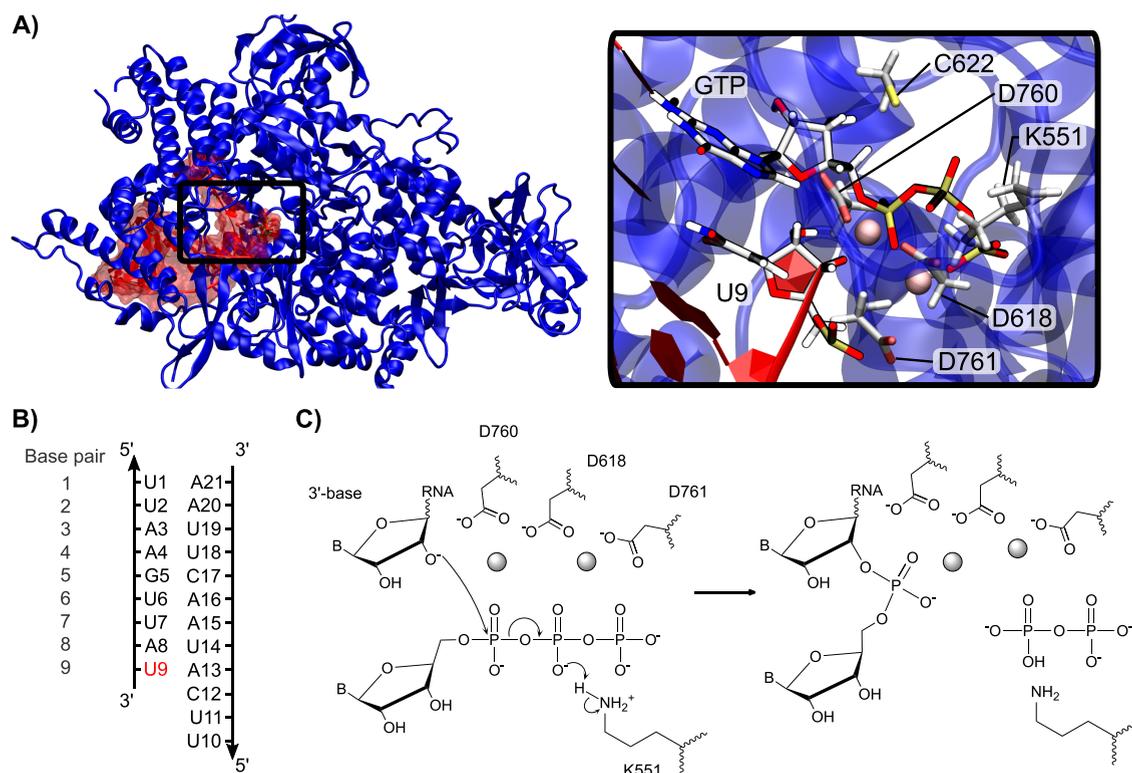
The emergence of a new  $\beta$ -coronavirus, reported for the first time at the end of 2019 in Wuhan, China, has led to the sudden outbreak of the COVID-19 pandemic.<sup>1–5</sup> By the beginning of 2022, COVID-19 has affected almost all of the continents and has pushed public governments to implement severe containment and social distancing measures, seriously affecting economic and social life. The novel  $\beta$ -coronavirus has been recognized as the causative agent of a severe acute respiratory stress (SARS) and is thus named SARS-CoV-2, to indicate its similarity with SARS-CoV whose outbreak is dated back to 2003. SARS-CoV-2 also shares a structural and phylogenetic analogy with the Middle East Respiratory Syndrome Coronavirus (MERS-CoV) that emerged in 2012. Differently from its predecessors, SARS-CoV-2 has a lower mortality ratio (estimated at around 2–3%),<sup>6–8</sup> and in some cases, it develops asymptotically, especially for its latest B.1.1.519 (omicron) variant.<sup>9–11</sup> However, its high contagiousness and the possible development of severe outcomes, especially in elder patients or subjects presenting comorbidities, have caused and still cause an impressive stress on health systems, only mitigated by the development of novel vaccines also including the mRNA strategy.<sup>12–14</sup>

As all  $\beta$ -coronaviruses, SARS-CoV-2 is an enveloped, rather large, positive-sense RNA virus.<sup>15</sup> As such, its genetic material is contained inside a virion protected by a lipid membrane and is constituted by a positive 5' to 3' single-stranded RNA consisting of about 30 000 nucleobases.<sup>16</sup> Upon infection, the viral genetic material is introduced into a cell by endocytosis or direct membrane fusion where it hijacks the cellular machinery to enforce the production of viral proteins before initiating its replication.<sup>17</sup> The analysis of the SARS-CoV-2 transcriptome<sup>18</sup> shows that, in addition to structural proteins, including the widely investigated Spike protein,<sup>19</sup> which form the viral envelope and allow the interactions with human receptors,<sup>20,21</sup> two large polyproteins are expressed and self-cleaved by two viral enzymes, i.e., the 3CL-like (3CL<sup>PRO</sup>) and the papain-like (PL<sup>PRO</sup>) proteases.<sup>22–25</sup> This leads to the maturation of several nonstructural proteins (NSP) which exert different (enzymatic)

Received: June 28, 2022

Published: August 19, 2022





**Figure 1.** (A) Representative structure extracted from the classical MD simulation of the RdRP (blue) interacting with an RNA double strand (red) composed of the template and the nascent strand. A close-up of the active site is also provided, highlighting GTP, the  $Mg^{2+}$  ions, and key protein residues such as K551. (B) Sequence of the RNA strand used consistently for the entire simulation. (C) Proposed chemical mechanisms for the inclusion of GTP on the nascent RNA strand.

functions allowing SARS-CoV-2 to complete its maturation and assembly into novel virions, while also eluding the innate cellular immune system. Among those, the RNA-dependent RNA polymerase (RdRP), composed of the complex NSP12/NSP7, is of particular importance.<sup>26,27</sup> It uses the original viral positive-sense RNA strand to produce an intermediate negative-sense strand which is further used as a template allowing the synthesis of a novel duplicated positive-sense RNA strand. Of note, the RNA replication takes place in the cytoplasm in double-membrane vesicles and differently from other RNA viruses, SARS-CoV-2 also disposes of a replication proof-reading system based on a viral exonuclease<sup>28,29</sup> which limits replication errors and considerably slows down its mutation rate.

Obviously, SARS-CoV-2 RdRP constitutes an ideal target for possible pharmacological development.<sup>30–32</sup> Its inhibition would, indeed, result in the arrest of the viral replication cycle and of the infection. As a matter of fact, some nucleotide analogues, also used against retroviruses such as Human Immunodeficiency Virus (HIV), have been proposed for SARS-CoV-2 treatment, and while they have shown interesting *in vitro* and *in vivo* efficiency, their clinical use is strongly limited also by rather severe side-effects and nontrivial pharmacokinetics and administration.<sup>31,32</sup> Among the proposed nucleotide analogues inhibiting SARS-CoV-2, we may cite remdesivir,<sup>33</sup> acting by impeding the RNA translocation,<sup>34</sup> which follows the inclusion of a new nucleotide to free the active site and slide along the template, and favipiravir,<sup>35,36</sup> which instead directly inhibits the RdRP catalytic site.<sup>37</sup>

While the structure of SARS-CoV-2 RdRP has been resolved,<sup>26,37</sup> also in the presence of inhibitors, studies focusing on the enzymatic mechanism from a molecular and biochemical

point of view are scarce.<sup>38</sup> However, the precise characterization of the rate-limiting steps of the enzymatic process, together with the identification of the structures of the key intermediates and transition states, would be of invaluable help in the design of efficient inhibitors. Indeed, understanding how the active site of the enzyme is organized during the reaction could drive the design of competitive binders able to inhibit the polymerase activity.

From a chemical standpoint, polymerases are metalloenzymes, sharing many structural and chemical similarities with exo- and endonucleases. Their active site is usually constituted by two  $Mg^{2+}$  ions which are stabilized by the interaction with hard electron donors such as the anionic oxygen atoms of aspartates (Figure 1). Furthermore, the presence of histidine residues, which in addition to stabilizing the magnesium ions may also participate in proton shuttle completing the catalytic cycle, is also widely conserved and appears as fundamental.<sup>39</sup> Upon reactive conformations, the  $Mg^{2+}$  ions also interact with the 3' oxygen of the nascent nucleic acid strand (3') and with the triphosphate form of the incorporated nucleotide. The selectivity and replication accuracy are assured by the fact that the triphosphate reactant nucleotide is Watson and Crick paired with the complementary base on the template strand. It has been shown that the 3' OH group of the nascent strand is easily deprotonated<sup>40</sup> by an environmental  $OH^-$  ion, leading to an intermediate that is prone to react and that is stabilized by the two  $Mg^{2+}$  ions. The incorporation of the nucleotide is obviously accompanied by the cleavage of P–O bonds on the triphosphate reactant. The release of the pyrophosphate intermediate and its protonation, usually assured by a catalytic histidine, are also necessary to complete the cycle.<sup>39</sup> If SARS-CoV-2 RdRP shares

strong similarities with other polymerases, particularly the presence of the  $Mg^{2+}$  ions, its active site is nonetheless peculiar due to the absence of any histidine in close proximity of the enzymatic pocket. This suggests that the precise chemical mechanisms leading to the inclusion of the triphosphate nucleotide in the nascent RNA strand may be different from those commonly admitted in analogous systems.<sup>39,41</sup> Aranda et al. recently published a computational investigation of UTP, ATP, and remdesivir incorporation mechanisms by SARS-CoV-2 RdRP. It is noteworthy that they propose the first hypothesis of an enzymatic mechanism featuring the deprotonation of the O3' by the pyrophosphate group resulting from the previous NTP incorporation, prior to the binding and incorporation of the new NTP.<sup>38</sup> However, their starting system was built using homology modeling prior to the release of several experimental structures of SARS-CoV-2 RdRP. In this contribution, starting from the prereactive structure of the SARS-CoV-2 RdRP in complex with an RNA double strand resolved by Naydenova et al.<sup>37</sup> (PDB code 7AAP, see Figure 1A), we apply molecular modeling and simulations techniques using both classical and hybrid quantum mechanical/molecular mechanical (QM/MM) strategies to understand both the structural features and the chemical mechanism of RdRP, which is crucial for the replication of SARS-CoV-2. In the following, after sampling the conformational space of the RdRP in complex with ds-RNA by using classical MD simulations, we perform QM/MM-MD simulations to unravel the catalytic mechanism leading to the inclusion of a nucleotide in the nascent RNA strand, also obtaining the reaction free energy profile. By coupling DFT/MM-MD to 2D-umbrella sampling techniques, we reveal the fundamental role of a lysine assisting the reaction and efficiently compensating for the absence of catalytic histidines nearby the active site, as well as the low activation barrier of the polymerase-catalyzed reaction which highlights the high efficiency of this polymerase.

## METHODS

All classical molecular dynamics (MD) simulations were carried out using NAMD3.<sup>42,43</sup> QM/MM-MD simulations were performed using the Terachem code<sup>44</sup> interfaced with Amber16.<sup>45</sup> Visualization, rendering of MD trajectories, and plots were performed using VMD,<sup>46</sup> Gnuplot, and RStudio software.

**System Setup and Classical MD Simulations.** The starting structure was generated from the crystal structure of SARS-CoV-2 RNA-dependent RNA polymerase complexed with the favipiravir inhibitor (PDB ID 7AAP). The favipiravir moiety was *in silico* mutated to the reactive guanine triphosphate (GTP), keeping the catalytic magnesium ions as positioned in the crystal structure. As the deprotonation of the terminal base has been reported to occur prior to the NTP binding<sup>38,39,41</sup> and is barrierless in similar systems,<sup>40</sup> we chose to focus on the reaction step involving the attack of the deprotonated 3' terminal nucleotide (rU9) of the synthesized RNA strand on the already bound GTP; hence, the O3' was coherently kept deprotonated. Force field parameters of the 3' deprotonated uracil were generated following the antechamber protocol, as described in the Supporting Information. The magnesium ion parameters were taken from Advist et al. JPC 1990.<sup>47</sup> The system was then soaked into a TIP3P<sup>48</sup> water box with a 10 Å buffer and potassium ions to ensure a neutral global charge, resulting in a total of ~181 000 atoms. The ff14SB force field<sup>47</sup> including the OL3 corrections<sup>49</sup> for RNA were used to describe

the protein and the nucleic acid. RdRP presents zinc fingers which have been modeled using nonbonded parameters.<sup>50</sup> The protonation states of the protein residues were assigned with respect to the  $pK_a$  prediction made with propka3.1, considering the system at pH 7.<sup>51</sup> Of note, a cysteine near the active site (C622) was considered as deprotonated for it is involved in the coordination of the magnesium ions.

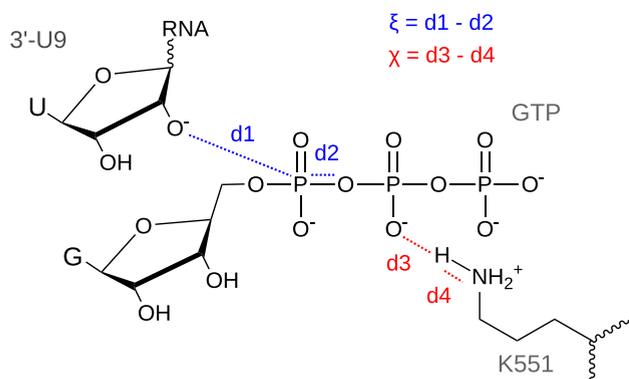
Hydrogen mass repartitioning (HMR)<sup>52</sup> was consistently applied to all hydrogens excluding water, thus allowing, in combination with Rattle and Shake,<sup>53</sup> the use of a 4 fs time-step to numerically solve the Newton equations of motion. Prior to the simulation, the system geometry was optimized in four sets of 10 000 steps using the conjugate gradient method with gradually decreasing restraints on the backbone atoms of the protein and RNA. Each of these steps was followed by a short 1.2 ns equilibration with the same restraints in order to ensure a proper relaxation of the system. A 150 ns production run was then performed at 300 K in the isotherm and isobaric (NTP) ensemble. The temperature was kept constant using the Langevin thermostat with a 1 ps<sup>-1</sup> collision frequency, and the electrostatic interactions were treated using the particle mesh Ewald (PME)<sup>54</sup> scheme with a 9 Å cutoff. A 10 kcal/mol restraint was imposed on the distance between the magnesium ions to compensate for the electrostatic repulsion and maintain the bimetallic active site in a prereactive conformation. Of note, this technique was successfully used in our previous studies.<sup>55</sup>

A structural analysis was performed using the cpptraj module of Amber16 and the Curves+ program.<sup>56</sup> A clustering of the trajectory was performed based on the RMSD of the protein/RNA complex, using the hierarchical agglomerative approach with the average linkage method. This allowed an extraction of the most abundant conformation sampled as a starting structure for the subsequent QM/MM-MD simulations.

**Equilibrium QM/MM-MD Simulations.** The representative frame of the most populated cluster issued from classical MD simulations was taken as a starting point for the QM/MM-MD approach. The QM partition was defined to encompass the triphosphate moiety and C5' atom of the GTP (excluding the ribose and the nucleobase); C2'; C4'; C3'; 3' uracil (rU9); magnesium ions; and K551, K798, D760, D761, D618, and C622 residues—see Figure S1 for a representation of the QM partition. The QM/MM boundary in the GTP was set between the C4' and C5' atom for this bond is not polarizable. All of the involved amino acids were truncated between the  $C\alpha$  and  $C\beta$  positions, resulting in a total of 81 atoms described at the QM level with a total charge of  $-3$  and in a singlet spin state. The rest of the system was described at the MM level. Dangling covalent bonds have been treated with the link atom (hydrogen) approach.<sup>57</sup> QM calculations were performed at the density functional theory (DFT) level using the  $\omega$ B97x-D exchange and correlation functional<sup>58</sup> and the double- $\zeta$  6-31G basis set.<sup>59</sup> Although the use of a relatively small basis set can be problematic especially for phosphorus energetics,<sup>60</sup> we decided to not include diffuse functions to avoid wave function overpolarization by the MM point charges while also maintaining a computational cost small enough to ensure an extended sampling by dynamics. The system was first relaxed at the QM/MM level for 10 ps, allowing the proper rearrangement of the active site and providing a relevant starting structure for the subsequent enhanced sampling free-energy calculation. Note that the constraints on the magnesium cluster used for the preliminary classical MD simulations have been lifted during all of the subsequent QM/MM-MD steps; both the bimetallic

cluster and its coordination sphere remained stable. The time-step for all of the QM/MM-MD simulations was set to 1 fs, and all of the Rattle and Shake constraints on the QM hydrogens were relaxed.

**Enhanced Sampling and Free Energy Calculation.** To allow for the exploration of chemical reactivity, we resorted to enhanced sampling and the more specific umbrella sampling (US) protocol as implemented in Amber16, thus obtaining a relevant free energy profile. Specifically, two collective variables were used to map the free energy surface (FES) corresponding to the incorporation of GTP into the nascent RNA strand (see Figure 2). The first one ( $\xi$ ) is defined as the difference between



**Figure 2.** Schematic definition of the collective variables  $\xi$  and  $\chi$  used for the US simulation.

the distances U3':O3'–GTP:PA (i.e., the phosphate forming bond) and GTP:PA–GTP:OP3 (i.e., the breaking bond). It is clear that  $\xi$  is a well-adapted degree of freedom to describe the advancement of the nucleophilic substitution leading to the incorporation of the nucleotide in the RNA strand. As previously mentioned, proton shuttles are usually fundamental to ensure polymerase efficiency, in particular assuring proton transfer to the pyrophosphate leaving group. To account for this, the second collective variable ( $\chi$ ) was defined as the difference between the distances K551:HZ1–GTP:O1B (forming bond) and K551:HZ1–K551:NZ (breaking bond), hence describing proton transfer from K551. US was performed partitioning the collective variable space into a regular grid spanning values of  $\xi$  from  $-2.0$  to  $1.8$  Å and values of  $\chi$  from  $-0.5$  to  $1.0$  Å with steps of  $0.1$  Å. This resulted in 624 windows. For each window, after a 1 ps equilibration, 5–10 ps of production was performed, for a total sampling of 3.2 ns. Values of  $\xi$  and  $\chi$  were constrained on each window by harmonic potentials with anchors of strength of 600 kcal/mol set at  $-10$  and  $10$  Å. The results were unbiased, and FES was reconstructed using the 2D version of the weighted histogram analysis method (WHAM2D).<sup>61</sup> The convergence of the WHAM2D procedure was confirmed by checking the proper overlap between the windows (see Figure S2).

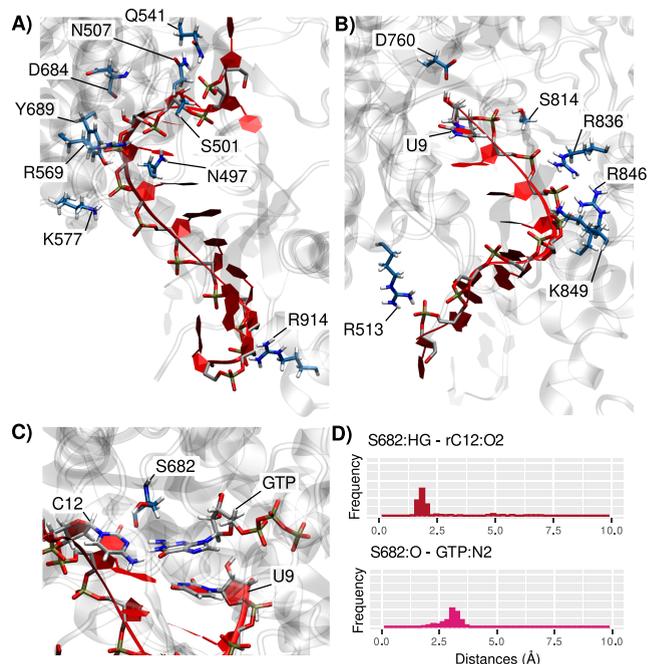
## RESULTS AND DISCUSSION

**Structural Features of SARS-CoV-2 RdRP.** The structural dynamic behavior of the RdRP:RNA complex was sampled by all-atom classical MD simulations. In particular, in the following, we describe the interaction network, leading to a stable protein/nucleic acid complex, as well as the main structural deformations experienced by the RNA strand and the enzymatic site. Importantly, to ensure the accuracy of the viral replication, RdRP relies on the Watson and Crick coupling between a

nascent RNA strand and its original template. Hence, the nucleic acid needs to assume a double-stranded secondary structure, while its precise tertiary folding is largely dictated by the interaction with the RdRP interface. In the following, we describe the protein–RNA interactions as nonspecific (involving the RNA phosphate groups) or specific (involving the sugar rings or the nucleobases).

At the interface between the nucleic acid and the enzyme, basic and polar amino acids develop nonspecific contact with the RNA backbone leading to a dense interaction network.

On the RdRP template strand side, positively charged (R569, R914, and K577) and polar (Y689, S501, Q541, N497, and N507) residues interact with the negatively charged phosphate groups of the RNA backbone (see Figures 3 and S3). While only



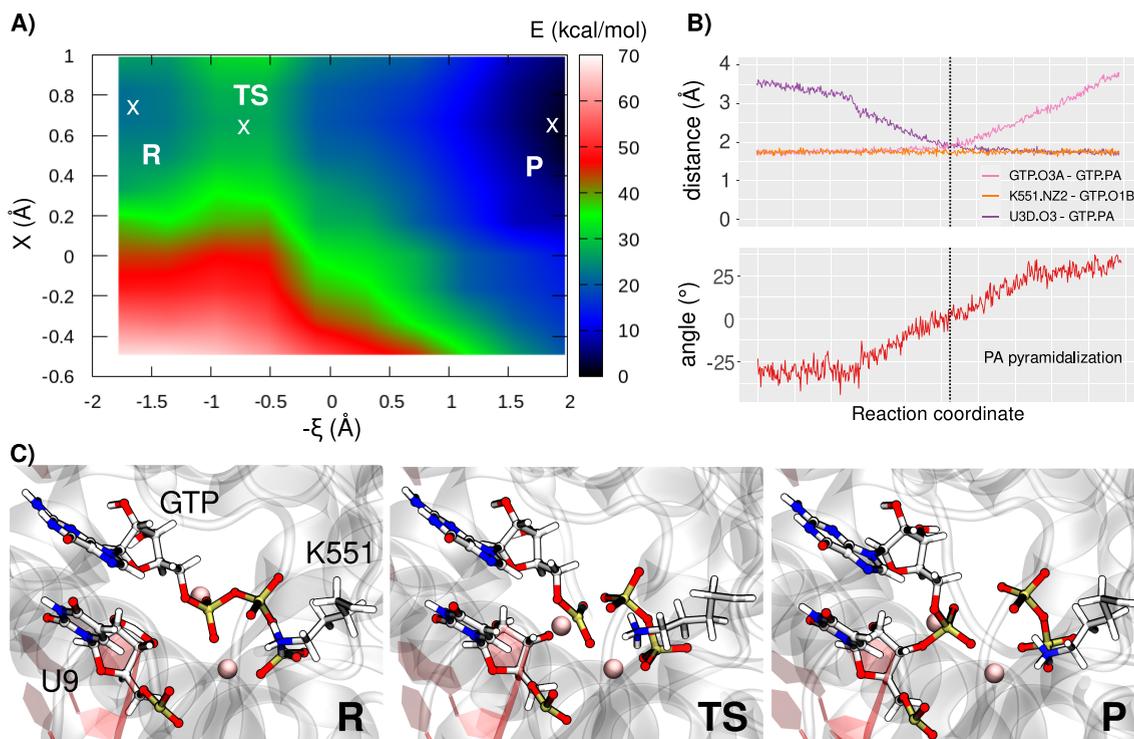
**Figure 3.** Main interactions between the enzyme (residues with blue carbon atoms) and (A) the RNA template strand or (B) the RNA nascent strand. The RNA strands are depicted in red, and the protein structure appears in light gray. (C) Zoom-in of the interaction of the key-residue S682 hydroxyl group with the nucleotide to be integrated (GTP) and the facing rC12 on the template strand. (D) Distribution of the distances between S682 and the base pair that is about to be formed.

one specific interaction is found between the protein and an RNA sugar ring, involving D684:O (backbone) and rA13:HO2', the latter is highly conserved along the simulation. Interestingly, S682:HG forms a hydrogen bond with rC12:O2, the nucleobase facing the incorporated GTP, facilitating the Watson and Crick pairing. This evidence suggests, as reported in previous works,<sup>26,37</sup> that this amino acid might play a role in the relative positioning of the template strand and the reacting nucleotide in the active site, hence being crucial for efficiency and accuracy of the replication.

Fewer interactions are observed involving the nascent strand. R513, R836, R858, and K849 develop persistent contacts with the RNA phosphates, probably helping the correct positioning of the strand and more importantly its translocation. Persistent interactions with the backbone and sugar ring of the terminal 3'-rU9 nucleotide are also observed that involve S814 and D760, averaged distance of  $2.25 \pm 0.07$  Å and  $1.94 \pm 0.04$  Å, for

**Table 1.** Averages and Standard Deviations of the Local and Global Bend Angles Calculated along the Classical MD Trajectory

	base pair								global
	2	3	4	5	6	7	8	9	
bend angle (deg)	6.8	4.8	3.9	3.6	3.4	2.4	1.7	1.4	25.4
standard deviation (deg)	$\pm 6.7$	$\pm 4.4$	$\pm 3.0$	$\pm 2.2$	$\pm 1.8$	$\pm 1.4$	$\pm 0.9$	$\pm 0.7$	$\pm 19.1$



**Figure 4.** (A) 2D free energy surface obtained from QM/MM umbrella sampling simulations. (B) Evolution along the reaction coordinate  $\xi$  of the distances between GTP and protein or RNA moieties (upper panel) and of the pyramidalization angle of the reacting phosphate (lower panel). Note that, due to the topology of the free energy surface, only the value of  $\chi = 0.7 \text{ \AA}$  has been considered. (C) Representative snapshots describing the critical points observed on the free energy maps; the position of the reactants (R), transition state (TS), and products (P) in the free energy basins is also provided in panel A.

rU9:OP1–S814:HG and rU9:HO3'–D760:OD1, respectively. Interestingly, since the latter residue is part of the active site, it is also coordinated to one of the catalytic magnesium ions.

The interaction between the double-strand nucleic acid and the enzyme, leading to a contact region spanning from base pair 5 to 9, is also translated into a stiffening and stabilization of the RNA. Indeed, the structural parameters of the RNA base-pairs as computed with Curves+ indicate a highly stable and mostly ideal B-conformation of the double strand when in contact with the polymerase interface. This observation is most probably due to a highly dense interaction network involving strong electrostatic interactions as described above. Concerning the global structural descriptors of duplex RNA, the bending remains low for the base-pairs buried in the protein (up to  $3.6 \pm 2.2^\circ$  for bp 5), while the regions more exposed to the solvent experience a higher flexibility with the local bending reaching  $6.8 \pm 6.7^\circ$  for bp 2 (see Table 1). This behavior is also reflected in the systematically higher deviations of the axis parameters for the regions of the double strand which are more exposed to the bulk water (see Figure S4). Overall, the protein embedding stabilizes and stiffens the RNA duplex structure, hence assuring the Watson and Crick coupling and offering an appropriate

organization of the active site for the efficient incorporation of the reacting nucleotides into the nascent strand.

As concerns the local organization of the active site, a first magnesium ion coordinates to D760, D618, the GTP reactive phosphate group (OP1), the terminal deprotonated O3' of the nascent RNA strand, and two water molecules (see Figure S5). This interaction clearly helps in maintaining the reactive moieties ideally placed for the reaction to take place (see Figure 1A). The second  $\text{Mg}^{2+}$  ion interacts with the two other phosphates of GTP, D618, as well as water molecules bridging with D761 and the 3' uracil phosphate group and O3' atom. The sugar moiety of GTP is also involved in a hydrogen bond with C622. This highly organized electrostatic network within the active site, maintained by the presence of the  $\text{Mg}^{2+}$  ions, provides a nearly ideal structure favoring the nucleophilic substitution. As a matter of fact, the reactive rU9:O3' and GTP:PA are kept only  $3.5 \text{ \AA}$  apart, while the corresponding nucleobases develop an ideal  $\pi$ -stacking further stabilizing their arrangement. A canonical Watson–Crick hydrogen bonding scheme is observed between GTP and the matching rC12 on the template strand. It is noteworthy that the composition of the coordination spheres around the magnesium ions in our simulations slightly differs from what was reported by Aranda et al.,<sup>38</sup> whose system

exhibits no coordination to any water molecules, yet they report similar O3'–PA distance values.

As already pointed out, the active site of RdRP lacks a histidine, which could potentially participate in the proton transfer necessary for the completion of the reaction.<sup>39</sup> While the analysis of the crystal structure of RdRP does not allow the identification of any suitable amino acid nearby the Mg<sup>2+</sup> ions capable of assuming this role, the evolution of the MD simulation reveals that a proton donor, namely, K551, rapidly approaches the pyrophosphate leaving group and is then stabilized in this conformation. K551 is not directed toward the active site in the crystal structure of RdRP with favipiravir (K551:NZ–GTP:PB at 7.16 Å); however, when the ligand is mutated to GTP, a reorganization takes place in which K551 forms strong hydrogen bonds with the pyrophosphate group (K551:H22–GTP:O2G at 1.74 Å). Therefore, we may hypothesize that K551 may compensate for the absence of the histidine, which in turn suggests an important role of this residue in the SARS-CoV-2 RNA polymerase enzymatic mechanism. This hypothesis will also be further confirmed by our QM/MM simulations. Of note, this lysine is highly conserved in other viral polymerases, hence suggesting a conserved biological role of this residue, though it would require additional investigations to be confirmed (see Figure S6). Importantly, the main characteristics of the active site of RdRP in complex with GTP are also retained during the QM/MM equilibration, in which only a slight adaptation of the Mg–Mg distance can be noted.

**Reaction Free Energy Profile.** The free energy profile of the enzymatic reaction catalyzed by RdRP was obtained to precisely quantify the associated activation energy and driving force. The common reaction mechanism for polymerases is based on a two-step process: (i) the activation of the 3' nucleotide of the nascent strand through its deprotonation, also coupled with the protonation of the leaving pyrophosphate group (see Figure 1C); and (ii) the nucleophilic attack involving the activated 3'-nucleotide and the incorporating triphosphate nucleotide, resulting in the elongation of the nascent RNA strand and the release of the pyrophosphate group.

Here, only the rate-limiting step of the global reaction, i.e., the nucleophilic attack,<sup>40</sup> was simulated by QM/MM-MD calculations. The free energy surface was determined along two reaction coordinates simultaneously describing the attack of rU9:O3' onto the GTP phosphate ( $\xi$ ) and the assistance and potential proton transfer of the  $-\text{NH}_3^+$  moiety of K551 facilitating the pyrophosphate departure ( $\chi$ ). Indeed, K551 has been pinpointed by structural consideration and sequence alignment as a key residue for the polymerase reactivity. Equilibrium MD simulations on the protonated U9:O3' show a high water accessibility of the OH group, with at least one solvent molecule present in the closest coordination shell of rU9:O3' for 72% of the global simulation time. The solvent exposition and the usual lability of alcohol moieties contribute to suggest a facile deprotonation of the 3'-terminal nucleotide and support our choice to concentrate on the second step only.

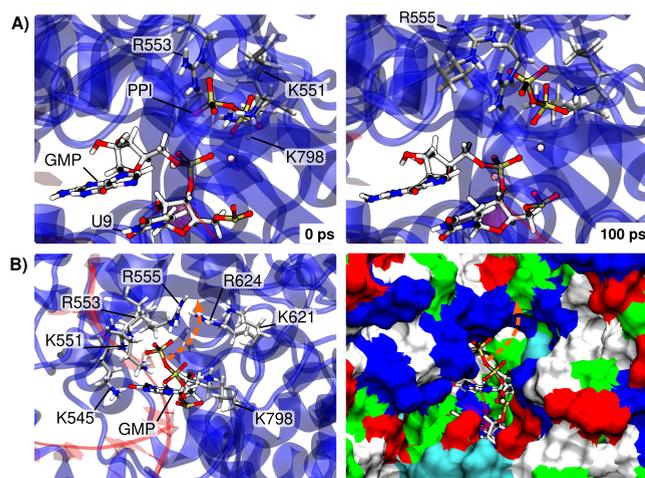
The 2D free energy surface along  $\xi$  and  $\chi$  is reported in Figure 4A, together with representative snapshots of critical points of the surface (Figure 4B). The reaction exhibits a classical nucleophilic substitution mechanism, going from the reactant (R) to the product (P) through a transition state (TS) without any intermediate. The TS is accessed bypassing only a moderate to low activation barrier  $\Delta G^\ddagger$  of 10 kcal/mol, hence confirm the efficiency of RdRP. Additionally, the TS exhibits a perfectly ideal planar reacting phosphate group, which is also equidistant from

O3' and O3A (1.9 Å) as was observed for ATP<sup>38</sup> and as can be appreciated from the evolution of the distance and the phosphate pyramidalization angle reported in Figure 4B,C. From a thermodynamic point of view, the reaction is found to be exergonic, with a driving force of  $\Delta G^\circ = 18$  kcal/mol. This value is higher than that reported for UTP and ATP by Aranda et al.<sup>38</sup> However, we should underline that their profile was conducted at a semiempirical level and only afterward corrected with DFT, leading also to slight structural differences compared to our model. This fact, as well as the use of different basis sets, does not allow making a direct quantitative comparison. However, both models preview general similarities in the reaction mechanisms and an easy polymerization reaction. Upon reaching the TS, the distance between the catalytic Mg<sup>2+</sup> ions is slightly increased from 4.2 to 4.8 Å. The first ion maintains its interaction with D760 and D618, but not anymore with the rU9:O3' which is no longer accessible and is instead coordinated to the phosphate's oxygen of the newly incorporated guanine monophosphate. The second ion also interacts with the backbone of the new 3' nucleotide, D168, and one end of the pyrophosphate.

Interestingly, the proton transfer from K551 to the pyrophosphate does not appear to be required for the completion of the reaction, as can be appreciated by the topology of the free energy surface and by the fact that the minimum energy path lays around values of  $\chi = 0.7$  Å. However, the role of K551 is crucial as it induces an important stabilization of the reaction product. Interestingly, the lysine proton is much more labile in the product region, and in stark contrast with the reactants, the presence of a large low-energy basin spanning values of  $\chi$  comprised between  $-0.6$  and  $1.0$  Å should be highlighted. Hence, once the nucleotide is incorporated into the nascent strand, K551 is able to partially share its proton with the leaving pyrophosphate group facilitating its departure. This peculiar topology of the free energy surface also confirms the different nature of RdRP enzymatic reactions. Indeed, the absence of the histidine precludes the possibility of a proton-shuttle mechanism; however, a lysine, in this case K551, may act as a partial proton donor leading to a product region in which the additional charge is largely delocalized over different moieties. To the best of our knowledge, this peculiar mechanism has not been identified in other polymerases and shows how unusual active sites may lead to highly efficient turnovers.

Starting from the structure of the reaction product (window  $\xi = 2.0$  and  $\chi = 0.7$ ), an additional unbiased QM/MM-MD trajectory of 100 ps was performed to sample at least the initial phases of the pyrophosphate group departure. In the course of the unbiased dynamics, K551 continues to interact with the leaving group and interestingly spontaneously drags it toward the bulk. In addition to K551, other basic amino acids, namely, K798, R553, and R555, approach the active site and develop interactions with the pyrophosphate globally favoring its departure from the active site (see Figure 5A).

Since in our simulation one of the phosphates remains coordinated to a magnesium ion, we have not observed a complete detachment of the leaving group. However, we may anticipate that subsequent classical MD simulations might reveal the complete exit path of the pyrophosphate from the active site, also thanks to the development of the interaction networks with the basic amino acids. The inspection of the basic amino acids and surface of the protein offering a possible exit path for PPI show an ideal way upward, with many lysines and arginines (K551, R553, R555, R624, and K798) that might favor the passage of the leaving group toward the bulk (see Figure 5B).



**Figure 5.** (A) Structure of the RdRP active site at the beginning (0 ps, left) and end (100 ps, right) of the unbiased QM/MM-MD simulations of RdRP after GMP incorporation into the nascent strand. The pyrophosphate (PPI) and basic amino acids within 5 Å are displayed, showing the movement of PPI upward, moving closer to R555 toward the exit. (B) Global view of the basic amino acids within 10 Å of the PPI moiety en route toward the exit (after 100 ps of unbiased sampling). On the left, the basic amino acids are explicitly depicted in licorice, and on the right, a surface representation of the protein is used with a color code of the amino acids corresponding to their type (basic in blue), highlighting the possible exit path for PPI. This path appears as an orange arrow on both representations.

Interestingly, Naydenova et al. report a nonproductive conformation of RdRP harboring favipiravir and a pyrophosphate above it that interacts with R555 and K621.<sup>37</sup> They suggest that this pyrophosphate group might be a byproduct from previous incorporation of NTPs into the nascent RNA strand. Its position corroborates the exit route that we hypothesize here. Furthermore, in addition to being stabilized by the positively charged residues, the pyrophosphate could also be protonated by another acidic group. However, this step, which remains at the moment rather speculative, was not proven by our study and goes beyond the scope of the present contribution, probably involving a redefinition of the QM partition and of the reaction collective variables. In addition, examples of polymerases have been reported in the literature,<sup>62</sup> also supported by the values of the experimental  $pK_a$ ,<sup>63–65</sup> in which the incoming nucleotide triphosphate is protonated prior to its binding to the nascent RNA strand or immediately after the binding itself. However, in all of these cases, the protonation of the phosphate is not the rate-limiting step, and hence, the general validity of our conclusions should hold.

## CONCLUSION

RdRP is a crucial enzyme in the viral cycle of SARS-CoV-2, and more generally, RNA-dependent RNA polymerases constitute a crucial target for the design of inhibitors against emerging viruses. While the structure of RdRP, in complex with different inhibitors, has been successfully resolved since the first outbreaks of the COVID-19 pandemics, its catalytic mechanism has not yet been fully elucidated from a biochemical point of view. With this contribution, following a multiscale approach combining classical and QM/MM simulations including enhanced sampling, we provide for the first time the description of the catalytic mechanism of RdRP, while quantifying the activation energy necessary for the RNA polymerization. From a

structural point of view, RdRP presents a high density of basic amino acids, producing a positively charged groove which is necessary to accommodate, in a rather nonspecific way, the Watson and Crick paired nascent and template strands. Interestingly, some specific interactions locking the template base in an ideal position to interact with the incorporated nucleotide are observed, thus increasing the specificity. If the RdRP active site presents many of the classical features of polymerases, including two  $Mg^{2+}$  ions, it lacks histidines susceptible to assist the polymerization by inducing a proton transfer to the leaving pyrophosphate group. The combination of classical and QM/MM-MD simulations has allowed a highlighting of the role of K551 in compensating for the absence of a histidine and in stabilizing the leaving pyrophosphate with the assistance of one of its protons. These observations differ from previous works based on homology modeling RdRP structures.<sup>38</sup> In particular, we have evidenced that, while the polymerization reaction necessitates an activation energy of only 10 kcal/mol, the presence of K551 is also fundamental to favor the departure of the pyrophosphate, also assisted by other basic amino acids. In agreement with previous works featuring UTP and ATP incorporation in a homology modeled RdRP,<sup>38</sup> we may conclude that RdRP is characterized by a high enzymatic efficiency which could translate to a high viral replication rate. Importantly, our work brings out new information about the unusual active site and a partially different chemical mechanism exhibited by SARS-CoV-2 RdRP. Our results, in providing a clear structural and mechanistic picture of the polymerization reaction, may open the way to the design of specific inhibitors, mainly based on nucleotide analogues. In particular, the key role played by K551 may also open possibilities for the design of inhibitors specifically targeting this residue.

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jcim.2c00802>.

Parameters and parametrization procedure used to define the nonstandard nucleic bases used in this study; representation of the atoms in the QM partition; and additional structural parameters for the RNA double strand obtained using Curves+ (PDF)

Movie extracted from the US simulation representing the structural evolution during the polymerase reaction (MPG)

## AUTHOR INFORMATION

### Corresponding Authors

**Emmanuelle Bignon** – Université de Lorraine and CNRS, LPCT UMR 7019, F-54000 Nancy, France; [orcid.org/0000-0001-9475-5049](https://orcid.org/0000-0001-9475-5049); Email: [emmanuelle.bignon@univ-lorraine.fr](mailto:emmanuelle.bignon@univ-lorraine.fr)

**Antonio Monari** – Université de Lorraine and CNRS, LPCT UMR 7019, F-54000 Nancy, France; Université de Paris, CNRS, ITODYS, F-75006 Paris, France; [orcid.org/0000-0001-9464-1463](https://orcid.org/0000-0001-9464-1463); Email: [antonio.monari@u-paris.fr](mailto:antonio.monari@u-paris.fr)

Complete contact information is available at <https://pubs.acs.org/doi/10.1021/acs.jcim.2c00802>

### Notes

The authors declare no competing financial interest.

**Data Availability.** Inputs files can be found on our GitHub repository: [https://github.com/emmanuellebignon/SARS-CoV-2\\_RdRP\\_reactivity-data](https://github.com/emmanuellebignon/SARS-CoV-2_RdRP_reactivity-data)

## ACKNOWLEDGMENTS

The authors thank GENCI (HPC resources of IDRIS under the allocation 2021-102152), Explor, and local LPCT computing centers for computational resources. E.B. thanks the CNRS and French Ministry of Higher Education Research and Innovation (MESRI) for her postdoc fellowship under the GAVO program. A.M. thanks ANR and CGI for their financial support of this work through Labex SEAM ANR 11 LABX 086, ANR 11 IDEX 05 02. The support of the IdEx “Université Paris 2019” ANR-18-IDEX-0001 and of the Platform P3MB is gratefully acknowledged. All of the software products used to perform this study are available online.

## REFERENCES

- (1) Liu, Y. C.; Kuo, R. L.; Shih, S. R. COVID-19: The first documented coronavirus pandemic in history. *Biomedical Journal* **2020**, *43*, 328–333.
- (2) Sohrobi, C.; Alsafi, Z.; O'Neill, N.; Khan, M.; Kerwan, A.; Al-Jabir, A.; Iosifidis, C.; Agha, R. World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). *International Journal of Surgery* **2020**, *76*, 71–76.
- (3) Harapan, H.; Itoh, N.; Yufika, A.; Winardi, W.; Keam, S.; Te, H.; Megawati, D.; Hayati, Z.; Wagner, A. L.; Mudatsir, M. Coronavirus disease 2019 (COVID-19): A literature review. *Journal of Infection and Public Health* **2020**, *13*, 667–673.
- (4) Tang, X.; Wu, C.; Li, X.; Song, Y.; Yao, X.; Wu, X.; Duan, Y.; Zhang, H.; Wang, Y.; Qian, Z.; Cui, J.; Lu, J. On the origin and continuing evolution of SARS-CoV-2. *National Science Review* **2020**, *7*, 1012–1023.
- (5) Zhu, N.; et al. A novel coronavirus from patients with pneumonia in China, 2019. *New England Journal of Medicine* **2020**, *382*, 727–733.
- (6) Gray, W. K.; et al. Variability in COVID-19 in-hospital mortality rates between national health service trusts and regions in England: A national observational study for the Getting It Right First Time Programme. *EclinicalMedicine* **2021**, *35*, 100859.
- (7) Tchicaya, A.; Lorentz, N.; Leduc, K.; de Lanchy, G. COVID-19 mortality with regard to healthcare services availability, health risks, and socio-spatial factors at department level in France: A spatial cross-sectional analysis. *PLoS One* **2021**, *16*, No. e0256857.
- (8) Luo, G.; Zhang, X.; Zheng, H.; He, D. Infection fatality ratio and case fatality ratio of COVID-19. *International Journal of Infectious Diseases* **2021**, *113*, 43–46.
- (9) Sun, Y.; Lin, W.; Dong, W.; Xu, J. Origin and evolutionary analysis of the SARS-CoV-2 Omicron variant. *Journal of Biosafety and Biosecurity* **2022**, *4*, 33–37.
- (10) Del Rio, C.; Omer, S. B.; Malani, P. N. Winter of Omicron-The Evolving COVID-19 Pandemic. *JAMA* **2022**, *327*, 319–320.
- (11) Karim, S. S. A.; Karim, Q. A. Omicron SARS-CoV-2 variant: a new chapter in the COVID-19 pandemic. *Lancet* **2021**, *398*, 2126–2128.
- (12) Rubin, E. J.; Longo, D. L. Covid-19 mRNA Vaccines - Six of One, Half a Dozen of the Other. *New England Journal of Medicine* **2021**, *386*, 183–185.
- (13) Sadarangani, M.; Marchant, A.; Kollmann, T. R. Immunological mechanisms of vaccine-induced protection against COVID-19 in humans. *Nature Reviews Immunology* **2021**, *21*, 475–484.
- (14) Forni, G.; Mantovani, A.; et al. COVID-19 vaccines: where we stand and challenges ahead. *Cell Death and Differentiation* **2021**, *28*, 626–639.
- (15) Hu, B.; Guo, H.; Zhou, P.; Shi, Z. L. Characteristics of SARS-CoV-2 and COVID-19. *Nature Reviews Microbiology* **2021**, *19*, 141–154.
- (16) Naqvi, A. A. T.; Fatima, K.; Mohammad, T.; Fatima, U.; Singh, I. K.; Singh, A.; Atif, S. M.; Hariprasad, G.; Hasan, G. M.; Hassan, M. I. Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: Structural genomics approach. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* **2020**, *1866*, 165878.
- (17) Zhang, Q.; Xiang, R.; Huo, S.; Zhou, Y.; Jiang, S.; Wang, Q.; Yu, F. Molecular mechanism of interaction between SARS-CoV-2 and host cells and interventional therapy. *Signal Transduction and Targeted Therapy* **2021**, *6*, 1–19.
- (18) Kim, D.; Lee, J.-Y.; Yang, J.-S.; Kim, J. W.; Kim, V. N.; Chang, H. The Architecture of SARS-CoV-2 Transcriptome. *Cell* **2020**, *181*, 914–921.e10.
- (19) Kang, S.; Yang, M.; Hong, Z.; Zhang, L.; Huang, Z.; Chen, X.; He, S.; Zhou, Z.; Zhou, Z.; Chen, Q.; Yan, Y.; Zhang, C.; Shan, H.; Chen, S. Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites. *Acta Pharmaceutica Sinica B* **2020**, *10*, 1228–1238.
- (20) Wrapp, D.; Wang, N.; Corbett, K. S.; Goldsmith, J. A.; Hsieh, C.-L. L.; Abiona, O.; Graham, B. S.; McLellan, J. S. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **2020**, *367*, 1260–1263.
- (21) García-Iriepa, C.; Hognon, C.; Francés-Monerris, A.; Iriepa, I.; Miclot, T.; Barone, G.; Monari, A.; Marazzi, M. Thermodynamics of the Interaction between the Spike Protein of Severe Acute Respiratory Syndrome Coronavirus-2 and the Receptor of Human Angiotensin-Converting Enzyme 2. Effects of Possible Ligands. *J. Phys. Chem. Lett.* **2020**, *11*, 9272–9281.
- (22) Ullrich, S.; Nitsche, C. The SARS-CoV-2 main protease as drug target. *Bioorg. Med. Chem. Lett.* **2020**, *30*, 127377.
- (23) Ramos-Guzmán, C. A.; Ruiz-Pernía, J. J.; Tuñón, I. Unraveling the SARS-CoV-2 Main Protease Mechanism Using Multiscale Methods. *ACS Catal.* **2020**, *10*, 12544–12554.
- (24) Francés-Monerris, A.; Hognon, C.; Miclot, T.; García-Iriepa, C.; Iriepa, I.; Terenzi, A.; Grandemange, S.; Barone, G.; Marazzi, M.; Monari, A. Molecular Basis of SARS-CoV-2 Infection and Rational Design of Potential Antiviral Agents: Modeling and Simulation Approaches. *J. Proteome Res.* **2020**, *19*, 4291–4315.
- (25) Świderek, K.; Moliner, V. Revealing the molecular mechanisms of proteolysis of SARS-CoV-2 M pro by QM/MM computational methods. *Chemical Science* **2020**, *11*, 10626–10630.
- (26) Hillen, H. S.; Kokic, G.; Farnung, L.; Dienemann, C.; Tegunov, D.; Cramer, P. Structure of replicating SARS-CoV-2 polymerase. *Nature* **2020**, *584*, 154–156.
- (27) Malone, B.; Urakova, N.; Snijder, E. J.; Campbell, E. A. Structures and functions of coronavirus replication–transcription complexes and their relevance for SARS-CoV-2 drug design. *Nat. Rev. Mol. Cell Biol.* **2022**, *23*, 21–39.
- (28) Riccio, A. A.; Sullivan, E. D.; Copeland, W. C. Activation of the SARS-CoV-2 NSP14 3–5 exoribonuclease by NSP10 and response to antiviral inhibitors. *J. Biol. Chem.* **2022**, *298*, 101518.
- (29) Liu, C.; Shi, W.; Becker, S. T.; Schatz, D. G.; Liu, B.; Yang, Y. Structural basis of mismatch recognition by a SARS-CoV-2 proof-reading enzyme. *Science* **2021**, *373*, 1142–1146.
- (30) Seifert, M.; et al. Inhibition of sars-cov-2 polymerase by nucleotide analogs from a single-molecule perspective. *eLife* **2021**, *10*, No. e70968.
- (31) Chien, M.; Anderson, T. K.; Jockusch, S.; Tao, C.; Li, X.; Kumar, S.; Russo, J. J.; Kirchdoerfer, R. N.; Ju, J. Nucleotide Analogues as Inhibitors of SARS-CoV-2 Polymerase, a Key Drug Target for COVID-19. *J. Proteome Res.* **2020**, *19*, 4690–4697.
- (32) Vicenti, I.; Zazzi, M.; Saladini, F. SARS-CoV-2 RNA-dependent RNA polymerase as a therapeutic target for COVID-19. *Expert Opinion on Therapeutic Patents* **2021**, *31*, 325–337.
- (33) Beigel, J. H.; et al. Remdesivir for the Treatment of Covid-19 — Final Report. *New England Journal of Medicine* **2020**, *383*, 1813–1826.
- (34) Kokic, G.; Hillen, H. S.; Tegunov, D.; Dienemann, C.; Seitz, F.; Schmitzova, J.; Farnung, L.; Siewert, A.; Höbartner, C.; Cramer, P. Mechanism of SARS-CoV-2 polymerase stalling by remdesivir. *Nat. Commun.* **2021**, *12* (1), 279.

- (35) Driouich, J. S.; Cochin, M.; Lingas, G.; Moureau, G.; Touret, F.; Petit, P. R.; Piorkowski, G.; Barthélémy, K.; Coutard, B.; Guedj, J.; de Lamballerie, X.; Solas, C.; Nougairède, A. Favipiravir antiviral efficacy against SARS-CoV-2 in a hamster model. *bioRxiv*, 2020. DOI: 10.1101/2020.07.07.191775.
- (36) Özlüßen, B.; Kozan, S.; Akcan, R. E.; Kalender, M.; Yaprak, D.; Peltek, I. B.; Keske, S.; Gönen, M.; Ergönül, Ö. Effectiveness of favipiravir in COVID-19: a live systematic review. *European Journal of Clinical Microbiology and Infectious Diseases* **2021**, *40*, 2575–2583.
- (37) 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. Structure of the SARS-CoV-2 RNA-dependent RNA polymerase in the presence of favipiravir-RTP. *Proc. Natl. Acad. Sci. U. S. A.* **2021**, *118* (7), 118.
- (38) Aranda, J.; Wieczór, M.; Terrazas, M.; Brun-Heath, I.; Orozco, M. Mechanism of reaction of RNA-dependent RNA polymerase from SARS-CoV-2. *Chem. catalysis* **2022**, *2*, 1084–1099.
- (39) Wang, B.; Feig, M.; Cukier, R. I.; Burton, Z. F. Computational simulation strategies for analysis of multisubunit RNA polymerases. *Chem. Rev.* **2013**, *113*, 8546–8566.
- (40) Carvalho, A. T.; Fernandes, P. A.; Ramos, M. J. The catalytic mechanism of RNA polymerase II. *J. Chem. Theory Comput.* **2011**, *7*, 1177–1188.
- (41) Genna, V.; Vidossich, P.; Ippoliti, E.; Carloni, P.; De Vivo, M. A self-activated mechanism for nucleic acid polymerization catalyzed by DNA/RNA polymerases. *J. Am. Chem. Soc.* **2016**, *138*, 14592–14598.
- (42) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781–1802.
- (43) Phillips, J. C.; et al. Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J. Chem. Phys.* **2020**, *153*, 044130.
- (44) Titov, A. V.; Ufimtsev, I. S.; Luehr, N.; Martinez, T. J. Generating efficient quantum chemistry codes for novel architectures. *J. Chem. Theory Comput.* **2013**, *9*, 213–221.
- (45) Götz, A. W.; Clark, M. A.; Walker, R. C. An extensible interface for QM/MM molecular dynamics simulations with AMBER. *J. Comput. Chem.* **2014**, *35*, 95–108.
- (46) Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graphics* **1996**, *14*, 33–38.
- (47) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.
- (48) Mark, P.; Nilsson, L. Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *J. Phys. Chem. A* **2001**, *105*, 9954–9960.
- (49) Zgarbová, M.; Otyepka, M.; Šponer, J.; Mládek, A.; Banáš, P.; Cheatham, T. E.; Jurečka, P. Refinement of the Cornell et al. Nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. *J. Chem. Theory Comput.* **2011**, *7*, 2886–2902.
- (50) Macchiagodena, M.; Pagliai, M.; Andreini, C.; Rosato, A.; Procacci, P. Upgraded AMBER Force Field for Zinc-Binding Residues and Ligands for Predicting Structural Properties and Binding Affinities in Zinc-Proteins. *ACS Omega* **2020**, *5*, 15301–15310.
- (51) Bas, D. C.; Rogers, D. M.; Jensen, J. H. Very fast prediction and rationalization of pKa values for protein-ligand complexes. *Proteins: Struct., Funct., Genet.* **2008**, *73*, 765–783.
- (52) Hopkins, C. W.; Le Grand, S.; Walker, R. C.; Roitberg, A. E. Long-time-step molecular dynamics through hydrogen mass repartitioning. *J. Chem. Theory Comput.* **2015**, *11*, 1864–1874.
- (53) Miyamoto, S.; Kollman, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **1992**, *13*, 952–962.
- (54) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N log (N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (55) Hognon, C.; Monari, A. Staring at the Naked Goddess: Unraveling the Structure and Reactivity of Artemis Endonuclease Interacting with a DNA Double Strand. *Molecules* **2021**, *26*, 3986.
- (56) Lavery, R.; Moakher, M.; Maddocks, J. H.; Petkeviciute, D.; Zakrzewska, K. Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Res.* **2009**, *37*, 5917–5929.
- (57) Monari, A.; Rivail, J.-L. L.; Assfeld, X. Theoretical modeling of large molecular systems. Advances in the local self consistent field method for mixed quantum mechanics/molecular mechanics calculations. *Acc. Chem. Res.* **2013**, *46*, 596–603.
- (58) Chai, J.-D.; Head-Gordon, M. Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615.
- (59) McLean, A. D.; Chandler, G. S. Contracted Gaussian basis sets for molecular calculations. I. Second row atoms, Z = 11–18. *J. Chem. Phys.* **1980**, *72*, 5639–5648.
- (60) Rudbeck, M. Basis set dependence of phosphate frequencies in density functional theory calculations. *Int. J. Quantum Chem.* **2012**, *112*, 2435–2439.
- (61) Grossfield, A. An implementation of WHAM: the weighted histogram analysis method. *Analysis* **2004**, 1–13.
- (62) Rungrotmongkol, T.; Mulholland, A. J.; Hannongbua, S. Active site dynamics and combined quantum mechanics/molecular mechanics (QM/MM) modelling of a HIV-1 reverse transcriptase/DNA/dTTP complex. *Journal of Molecular Graphics and Modelling* **2007**, *26*, 1–13.
- (63) Castro, C.; Smidansky, E.; Maksimchuk, K. R.; Arnold, J. J.; Korneeva, V. S.; Götte, M.; Konigsberg, W.; Cameron, C. E. Two proton transfers in the transition state for nucleotidyl transfer catalyzed by RNA- and DNA-dependent RNA and DNA polymerases. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 4267–4272.
- (64) Bojin, M. D.; Schlick, T. A quantum mechanical investigation of possible mechanisms for the nucleotidyl transfer reaction catalyzed by DNA polymerase  $\beta$ . *J. Phys. Chem. B* **2007**, *111*, 11244–11252.
- (65) Lin, P.; Pedersen, L. C.; Batra, V. K.; Beard, W. A.; Wilson, S. H.; Pedersen, L. G. Energy analysis of chemistry for correct insertion by DNA polymerase  $\beta$ . *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 13294–13299.